2012
Electronic Imaging

SCIENCE AND TECHNOLOGY

22–26 January 2012

Technical Summaries

Conferences and Courses

22–26 January 2012

Hyatt Regency
San Francisco Airport Hotel

Burlingame, California, USA

electronicimaging.org

IS&T
imaging.org

SPIE
Connecting minds. Advancing light.

Contents

3D Imaging, Interaction, and Measurement

Imaging, Visualization, and Perception

Image Processing

Digital Imaging Sensors and Applications

Multimedia Processing and Applications

Visual Information Processing and Com munication

Conference 8288 · Stereoscopic Displays and Applications XXIII

Monday-Wednesday 23-25 January 2012

Part of Proceedings of SPIE Vol. 8288 Stereoscopic Displays and Applications XXIII

8288-01, Session 1

Matching stereo 3D reconstructions of CT/MRI data and intraoperative stereo video for medical education and treatment planning

J. F. Ilgner, M. Westhofen, Univ. Hospital Aachen (Germany)

Introduction: Anatomy of soft tissue structures in the Head and Neck is complex. While navigated surgery can maintain the surgeon's orientation intraoperatively, treatment planning is needed to choose the correct strategy for a successful surgical procedure. MRI and CT data usually rely on 3D reconstructions to obtain this objective.

Material and Methods: We took MRI and CT data from three different patients presenting with complex lesions in the Head and Neck. 3D reconstruction was already available by the time of pre-operative scanning, while stepping in rotated reconstructions was small enough to allow 3D stereo matching. Existing CT and MRI data had to be re-oriented to match the surgical situation.

Results: Re-orienting the radiological data required extra effort and about 30 min. per case, involving the radiologist as well as the ENT surgeon. The result, however, allowed CT and MRI data to be included in case presentations for junior and advanced surgeons, thus developing a common strategy for treatment planning and surgical setup.

Conclusion: The extra time and effort in matching stereo CT and MRI data with intraoperative stereo 3D video recordings data is justified for selected cases which serve as models for educating doctors in treatment planning.

8288-02, Session 1

Stereoscopic desktop VR system for tele-maintenance

M. Kleiber, Fraunhofer FKIE (Germany)

We have developed a tele-maintenance system which consists of an augmented reality (AR) system for the mechanic and a desktop virtual reality (VR) system for the expert. The expert has access to different data sources captured at the remote site and 3D construction data. The remote view of the mechanic is reconstructed in the VR system by using the tracking data from the AR system. The expert can use the 3D model to highlight parts, attach annotations or create animated instructions. To support the expert's spatial perception we have designed the system to produce stereoscopic output. The aim of the development was the integration of all necessary functions into a single stereoscopic application. Therefore, all necessary functions of the system had to be usable when viewed stereoscopically. Good stereoscopic viewing conditions were thus very important. This was the major challenge in regard to integrating the monoscopic sensor data and providing an intuitive user interface. We did not want to globally limit the stereoscopic effect. To control the amount of disparity we employ a shader based depth of field effect which is coupled to the point of focus of the user. The concept, implementation and evaluation will be described in the paper.

8288-03, Session 2

Angle-sensitive pixels: a new paradigm for low-power, low-cost 2D and 3D sensing

A. Wang, S. S. Hemami, A. Molnar, Cornell Univ. (United States)

We demonstrate a light-field camera which uses an image sensor composed of angle-sensitive pixels and a conventional camera lens. Angle-sensitive pixels are micro-scale devices which capture information about both the intensity and incident angle of the light they see. Because these pixels acquire a richer description of incident light that conventional intensity-sensitive pixels, our sensor only requires a simple camera objective to recover light-field information from a visual scene. Single images captured by our camera can be directly used for both synthetic refocus and depth map computation. The algorithms used for refocus and range-finding rely on the fact that each angle-sensitive pixel provides a response similar to one component of a 2D Hartley transform with a distinct frequency and orientation. By using several kinds of different pixels throughout the image sensor, we obtain a full, low-order Hartley transform of local angle, which is mapped to a local, spatial Hartley transform by the camera lens.

8288-04, Session 2

Polarizing aperture stereoscopic cinema camera

L. Lipton, Leonardo IP (United States)

SUMMARY: A new stereoscopic digital camera/lens for cinematography has been designed which is especially well suited for feature film and television applications where reduced interaxial separations are required. The current technology employed for such cameras uses so-called beam-splitter rigs that require the coordination of two cameras working together as if they were one stereoscopic camera. There are many engineering difficulties to overcome to make such a device work to specification, and given the assessment of the state of the art by many in the field, a new approach is required.

The new approach to be described in this paper uses polarizing left and right apertures of a suitably designed lens or lenses, working in combination with a digital sensor or sensors to provide a way to select the left and right images. The design allows for sensors of super 35 size, for example, to give cinematographers the ability to control depth of field as they have from the inception of 35mm cinematography. The camera/lens technology allows for a package to have a form factor and appearance similar to that of conventional production cameras with vastly simplified instrumentation and a high degree of confidence that what comes out of the camera does not need to be rectified in post-production.

8288-05, Session 2

Hardware-efficient, real-time 3D image processing for single lens color/depth-unified sensor: from depth noise reduction, depth super-resolution, to stereoscopic image synthesis

H. Lim, Y. S. Kim, O. Choi, B. Kang, S. Kim, K. Lee, J. D. Kim, C. Kim, Samsung Advanced Institute of Technology (Korea, Republic of)

The single lens color/depth-unified sensor [1] can provide real time color and time of flight depth images. In this paper, we present hardware-efficient algorithms for 3D image processing used in color/depth-unified sensors, including depth noise reduction, depth super-resolution, and stereoscopic image synthesis for instant watching in stereoscopic displays. For the effective depth noise reduction we utilize a separable nonlocal means filter that collects patches along vertical and horizontal directions from the target pixel position and performs weighted averaging of their center pixels with patch-wise depth similarity. For the super-resolution of the low depth image up to that of the color image we apply a hierarchical trilateral filter that uses three weights in term of pixel proximity, color similarity, and depth

similarity. Finally for the stereoscopic view synthesis the left side color image is generated by warping the original color image based on the disparity map converted from the depth image and we then recover the disoccluded regions using only the background boundary pixels without intrusion of the foreground pixel value.

All the above algorithms were implemented on Xilinx Virtex-5 LX330 and the computation time for a 1280x960 color image and a 320x240 depth image is 87ms at 54Mhz operating clock. As demonstrated in the attached figures, the proposed framework produces realistic stereoscopic image in real time.

[1] S.-J. Kim, J.D.K. Kim, S.-W. Han, B. Kang, K. Lee, and C.-Y Kim, "A 640×480 Image Sensor with Unified Pixel Architecture for 2D/3D Imaging in 0.11μm CMOS," in Symposia on VLSI circuits Dig. Tech. Papers, pp. 92-93, June, 2011.

8288-06, Session 2

Unassisted 3D camera calibration

K. Atanassov, V. Ramachandra, S. Goma, Qualcomm Inc. (United States)

With the rapid growth of 3D technology, 3D capture has becomes a critical part of the 3D feature. 3D quality is affected by the scene geometry as well as the on-the-device processing. An automatic 3D system usually assumes known cameras poses accomplished by a factory calibration using a special chart. In a real life operation, the result from the factory calibration can be negatively impacted by movements of the lens barrel due to shaking, focusing, or camera drop. If any of those factors displaces cameras optical axis such as vertical disparity exceeds the maximum tolerable margin, the 3D user may experience eye strain or headache. To make 3D capture more practical, one needs to consider unassisted (on arbitrary scene) calibration. In this paper we propose an algorithm that relies on detection and 2D matching of keypoints between left and right images. Frames containing erroneous matches are detected and discarded, as well as frames with insufficiently rich keypoints constellation. Roll, pitch yaw , and scale differences between left and right frames are then estimated. The algorithm performance is evaluated in terms of remaining vertical disparity as compared to the maximum tolerable vertical disparity.

8288-07, Session 2

Auto convergence for stereoscopic 3D mobile cameras

B. Zhang, S. Kothandaraman, A. U. Batur, Texas Instruments Inc. (United States)

The commercial success of 3D movies is generating great interest in stereoscopic 3D capture and display technologies. 3D capable TVs, digital cameras, and mobile devices are entering the consumer electronics market, which enable consumers capture and display their own 3D content. However, a major challenge to the success of these 3D capable devices is the viewing comfort. Consumer 3D cameras have fixed camera separation and orientation, and the 3D display viewing distance is typically short. For these reasons, the vergence-accommodation conflict is particularly pronounced, which causes discomfort and eye fatigue. In this paper, we present a Stereo Auto Convergence (SAC) algorithm for consumer 3D mobile cameras that reduces the vergence-accommodation conflict on the 3D display by adjusting the depth of the 3D scene automatically. Our algorithm processes stereo video in real-time and shifts each stereo frame horizontally by an appropriate amount in order to converge on a chosen object in that frame. The algorithm starts by estimating disparities between the left and right image pairs using correlations of the horizontal projections of the frame. The estimated disparities are then analyzed by the algorithm to select a point of convergence according to a center-convergence or surround-convergence strategy. The current and the target disparities of the chosen convergence point determine how much horizontal shift is needed. A disparity safety check is then performed to determine whether or not the maximum and minimum disparity limits have been exceeded after auto convergence. If the limits have been exceeded, further adjustments are made to satisfy the safety limits. Finally, convergence is performed by shifting the left and the right frames accordingly. Our solution has been implemented on a TI OMAP® processor running in real-time at 30 fps. It significantly improves the 3D viewing comfort.

8288-08, Session 3

Beyond parallax barriers: applying formal optimization methods to multi-layer automultiscopic displays

D. Lanman, Massachusetts Institute of Technology (United States); G. Wetzstein, Bauhaus Univ. Weimar (Germany); M. Hirsch, Massachusetts Institute of Technology (United States); W. Heidrich, The Univ. of British Columbia (Canada); R. Raskar, Massachusetts Institute of Technology (United States)

This talk focuses on resolving long-standing limitations of parallax barriers by applying formal optimization methods. This line of research, being pursued by the Camera Culture group at the MIT Media Lab, has led to two new attenuation-based displays: "HR3D" (http://www.hr3d. info) and "Layered 3D" (http://www.layered3d.info). The "High-Rank 3D (HR3D)" display consists of a stacked pair of modified LCD panels; rather than using heuristically-defined parallax barriers, both layers are jointly-optimized using low-rank light field factorization, resulting in increased brightness, refresh rate, and battery life for mobile applications. The "Layered 3D display" extends this approach to multi-layered displays composed of compact volumes of light-attenuating material. Such volumetric attenuators recreate a 4D light field or high-contrast 2D image when illuminated by a uniform backlight. Together, these projects reveal new generalizations to parallax barrier concepts, uniquely leveraging the compressive nature of natural 3D scenes for display applications. This talk primarily focuses, however, on introducing a new optimization framework that extends that developed in these prior works, being the first to allow joint multi-layer, time-multiplexed optimization. Simulations are compared to laboratory prototypes, establishing new capabilities for multi-layer, attenuation-based displays.

8288-09, Session 3

3D image quality of 200-inch glasses-free 3D display system

M. Kawakita, NHK Science & Technical Research Labs. (Japan); S. Iwasawa, National Institute of Information and Communications Technology (Japan); M. Sakai, Y. Haino, M. Sato, JVC KENWOOD Holdings, Inc. (Japan); N. Inoue, National Institute of Information and Communications Technology (Japan)

We have proposed a large glasses-free 3D display that could display 3D images on large screens (tens to hundreds of inches diagonally) by using multi-projectors and an optical screen consisting of a special diffuser film and large condenser lens. To achieve high-presence communication using 3D images, we numerically analyzed the factors responsible for degrading the image quality to increase the image size. A major factor used for determining the 3D image quality was the arrangement of component units, such as projector array and condenser lens, and the diffuser film characteristics. We designed and fabricated a prototype 200-inch glasses-free display system using the numerical results. We selected a suitable diffuser film and combined it with an optimally designed condenser lens. We used 64 high-definition projector units and obtained viewing angles of 13.5°. The gap between the parallax images was <2.5 cm and helped us realize natural 3D images with smooth moving parallax. The width of the viewing area was 1.3 m at a distance of 5.5 m; an observer could see the 3D images even though he/she moved more than ±2 m in the depth direction. This system could display life-size glasses-free 3D images of a car using natural parallax images.

8288-10, Session 3

Implementation and analysis of an autostereoscopic display using multiple liquid crystal layers

H. Gotoda, National Institute of Informatics (Japan)

Multilayer displays are usually constructed by stacking multiple liquid crystal panels along with a pair of linear polarizers and a light source. Previous theoretical analyses have indicated that such a display could show the light field of a 3D scene if the panels in the display are properly controlled. However, as far as we know, implementations realizing this feature have not been reported yet. In this study, we present an implementation of a monochromatic multilayer display using IPS-mode LCD panels, and report that a reasonable approximation of the light field is actually observed from the display.

A major obstacle to constructing a multilayer display is the complicated optical property of liquid crystal panels. Since an accurate analytic representation of the pixel's function is difficult to obtain, we have developed two models that approximately describes it. These two models are then integrated into a two-stage method, using which nearly real-time control of the panels in the display has been achieved. Finally, the quality of the light field observed from the prototype implementation is evaluated by examining the images photographed at several camera positions. The results show that fairly good images are observed within the viewing angle of 30 degrees.

8288-11, Session 3

360-degree, three-dimensional table-screen display using small array of high-speed projectors

S. Uchida, Y. Takaki, Tokyo Univ. of Agriculture and Technology (Japan)

A new 360-degree three-dimensional table-screen display is proposed, which combines the advantages of the previously proposed high-speed projector and multi-projector systems. The proposed system consists of a small number of high-speed projectors and a rotating screen. Because each high-speed projector is located outside the rotating axis of the screen, multiple projectors can be aligned above or below the rotating screen. The lens shift technique is used to superimpose multiple images generated by all projectors on the rotating screen. The screen has an off-axis lens function such that the rotation of the screen generates numerous viewpoints on a circle around the rotating screen. Each projector generates numerous viewpoints on a different circle. The use of multiple projectors enables the reduction of the rotation speed of the screen, the increase in the number of colors, and the increase in the number of viewpoints. An experimental system employing two DMD projectors was demonstrated. The rotating speed of the screen was reduced from 3,333 to 1,666 rpm in order to generate a 3D image with a frame rate of 56 Hz. Each projector generated 800 viewpoints. The diameter of the screen was 300 mm. The 3D resolution was 768 × 768.

8288-12, Session 4

Focus mis-match detection in stereoscopic content

F. Devernay, S. Pujades, V. Ch.A.V., INRIA Rhône-Alpes (France)

Live-action stereoscopic content production requires a stereo rig with two cameras that are precisely matched (in terms of focal length, colorimetry, etc.) and aligned (so that the stereo disparity is essentially horizontal). While most deviations from this perfect setup can be corrected either live or in post-production, a difference in the focus distance or focus range between the two cameras, which is difficult to check visually, will lead to unrecoverable degradations of the stereoscopic footage. We propose algorithms that check for focus differences between the left and right images, and give directions to the

operator to solve these problems. The proposed method was evaluated on ray-traced photorealistic scenes, and on real stereoscopic image pairs, and proves to perform well in most cases.

8288-13, Session 4

2D-to-3D image conversion: leveraging 3D data on the net

J. Konrad, G. Brown, M. Wang, P. Ishwar, Boston Univ. (United States)

Many methods have been proposed to date to convert 2D images to 3D stereopairs. The most successful ones involve human operators and, therefore, are time-consuming and costly. Fully-automatic 2D-to-3D conversion methods have not achieved yet the quality of operator-supervised methods. The main reason for this is the fact that automatic methods usually rely on assumptions about the captured 3D scene that are often violated in practice. In this paper, we explore a radically different idea inspired by our work on saliency detection in images; instead of deterministically specifying a scene model for the input 2D image, we propose to "learn" the model, and more specifically to learn the scene depth, from a large dictionary of stereopairs, such as YouTube 3D. The main contribution of our work is the realization that among millions of stereopairs on-line there likely exist very many stereopairs with similar content to that of the 2D query. We exploit this observation by constructing, from the on-line stereopairs, the depth of 3D scene captured by the 2D query. First, we perform the so-called "k nearest neighbor" (kNN) search for our 2D query using, for example, the left image from all on-line stereopairs. This returns an ordered list of left images, from the most to the least similar vis-a-vis the 2D query. We discard all but the top k matches (kNNs). Then, we estimate a geometric transformation between the 2D query and each of the kNNs using the SIFT flow. This step is needed to account for differences in scale, orientation, position, etc., of objects between the 2D query and each kNN. Finally, assuming that depth is known for each on-line stereopair, we use the inverse of the above transformation to obtain the depth of the 2D query. In practice, in the last step we use disparity instead of depth and since the geometric transformation is not invertible we use an approximate inverse. Having obtained k disparity fields from k nearest neighbors, we fuse them by computing the median. We use the obtained disparity field to generate the right image from the 2D query while handling occlusions and newly-exposed areas in the usual way. To date, we have implemented a simplified version of the above algorithm that demonstrates the proof-of-concept. With the continuously increasing amount of 3D data on-line and with the rapidly growing computing power in the cloud the proposed framework seems a promising alternative to operator-assisted 2D-to-3D conversion.

8288-14, Session 4

3D cinema to 3DTV content adaptation

D. Doyen, L. Yasakethu, L. Blondé, Q. Huynh-Thu, Technicolor S.A. (France)

3D cinema and 3DTV have grown in popularity in recent years. Filmmakers have a significant opportunity before them given the recent success of 3D films. In this paper we investigate whether this opportunity could be extended to the home in a meaningful way.

"3D" perceived from viewing stereoscopic content depends on the viewing geometry. This implies that the stereoscopioc-3D content should be captured for a specific viewing geometry in order to provide a satisfactory 3D experience. However, although it would be possible, it is clearly not viable, to produce and transmit multiple streams of the same content for different screen sizes. In this study to solve the above problem, we analyze the performance of six different disparity based transformation techniques, which could be used for cinema-to-3DTV content conversion. Subjective tests are performed to evaluate the effectiveness of the algorithms in terms of depth quality, visual comfort and artefacts such as crosstalk, cardboard effect and puppet theatre effect. The resultant 3DTV experience is also compared to that of cinema. We show that by selecting the proper transformation technique, based on content type, it is possible to enhance the 3DTV

experience from a common input signal which was originally captured for cinema.

8288-83, Session 4

Video retargeting for stereoscopic content under 3D viewing constraints

C. Chamaret, Technicolor S.A. (France)

The imminent deployment of new devices such as TV, tablets, smartphones supporting stereoscopic display creates a need for retargeting the content. New devices bring their own aspect ratio and potential small screen size. Aspect ratio conversion becomes mandatory and an automatic solution will be of high value especially if it maximizes the visual comfort.

Some issues inherent to 3D domain are considered: no vertical disparity, no object having negative disparity (outward perception) on the border of the cropping window.

An efficient visual attention model is applied on each view and provides saliency maps with most attractive pixels. Dedicated 3D retargeting correlates the 2D attention maps for each view as well as additional computed information to ensure the best cropping window. Specific constraints induced by 3D experience are fixed by computing a map presenting objects that should not be cropped.

The comparison with original content of 2:35 ratio having black stripes provide limited 3D experience on TV screen, while the automatic cropping and exploitation of full screen show more immersive experience.

The proposed system is fully automatic, ensures a good final quality without missing fundamental parts for the global understanding of the scene. In addition, it improves the immersive experience.

8288-16, Session 5

Stereoscopic-3D display design: a new paradigm with Intel adaptive stable image technology [IA-SIT]

S. K. Jain, Intel Corp. (United States)

Stereoscopic-3D (S3D) proliferation on personal computers is limited by several challenges: 1) lower effective resolution, luminance and color gamut of current S3D displays compared to 2D; 2) viewing discomfort due to cross-talk amongst stereo images; 3) high system cost; and 4) restricted content availability. Intel Adaptive Stable Image Technology (IA-SIT) is a system architecture approach developed to address such S3D system design challenges. Our goal is to deliver high quality 3D visualization at PC price points. Optimizations in display driver, timing firmware, backlight hardware, eyewear optical stack and synch mechanism can together help accomplish this goal. Agnostic to refresh rate, IA-SIT naturally scales with shrinking of display transistors and improvements in liquid crystal and LED materials. We propose: 1) new IA-SIT display timings in 3D mode (through VESA); 2) IA-SIT optical stack for 3D Eyewear that is truly universal across active/passive and PC/TV devices (through CEA); and 3) IA-SIT real time control-synchronization profile (through BlueTooth Sig).

8288-17, Session 5

A real-time misalignment correction algorithm for stereoscopic 3D cameras

I. E. Pekkucuksen, B. Zhang, A. U. Batur, Texas Instruments Inc. (United States)

Camera calibration in a 3D stereo setup is an important problem since the misalignment between the two views can significantly degrade the 3D quality. However, hardware calibration is not always an option especially for mass produced stereo cameras. Even if one-time calibration is performed, it cannot be maintained indefinitely because

of environmental factors. That is why software calibration offers a more flexible option to address the problem. In this paper, we propose a stereo misalignment correction algorithm that does not require the use of special charts. The algorithm first estimates the disparity in horizontal and vertical directions between corresponding blocks from both images. Then, this initial estimate is refined with two dimensional search using smaller sub-blocks. The displacement data and block coordinates are fed to a modified affine transformation model and outliers are discarded to keep the modeling error low. Finally, the estimated affine parameters are portioned between the two views to compensate for the misalignment. The proposed algorithm can significantly reduce the misalignment between stereo frames and enable a better 3D experience in real time.

8288-18, Session 5

Full-color stereoscopy with little flicker at low-refresh rate by time-division multiplexing anaglyph

H. Kakeya, H. Kodaira, Univ. of Tsukuba (Japan)

In the present paper we propose a time-division multiplexing anaglyph method to realize full color stereoscopy with little flicker at low refresh rate of around 60Hz, which is compatible with the conventional 2D displays. To let more people experience 3D contents at home, it is preferable that people can watch 3D contents with the conventional 2D TV sets. Because of the low refresh rate, however, applying time-division multiplexing method using shutter glasses to conventional displays results in severe flicker, which is far from acceptable to almost all consumers. To overcome this problem, we propose a time-division multiplexing anaglyph method, where the red and blue components of right image is shown to the right eye and the green component of left image is shown to the left eye at odd frames, while the green component of right image is shown to the right eye and the red and blue components of left image is shown to the left eye at even frames. We carry out an experiment to let the subjects see time-division multiplexing anaglyph images and the result shows that flicker can be reduced to an acceptable level by the proposed method.

8288-19, Session 5

Supervised disparity estimation

P. Vandewalle, C. Varekamp, Philips Research Nederland B.V. (Netherlands)

Automated disparity estimation from stereo input video has been widely investigated. While the results using such methods have strongly improved in recent years, they are typically not sufficiently accurate yet for general usage. The main remaining problems occur in occlusion regions, homogeneous areas, and depth transitions between objects with low color contrast. User interaction can be applied to overcome these problems and achieve any desired quality level.

We introduce a supervised disparity estimation method in which an operator can steer the disparity estimation process. Instead of correcting errors, we view the estimation process as a constrained process where the constraints are indicated by the user in the form of control points, scribbles and contours. Control points are used to obtain accurate disparity estimates that can be fully controlled by the operator. Scribbles are used to force regions to have a smooth disparity, while contours create a disparity discontinuity in places where automatic methods fail. Control points, scribbles and contours are propagated through the video sequence using motion compensation to create temporally stable results.

8288-20, Session 5

Analysis of brain activity and response during monoscopic and stereoscopic visualization

E. Calore, R. Folgieri, D. Gadia, D. Marini, Univ. degli Studi di Milano (Italy)

Stereoscopic visualization in cinematography and VR creates an illusion of depth by means of two bidimensional images corresponding to different views of a scene. This perceptual "trick" is used to enhance the emotional response and the sense of presence and immersivity of the observers.

An interesting question is if and how is possible to measure and analyze the level of emotional involvment of the observers during a stereoscopic visualization of a movie or of a virtual environment.

The final goal of this research is a challenge, due to the large number of sensorial, physiological and cognitive stimuli involved.

In this paper we begin this research by analyzing eventual differences in the brain activity of subjects during the observation of monoscopic or stereoscopic contents. To this aim, we have performed some experiments collecting EEG data using a Brain-Computer Interface device from two groups of users, during the observation of stereoscopic and monoscopic short movies inside the Virtual Theater of the University of Milan.

From the analysis of the collected data, it seems that interesting differences are present in the average brain activity among the observers in the two groups, with a significative effect of stereoscopic visualization.

8288-21, Session 6

Motion in-depth constancy in stereoscopic displays

S. Laldin, L. Wilcox, C. Hylton, R. S. Allison, York Univ. (Canada)

In a stereoscopic 3D scene, non-linear mapping between real space and disparity can produce distortions of space when camera geometry differs from natural stereoscopic geometry. If an object traverses a distorted stereoscopic space at constant velocity, one might anticipate distortion of the perceived trajectory. To determine if the predicted distortions are in fact perceived, we assessed observers' percepts of acceleration and deceleration using an animation of a ball moving in depth through a simulated environment, viewed stereoscopically.

The method of limits was used to measure transition points between perceived acceleration and deceleration as a function of IA and context (textured vs. non-textured background).

Based on binocular geometry, we predicted that the transition points would shift toward deceleration for small IA and towards acceleration for large IA. This effect should be modulated by monocular depth cues. However, we found that the average transition values were not influenced by IA or the simulated environment. These data suggest that observers are able to discount distortions of stereoscopic space in interpreting the trajectory of objects moving through simple environments.

8288-22, Session 6

Quality evaluation of depth image-based rendering methods for stereoscopic video content

M. Eisenbarth, Technische Univ. Wien (Austria) and emotion3D (Austria); F. H. Seitner, emotion3D (Austria) and Technische Univ. Wien (Austria); M. Gelautz, Technische Univ. Wien (Austria)

Depth image-based rendering (DIBR) allows the generation of new camera viewpoints from stereoscopic footage. This enables a wealth of new applications for 3D film post-production such as depth correction and grading, content remastering and multi-view generation for auto-stereoscopic displays.

We compared state-of-the-art DIBR techniques and their warping and in-filling methods. Every method has its own advantages when applied to specific content, but generates individual artifacts if content changes over time. We made objective and subjective comparisons of DIBR

generated stereoscopic content on 3D displays. We designed a test image to measure the depth boundaries of 3D displays. Measurement of depth impression in video input and of displayed output allows the DIBR challenges to be quantized. Based on this method, an evaluation of DIBR techniques with different warping and in-filling techniques was carried out.

Our results demonstrate both a proper objective and subjective comparison of DIBR algorithms according to the visual perception of warping and in-filling artifacts. The study also shows the critical factors in generating virtual camera viewpoints or performing stereo-base and convergence corrections given a stereo video.

8288-23, Session 6

Disparity profiles in 3DV applications: overcoming the issue of heterogeneous viewing conditions in stereoscopic delivery

G. Boisson, C. Chamaret, Technicolor S.A. (France)

More and more numerous 3D movies are released each year. Thanks to the current spread of 3D-TV displays, these 3D Video (3DV) contents are about to enter massively the homes. Yet viewing conditions determine the stereoscopic features achievable for 3DV material. Because the conditions at home - screen size and distance to screen - differ significantly from a theater, 3D Cinema movies need to be repurposed before broadcast and replication on 3D Blu-ray Discs for being fully enjoyed at home. In that paper we tackle that particular issue of how to handle the variety of viewing conditions in stereoscopic contents delivery. To that extend we first investigate what is basically at stake for granting stereoscopic viewers' comfort, through the well-known - and sometimes dispraised - vergence-accommodation conflict. Thereby we define a set of basic rules that can serve as guidelines for 3DV creation. We propose disparity profiles as new requirements for 3DV production and repurposing. Meeting proposed background and foreground constraints prevents from visual fatigue, and occupying the whole depth budget available grants optimal 3D effects. We present an efficient algorithm for automatic disparity-based 3DV retargeting depending on the viewing conditions. Variants are proposed depending on the input format (stereoscopic binocular content or depth-based format) and the level of complexity achievable.

8288-24, Session 6

Investigation of object thickness for visual discomfort prediction

in stereoscopic images

H. Sohn, Y. J. Jung, S. Lee, H. W. Park, Y. M. Ro, KAIST (Korea, Republic of)

With the rising popularity of stereoscopic 3D content services, concerns on the 3D viewing safety issues have been raised. To address the viewing safety issues, the automatic prediction of visual discomfort for stereoscopic 3D contents has increasingly gained its importance. It has been well known that stereoscopic 3D displays induce the conflict between accommodation and vergence [1, 2]. Specifically, excessive amount of binocular disparity (i.e., disparity magnitude) leads to high degrees of the conflict between accommodation and vergence, which may cause visual discomfort and fatigue [1, 2]. In addition, it was reported that as disparity magnitude of visual stimulus increased, more visual discomfort was induced [1, 3]. As a result, disparity magnitude has been regarded as a central determinant for visual discomfort prediction of stereoscopic 3D images [1-4].

In stereoscopic viewing, human visual system has a limited ability to fuse binocular disparity, i.e., binocular fusion limit. In literature, it has been known that the binocular fusion limit is affected by not only disparity magnitude but also a horizontal size of visual stimulus [5]. In particular, given the disparity magnitude, decreases in stimulus size also decrease binocular fusion limit. In viewing stereoscopic 3D displays, it was reported that the crosstalk of thin objects was more visible and affected depth perception than that of thicker objects [6]. As

a result, these clues indicate that visual comfort may be affected by the width of objects in natural stereoscopic 3D scene contents.

The purpose of this paper is to investigate the influence of object thickness on visual discomfort prediction for stereoscopic scenes. In particular, we demonstrate that the combined use of disparity magnitude and object thickness substantially improves the performance in visual discomfort prediction.

8288-25, Session 7

History of autostereoscopic cinema

W. Funk, Hologlyphics (United States)

This paper covers the history of autostereoscopic cinema, from the beginnings of autostereoscopy in the 1800's, the development of motion capability and it's subsequent evolution to present techniques. Public viewings of autostereoscopic movies have occurred on a semi-ongoing basis since the early 1940's.

A theater in Moscow, called the Stereokino, was constructed for

showing autostereoscopic films, with specially positioned seating for proper viewing. In 1941, the first autostereoscopic movie was shown there, Kontsert directed by Aleksandr Andreyevskiy. The first autostereoscopic feature film was Robinson Kruzo, released in 1947. Additional stereokinos were opened across the Soviet Union, with films produced and shown until the late 1950s.

The Cyclostéréoscope was an autostereoscopic cinema system invented by François Savoye in France. It was based around a drum made of metal bars that revolve around a screen. For several years in the 1940's and 1950's, it was open to the public in Paris. Any film made in a dual film format could be shown.

Besides dedicated theaters in Russia and France, exhibits of content have occurred outside devoted theaters. The paper focuses on the history of autostereoscopic technology developed for entertainment, public viewing of content, the individuals involved and the content itself.

8288-26, Session 7

Full optical characterization of auto-stereoscopic 3D displays using local viewing angle and imaging measurements

P. M. Boher, T. Leroux, T. Bignon, ELDIM (France)

Different commercial auto-stereoscopic 3D displays are characterized using Fourier optics viewing angle system and imaging video-luminance-meter. The emissive properties of the display are first measured locally at three positions (center, right and left). Qualified monocular and binocular viewing spaces at deduced for an observer in front of the display and the best working distance is deduced. Then the imaging system is positioned at this working distance and laterally shifted to take a series of luminance measurements. Crosstalk homogeneity on the entire surface of the display is deduced for all the observer locations along the horizontal. Results are correlated to those obtained locally using high angular resolution Fourier optics viewing angle instrument. We show that the crosstalk is generally optimized on most of the surface of the display except at the corners depending on the display technology and the alignment of the parallax barriers or cylindrical lenses. Local imperfections like scratches and mark generally increase drastically the crosstalk and demonstrating the cleanliness requirements of this type of display.

8288-27, Session 7

Analysis on monocular accommodation in horizontal-parallax-only super-multiview display

J. Hong, Y. Kim, Seoul National Univ. (Korea, Republic of); J. Park, Chungbuk National Univ. (Korea, Republic of); B. Lee, Seoul National Univ. (Korea, Republic of)

Accommodation response to the horizontal-only super-multiview display was measured to assess if the super-multiview condition holds for the horizontal-only multiview display. Experimental results show that the monocular depth cue does not perfectly match to the virtual image of horizontal-only super-multiview display, but the degree of accommodation-convergence mismatch can be reduced because accommodation response tends to approach the virtual image. We are planning to quantify the degree of reducing the accommodation-convergence mismatch by the horizontal-only super-multiview display and the data will be helpful in commercializing the multiview display system.

8288-28, Session 7

A novel time-multiplexed, autostereoscopic, multi-view, full-resolution 3D display

J. Liou, Industrial Technology Research Institute (Taiwan)

We propose and experimentally demonstrate a novel time-multiplexed autostereoscopic multi-view full resolution 3D display based on the lenticular lens array in association with the control of the active dynamic LED backlight. The lenticular lenses of the lens array optical system receive the light and deflect the light into each viewing zone in a time sequence. The crosstalk under different observation scanning angles is showed, including the cases of 4-views field scanning. The crosstalk of any view zones is about 5% respectively; the results are better than other 3D type.

This paper addresses the specific technological challenges of autostereoscopic 3D displays and presents a novel system that integrates a time-multiplex autostereoscopic display based on active directional backlight (active dynamic backlight) with an autostereoscopic display. Our successfully designed prototype utilized a FPGA system to synchronize between a display panel and backlight slit panel.

Multiple viewers can view the 3D image display with different contents at different display regions of the round display panel. Taking four viewing zones as an example to be created, each viewing zone occupies 1/240 second for display playing one 2D image. One eye receives the first image in 1/240 second at one viewing zone and another eye receives another image in next 1/240 second at the adjacent viewing zone.

8288-29, Session 8

Crosstalk reduces the amount of depth seen in 3D images of natural scenes

I. Tsirlin, R. S. Allison, L. M. Wilcox, York Univ. (Canada)

Crosstalk remains an important determinant of S3D image quality. Defined as the leakage of one eye's image into the image of the other eye it affects all commercially available stereoscopic viewing systems. Previously we have shown that crosstalk affects perceived depth magnitude in S3D displays. We found that perceived depth between two lines separated in depth decreased as crosstalk increased. The experiments described here extend our previous work to complex images of natural scenes. We controlled crosstalk levels by simulating them in images presented on a zero-crosstalk mirror stereoscope display. The observers were asked to estimate the amount of stereoscopic depth between pairs of objects in stereo-photographs of clattered rooms. Data show that as crosstalk increased perceived depth decreased; an effect that occurred at all disparities. Interestingly, observers overestimated the depth in displays that contained no crosstalk. A control experiment shows that it is likely due to the presence of pictorial cues to depth. Taken together these results demonstrate that our previous findings generalize to natural scenes showing that crosstalk affects perceived depth magnitude even in the presence of pictorial depth cues.

8288-30, Session 8

Method and simulation to study 3D crosstalk perception

D. A. Khaustova, L. Blondé, Q. Huynh-Thu, C. Vienne, D. Doyen, Technicolor S.A. (France)

3D image/video quality depends on the amount of crosstalk exhibited by displays. A good understanding of the crosstalk phenomenon may lead to improved 3D technologies. With this perspective, our work took into account display technologies, content characteristics and human perception. On tested displays, system crosstalk was described as a 2D surface (XT-surface) by measuring luminance variations compared to a no-crosstalk reference. The XT-surface represents light intensities for intended (deliberate signal) and unintended (leaking signal from the other view) pixel values. This data surface was used for the creation of a controlled simulation of crosstalk. The set-up of a test bench used high-contrast displays and mirrors to realize a zero crosstalk stereoscopic system. By dividing the XT-surface into areas, it was possible to control the simulation and examine the influence of different types and levels of crosstalk on human perception. Crosstalk perception was studied through a series of psychovisual tests. Results demonstrate that not only white-to-black crosstalk has crucial influence on image quality. By modifying parameters of the XT-surface it was possible to find relations between crosstalk amplitude and human perception, as well as to see how modifications influence judgments about image quality.

8288-31, Session 8

Reproducibility of crosstalk measurements on active glasses 3D LCD displays based on temporal characterization

S. Tourancheau, Mid Sweden Univ. (Sweden); K. Wang, Acreo AB (Sweden) and Mid Sweden Univ. (Sweden); L. Janowski, J. Bulat, AGH Univ. of Science and Technology (Poland); K. Brunnström, Acreo AB (Sweden); M. Barkowsky, Polytech' Nantes (France)

Crosstalk is one of the main display-related perceptual factors degrading image quality and causing visual discomfort on 3D-displays. It causes visual artifacts such as ghosting effects, blurring, and lack of color fidelity which are considerably annoying and can lead to difficulties to fuse stereoscopic images. On stereoscopic LCD with shutter-glasses, crosstalk is mainly due to dynamic temporal aspects: imprecise target luminance (highly dependent on the combination of left-view and right-view pixel color values in disparity regions) and synchronization issues between shutter-glasses and LCD. These different factors influence largely the reproducibility of crosstalk measurements across laboratories and need to be evaluated in several different locations involving similar and differing conditions.

In this paper we propose a fast and reproducible measurement procedure for crosstalk based on high-frequency temporal measurements of both display and shutter responses. It permits to fully characterize crosstalk for any right/left color combination and at any spatial position on the screen. Such a reliable objective crosstalk measurement method at several spatial positions is considered a mandatory prerequisite for evaluating the perceptual influence of crosstalk in further subjective studies.

8288-32, Session 8

Crosstalk and brightness in projection-based, multi-view systems

R. Blach, A. Pross, Fraunhofer-Institut für Arbeitswirtschaft und Organisation (Germany); A. Kulik, Bauhaus Univ. Weimar (Germany); O. Stefani, Fraunhofer-Institut für Arbeitswirtschaft und Organisation (Germany)

Immersive working environments generally use single user stereoscopic projections. We have extended these systems to multi view systems from two to six users with individual perspectively correct stereoscopic views for each user. To separate the images we have used a hybrid approach which combines active and passive filtering.

In projection based view systems brightness and crosstalk are one of the major issues which have to be handled as the noise increases and the signal brightness decreases with the increase of views. To be able to evaluate different technical approaches for optimization and also compare these systems with existing single user 3D systems, we have extended the crosstalk measurement approach by Weissman and Woods to our specific multi view case which is: a) projection based, b) eye separation with optical filters and c) user separation with active shutter elements.

The main difference compared to the single user stereo approach is that crosstalk or noise is not only evoked by a single unintended image but by several.

Our contribution is a simple approach for defining and measuring crosstalk for projection based multi view systems. As these systems are still rare, we are not aware that this has been explicitly shown elsewhere.

8288-33, Session 8

How much crosstalk can be allowed in a stereoscopic system at various grey levels?

S. A. Shestak, D. Kim, Samsung Electronics Co., Ltd. (Korea, Republic of)

We derived an expression for physiological limit of crosstalk perception, based on Barten's model of contrast sensitivity. The physiological limit varies with the displayed luminance and with maximum display's luminance. The calculated threshold values are close to the previously published level 0.5-1% just within the certain range of luminance. At low luminances the perceptual threshold can be noticeably lower. We propose to specify the allowable crosstalk X_a as a number (N) of physiological limits $X_a=NX_p$, where N=2..3.

8288-19, Session 9A

Apparent stereo: the Cornsweet illusion can enhance perceived depth

P. Didyk, Max-Planck-Institut für Informatik (Germany); T. Ritschel, E. Eisemann, Telecom ParisTech (France); K. Myszkowski, H. Seidel, Max-Planck-Institut für Informatik (Germany)

It is both a technical and an artistic challenge to depict three-dimensional space using a flat two-dimensional medium. Fitting within the limits of a given display technology and at the same time achieving comfortable viewing can require modifications to stereo content, e. g., to flatten or even remove binocular disparity. However, in such scenarios crucial depth details might get lost. Recently, a backward-compatible stereo technique was proposed, which builds upon the Craik-O'Brien-Cornsweet effect - a visual illusion, which was applied to brightness in CG before. The approach replaces disparity by Cornsweet profiles in disparity where depth discontinuities occur. This allows for the reduction of the overall disparity range while still conveying a comfortable stereo experience. The present work extends the previous idea by showing that Cornsweet profiles can also be used to enhance the 3D impression in regions where the disparity range was compressed. A user study measures the performance of backward-compatible stereo as well as our disparity enhancement.

8288-20, Session 9A

Perceived depth of multi parallel, overlapping, transparent, stereoscopic surfaces

S. Aida, K. Shimono, Tokyo Univ. of Marine Science and Technology (Japan); W. J. Tam, Communications Research Ctr.

Canada (Canada)

The geometric relational expression of horizontal disparity, viewing distance, and depth magnitude between objects in stereopsis suggests that, for a given viewing distance, the magnitude of perceived depth of objects would be the same as long as the disparity magnitudes are the same. However, we found that this is not necessarily the case for random dot stereograms depicting two or more parallel-transparent-overlapping-stereoscopic-surfaces (POTS). Data from two experiments indicated that, when the stimulus size is relatively large (e.g., 13 x 20 arc deg), the magnitude of reproduced depth between two stereo-surfaces of a two-POTS configuration can be larger than that between an identical pair of stereo-surfaces but with an additional stereo-surface located between the pair. In follow-up experiments, dot density, the size of the random dot stereograms and the number of stereo-surfaces lying between the two outermost stereo-surfaces were manipulated. Overall, the results indicate a difference of approximately 10% in the reproduced depth, depending on whether there is an additional stereo-surface between the two surfaces of a two-POTS configuration or not. The results are discussed in terms of global stereopsis which "operates" for relatively large stimulus sizes (such as 13 x 20 arc deg).

8288-21, Session 9A

Diagnosing perceptual distortion present in group stereoscope viewing

M. M. Burton, B. B. Pollock, J. W. Kelly, S. B. Gilbert, E. Winer, Iowa State Univ. (United States); J. de la Cruz, U.S. Army RDECOM/STTC (United States)

Stereoscopic virtual environments are frequently employed to immerse viewers in virtual reality visualizations, simulations, and training systems. Stereo systems render the scene graph from the perspective of a single person or static viewpoint which can be defined as a leader, while others see inaccurate and distorted views from a follower position. Greater understanding of distortion from the follower perspective can lead to more effective multi-view solutions and configurations of leader-follower systems.

An experiment at Iowa State University was conducted to gain insight into the perceptual distortion present in leader-follower systems where multiple users view the same stereoscopic, projection-based virtual environment. The experiment was comprised of tasks reporting quantities of angles in degrees from positions parallel to the screen and reporting the depth dimension of a rectangle at positions centered and perpendicular to the screen. Results from this study indicated a relationship of angular distortion with leader-follower offsets parallel to the screen and distance to the screen affects the quantity of depth distortion in addition to perpendicular offsets from the rendered viewpoint. Further work will attempt to understand the effects of this distortion on group collaboration and further develop relationships of distortion and leader-follower offsets.

8288-34, Session 9A

Visual discomfort and the timing of vergence-accommodation conflicts

J. Kim, D. Kane, M. S. Banks, Univ. of California, Berkeley (United States)

Stereo 3D (S3D) displays cause visual discomfort in many viewers, so it is very important to understand the causes of the discomfort so one can minimize it. We know that the vergence-accommodation conflict causes visual discomfort and eventually visual fatigue. Vergence and accommodation both have phasic and tonic components that adapt at different rates. To determine if the different adaptation rates in phasic and tonic components affect visual discomfort, we varied the temporal properties of the vergence-accommodation conflict and assessed discomfort.

Observers viewed random-dot stereograms in a novel volumetric display that allows us to independently manipulate the stimulus to vergence and the stimulus to accommodation. There were two main conditions: natural viewing (cues-consistent) and S3D viewing (cues-inconsistent). The stimuli to vergence and accommodation were varied at different three different rates, presented in different sessions.

We found that observers experienced worse symptoms in the S3D viewing than in the natural viewing. We also found that more frequent changes in the stimulus to vergence yielded worse symptoms. Thus, the time-average magnitude of the vergence-accommodation conflict is not the sole predictor of the amount of discomfort. It also matters how the conflict is distributed over time.

8288-22, Session 9B

Measuring 3D discomfort from vertical and torsional disparities in natural images

C. W. Tyler, L. T. Likova, The Smith-Kettlewell Eye Research Institute (United States); K. Atanassov, V. Ramachandra, S. Goma, Qualcomm (United States)

The two major aspects of camera misalignment that cause visual discomfort when viewing images on a 3D display are vertical and torsional disparities. The goal of the study was to determine the discomfort ranges for the kinds of natural image that people are likely to take with 3D cameras. The results showed that, as a rule of thumb, vertical disparity between the two eyes needs to be held below about 15 arcmin, or a quarter of a degree, to be within an acceptable range of discomfort, while human vision is relatively insensitive to torsional disparities below about 30 deg,

8288-35, Session 9B

Visual fatigue versus eye-movements

C. Vienne, L. Blondé, D. Doyen, Technicolor S.A. (France)

Observing 3D content on a cinema or TV screen potentially generates fatigue. In research, experience of visual symptoms following the observation of stereo-content is usually assessed thanks to questionnaires and subjective reports. We attempted to explore the occurrence of visual fatigue using more objective methods, namely by using binocular eye-tracking and psychophysics. A main objective was to study the emergence of visual fatigue in relation with eye-movement knowing the stimulation of the oculomotor system and its response. We designed an experiment in which participants were asked to perform a repeated vergence effort task, just followed by a 3D space perception task. Participants' eye movements were recorded during the whole session using an eye-tracking system. The analysis revealed that the perception of 3D shape was gradually affected by the intensity of the vergence effort task. The effect on stereo-estimation was actually due to the occurrence of visual fatigue, 3D objects are perceived flatter. Results on the subjective reports of SSQ revealed that oculomotor factors were predominant in the visual symptoms. In addition, some effects and correlations on the micro-saccadic rate were obtained. It offers a perspective to characterize objectively visual fatigue when watching 3D content on stereoscopic displays.

8288-36, Session 9B

Visual comfort: stereoscopic objects moving in the horizontal and mid-sagittal planes

W. J. Tam, F. Speranza, C. Vázquez, R. Renaud, Communications Research Ctr. Canada (Canada); N. Hur, Electronics and Telecommunications Research Institute (Korea, Republic of)

The present study examines the effect of motion of stereoscopic objects on visual comfort using computer-generated video sequences based on a parallel-camera setting. In the first experiment, the visual comfort of a set of stereoscopic objects moving in the z-plane with different disparities and velocities is examined. Three types of movement in the z-plane were included: both uncrossed and crossed disparities; crossed only; and uncrossed only. In a second experiment, the visual comfort of a set of stereoscopic objects moving at a given

depth in the horizontal plane is examined. Also in this case, three types of movement in the z-plane were included: both uncrossed and crossed disparities; crossed only; and uncrossed only. In both experiments, the movements (along the z-plane for Exp. 1 and along the x-plane for Exp. 2) involved three levels of velocity and five levels of disparity. Preliminary results appear to confirm that visual comfort is influenced more by object velocity than by disparity magnitude (for the range of magnitudes used); an interaction of velocity and disparity is also apparent. Therefore, the preliminary results appear to confirm that changes in disparity over time are a significant contributor to visual comfort of stereoscopic video images; the effect appears to be present also when the changes are confined within the generally accepted comfort zone.

8288-37, Session 9B

Visual discomfort with stereo 3D displays when the head is not upright

D. Kane, R. Held, M. Banks, Univ. of California, Berkeley (United States)

Stereo 3D (S3D) displays can cause visual discomfort. In part, discomfort is due to the well-known vergence-accommodation conflict. Additionally, discomfort may be caused by viewing S3D displays with the head rotated to the side (roll rotation). Head roll converts horizontal disparity on the screen into vertical disparity on the retina that in turn should elicit vertical vergence eye movements to minimize vertical disparity. Because such movements are small in natural viewing, the larger ones required with head roll are likely to cause discomfort. We investigated this by simulating head roll in a S3D display. We kept the observers' head upright, but rotated the stimulus. The stimulus was a random-dot stereogram depicting sinusoidal depth corrugations. It contained either no variation in overall disparity or a variation in overall disparity over time. Observers indicated the amount of visual discomfort experienced when viewing these stimuli. They reported worse symptoms in the simulated head-roll conditions and when the disparity pedestal changed by a greater amount. We attribute the increased symptoms to the vertical vergence eye movements elicited by head roll.

8288-54, Poster Session

Correlation between a perspective distortion in a S3D content and the visual discomfort perceived

D. Doyen, L. Blondé, J. Sacré, Technicolor S.A. (France)

Perspective distortion will occur in S3D when the relative disparity between elements generates a depth that is not in accordance with the relative size of objects presented. Subjective tests have been conducted using test sequences where shooting parameters are perfectly known and where vergence/accommodation conflict is not predominant. Perspective distortions will occur in some of the contents regarding viewing conditions. People were asked to qualify sequences in term of naturalness and visual comfort. Results of test revealed a clear correlation between perspective conflict and visual discomfort perceived. Whatever the shooting condition, parallel or toed-in cameras, results are similar. A factor between depth and perspective can be calculated from shooting configuration and viewing condition. It seems to be relevant to estimate the comfort of S3D content perception. Subjective tests have permitted to better understand the link between perspective conflicts and visual comfort. A specific factor has been defined taking into account shooting parameters and viewing conditions to predict comfort or discomfort. Next, studies will be conducted to extend these tests to cinema conditions were the range of viewing conditions is larger.

8288-55, Poster Session

Evaluation of quality of experience in interactive 3D visualization: methodology and results

S. Tourancheau, M. Sjöström, R. Olsson, Mid Sweden Univ. (Sweden); A. Persson, Ctr. for Medical Image Science and Visualization (Sweden); T. Ericson, Setred AB (Sweden)

Human factors are of high importance in 3D visualization, but subjective evaluation of 3D displays is not easy because of a high variability among users. This study aimed to evaluate and compare two different 3D visualization systems (a market stereoscopic display, and a state-of-the-art multi-view display) in terms of quality of experience (QoE), in the context of interactive visualization.

An adapted methodology has been designed in order to focus on 3D differences and to reduce the influence of all other factors. 46 subjects took part in the experiment. The influence of their personal background, prior experience, and physiological characteristics has been investigated. After solving different tasks on a synthetic 3D scene, they were asked to judge the quality of the experience, according to specific features.

Results showed that visual comfort was significantly better on the multi-view display. This study has permitted to identify some factors influencing QoE: external factors such as length of the experiment, as well as user-centered factors such as prior experience, and stereopsis threshold. This paper provides an analysis of these factors and suggestions for test methodologies of future studies.

8288-56, Poster Session

Objective view synthesis quality assessment

P. Conze, Technicolor S.A. (France) and Institut National des Sciences Appliquées de Rennes (France); R. Philippe, Technicolor S.A. (France); L. Morin, Institut National des Sciences Appliquées de Rennes (France)

View synthesis brings geometric distortions which are not handled efficiently by existing image quality assessment metrics. Despite the widespread of 3-D technology and notably 3D TeleVision (3DTV) and Free-viewpoints TeleVision (FTV), the field of view synthesis quality assessment has not yet been widely investigated and new quality metrics are required.

In this study, we propose a new full-reference objective quality assessment metric called View Synthesis Quality Assessment (VSQA) metric. Our method is dedicated to artifacts detection in synthesized view-points and aims to handle areas where disparity estimation may fail: thin objects, object borders, transparency, variations of illumination or color differences between left and right views, periodic objects...

The key feature of the proposed method is the use of three visibility maps which characterize complexity in terms of textures, diversity of gradient orientations and presence of high contrast. Moreover, the VSQA metric may be defined as an extension of any existing 2D image quality assessment metric. Experimental tests have shown that our method allows a relevant detection of view synthesis artifacts in a wide range of situations.

8288-57, Poster Session

Depth enhancement of S3D content and the psychological effects

M. Hirahara, S. Shiraishi, T. Kawai, Waseda Univ. (Japan)

In recent, stereoscopic 3D (S3D) imaging technologies are widely used to create content for movies, TV programs, games etc. Although difference between S3D and 2D content is depth sensation induced by binocular parallax, the relation between depth control and user experience is still unclear.

Purpose of this study is to examine the effectiveness of depth control,

such as expansion or reduction, shifting forward or backward of the range of maximum parallactic angles between cross and uncross direction (depth bracket), by conducting a subjective evaluation on user experience. From the results of the evaluation, following two points were suggested.

(1) Expansion / reduction of depth bracket affected preference to the S3D content

(2) The effects of expansion / reduction of depth bracket were more notable than shifting to cross / uncross direction.

8288-58, Poster Session

Perception of size and shape in stereoscopic 3D imagery

M. D. Smith, MDS Consulting (United States); B. T. Collar, Warner Bros. Entertainment Inc. (United States)

3D content creation in the entertainment industry has always been somewhat of a "black art." Stereographers utilize a mix of creative and technical methods to produce 3D content that meets the creative intent of the film maker. However while the natural laws of physics governing how 3D imagery is acquired, transformed, and ultimately perceived by a viewer is well known, in our experience few stereographers make use of it in the course of day-to-day 3D film making. Stereographers turn their knobs in an iterative fashion until the desired result is achieved - great quality 3D films. This paper posits that if the mathematical relationships governing how 3D works were more often utilized the resulting quality would be the same, but with less iteration, i.e. same quality with less time and cost.

8288-59, Poster Session

Quality of experience model for 3DTV

W. Chen, France Telecom R&D (France) and Polytech' Nantes (France); J. Fournier, France Telecom R&D (France); M. Barkowsky, P. Le Callet, Polytech' Nantes (France)

Abstract:

With the advantage of enhanced 3D sensation brought by 3DTV to the viewers, how to assess these additional values subjectively is still an open question. In [1], the author tired to model the viewing experience and naturalness in terms of image quality and depth, however, visual comfort is not considered. In our study [2], the results revealed that visual experience is not a linear function in terms of visual comfort and depth rendering.

In this paper, we aim to modeling more precisely the quality of experience for 3DTV in terms of image quality, depth rendering and also visual comfort.

Synthetic contents and nature contents are generated and captured; each part has at least 5 scenes. For each scene, we generate several test conditions with combination of Depth of Focus-DOF (0.1, 0.2 and 0.3 diopter), stereoscopic shape distortion[2] visual artifacts (video bitrates) and views asymmetry (rotation, focal length difference, and etc). The test methodology is based on the SAMVIQ including four different perceptual scales: visual experience, image quality, depth rendering and visual discomfort. The analysis of result will define the interaction between these technique parameters and the subjective scales. It will lead to a new subjective quality of experience model. Furthermore, categorization of the different technical parameters is proposed based on their influence towards different subjective scales. Finally, a preliminary objective model will be proposed.

In summary, this paper will propose a new model of quality of experience for 3DTV considering different subjective components. From this model, a preliminary objective model will be derived taking into account the categorizations of technique parameters.

[1] P. Seuntiëns, "Visual experience of 3D TV," Doctor Thesis, Eindhoven University of Technology, 2006.

[2] W. Chen, et al., "New stereoscopic video shooting rule based on stereoscopic distortion parameters and comfortable viewing zone," in Stereoscopic Displays and Applications XXII, San Francisco Airport, California, USA, 2011, pp. 78631O-13.

8288-60, Poster Session

New approach on calculating multi-view 3D crosstalk for auto-stereoscopic displays

S. Jung, K. Lee, J. Kang, S. Lee, K. Lim, LG Display (Korea, Republic of)

In this study, we suggest a new concept of 3D crosstalk for auto-stereoscopic displays and obtain 3D crosstalk values of several multi-view systems based on the suggested definition. First, we first measure the angular dependencies of the luminance for auto-stereoscopic displays under various test patterns corresponding to each view of a multi-view system and then calculate the 3D crosstalk based on our new definition with respect to the measured luminance profiles. Our new approach gives just a single 3D crosstalk value for single device without any ambiguity and shows similar order of values to the conventional stereoscopic displays. These results are compared with the conventional 3D crosstalk values of selected auto-stereoscopic displays such as 2-view, 4-view and 9-view system. From the result, we believe that this new approach is very useful for controlling 3D crosstalk values of the 3D displays manufacturing and benchmarking of the 3D performances among the various auto-stereoscopic displays.

8288-61, Poster Session

Towards adapting current 3DTV for an improved 3D experience

L. Blondé, D. Doyen, C. Thébault, Q. Huynh-Thu, Technicolor S.A. (France); D. Stoenescu, E. Daniel, J. de Bougrenet de la Tocnaye, TELECOM Bretagne (France); S. Bentahar, EyesTripleShut (France)

Recent upgrades of HDTV into 3DTV resulted in impairments in displaying stereo contents. One of the most critical flaws is probably crosstalk and the resultant ghosting effect impairing the 3D experience. The purpose of this study is to identify the primary source of crosstalk, throughout the full image generation and viewing chain, for a selection of 3D displays (LCD and PDP) combined with different active glasses technologies. Temporal measurements have been carried out on various display panels and shutter glasses technologies. For each technology, the crosstalk is a complex combination of several factors depending on display panels, shutter glasses and their synchronization, and ghost busting. The study tried to discriminate the main sources of crosstalk in each case, and to simulate the effect of various display panel or shutter glasses performance optimizations. Conclusions vary depending on the display technology. For LCD displays, light leakage reduction at the panel level appears the first cause of crosstalk, and, in a second step, optimization of the shutter glasses. For plasma displays the use of more efficient shutter glasses can reduce color distortion effects.

8288-62, Poster Session

Integral volumetric imaging with high-resolution and smooth-motion parallax

S. Sawada, H. Kakeya, Univ. of Tsukuba (Japan)

This paper proposes a high resolution integral imaging system using a lens array composed of non-uniform elemental lenses. One of the problems of integral imaging is the trade-off relationship between the resolution and the number of views. The only conventional way to solve this problem is to use a display panel with a finer pixel pitch. In the proposed method large display area is used to show a smaller and finer 3D image. To realize it, the elemental lenses should be smaller than the elemental lenses. To cope with the difference of sizes between the elemental images and the elemental lenses, the lens array is designed so that the optical centers of elemental lenses are located in the centers of elemental images, not in the centers of elemental lenses. In addition, new image rendering algorithm is developed so that undistorted 3D image can be presented with a non-uniform lens array. The proposed design of lens array can be applied to integral volumetric imaging, where display panels are layered to show volumetric images in

the scheme of integral imaging. Realization of smooth motion parallax is also realized by adding another layer of lens array composed of finer elemental lenses.

8288-63, Poster Session

Enlargement of viewing freedom of reduced-view SMV display

J. Nakamura, T. Takahashi, Y. Takaki, Tokyo Univ. of Agriculture and Technology (Japan)

In order to reduce the resolution of a flat-panel display used for the SMV display, we previously proposed a reduced-view SMV display generating viewing zones only around viewer's left and right eyes. We combined a head tracking system employing a camera with the reduced-view SMV display to enlarge the viewing freedom simply. However, the longitudinal viewing freedom was limited. In the present study, we propose a SMV display system which generates viewing zones corresponding to 3D positions of viewer's left and right eyes to enlarge both the horizontal and longitudinal viewing freedoms. An eye tracking system with a stereo camera is combined with the previously developed reduced-view SMV display. Pixels on the flat-panel display which are seen through each cylindrical lens by the left and right eyes are identified geometrically for all of the cylindrical lenses. Parallax images are generated corresponding to the 3D positions of both eyes. Two or more consecutive pixels including the identified pixels are used to display the parallax images to achieve the SMV display condition. The 3D resolution of the display was 256×192, and the screen size was 2.57 inches. The eye tracking system was composed of two USB cameras. The frame rate of the eye tracking and the image update was 7 Hz.

8288-64, Poster Session

Development of high-speed, phase-modulating spatial light modulators for 3D-holographic displays

H. Takagi, T. Goto, A. Baryshev, M. Inoue, Toyohashi Univ. of Technology (Japan)

We have demonstrated spatial light modulators (SLMs) with a new design-so-called phase-modulating SLMs (eMO-SLM). Working element of these SLMs comprised magneto- and electro-optical bilayer introduced into a Fabri-Perot microcavity. Responses from these cavities were analyzed theoretically and measured experimentally, when controlling applied electric fields.

Calculated prototype of eMO-SLMs had the structure of anti-reflection coating/SGGG substrate/(Ta2O5/SiO2)f/Bi:YIG/PLZT/(SiO2/Ta2O5) r, where the response of Bi:YIG/PLZT magneto/electro-optical bilayer was controlled by the electric field applied to PLZT. Calculated voltage-driven responses of the considered structures have demonstrated that phase change over 360 deg. at a voltage of 0-0.5 V can be achieved for the reflected light intensity of 99%. Since eMO-SLMs comprising Bi:YIG cannot provide good performance in the blue spectral range because of a large absorption in Bi:YIG, calculations for eMO-SLM with paramagnetic MO materials were done. It was shown that the phase modulation of 180 deg. at reflectivity of 90 % was achieved in the short-wavelength range (405 nm). Analysis showed that eMO-SLMs provided a large phase modulation, high optical efficiency and high switching speed for wavelengths from the full visible spectral range. The modulation was also observed experimentally.

8288-65, Poster Session

Color hologram generation using depth map of real objects with viewing-zone-angle expansion

K. Nomura, Tokyo Univ. of Science (Japan) and National Institute of Information and Communications Technology (Japan); R. Oi, T. Kurita, National Institute of Information and Communications Technology (Japan); T. Hamamoto, Tokyo Univ. of Science (Japan)

In this paper, we propose the hologram generation method applying a viewing-zone-angle expansion method using color and depth information of real objects. First, we parallel-project the objects using color and depth information and calculate the light propagation from each pixel. At this time, to expand horizontal viewing-zone triply, we calculate the propagated light which satisfies the max diffraction angle to horizontal three directions from each pixel. At the same time, to eliminate the conjugate image, we apply the half-zone-plate processing to our proposed method. Second, we made the optical reproduction experiment using a 4f optical system with the single sideband method to eliminate the undesirable light. Also, the 3D image reconstructed from a hologram was so clear without color aberration. In addition, when we observed it from the expanded viewing-zone, we could observe the corresponding 3D images from each viewing-point and we showed the effectiveness of our proposed method. Moreover, applied Fast-Fourier-Transform (FFT) to this viewing-zone-angle expansion method, the 3D image is reconstructed from the hologram using this method with no degradation. Finally, we also confirmed the processing time is nine times faster than previous.

8288-66, Poster Session

Stereoscopic surround displays using interference filters

S. Peikert, J. Gerhardt, Fraunhofer FIRST (Germany)

Stereoscopy has become a common technology in modern media. However planar stereoscopic screens suffer from a limited field of view. The usage of multiple projectors allows to extend the field of view and to setup displays that surround the viewers. In that way the viewer can be put amidst the medial content instead of playing it in front of him.

To achieve stereoscopy on surround displays interference filters have some advantages over other techniques. However these filters introduce strong color differences between the projectors, which may reveal that the display is compound by multiple projectors. This article presents methods for a computationally efficient correction of the colorimetric properties of multi-projector surround displays. This correction is based on automated measurements by multiple cameras and a spectrometer.

The described methods were validated by applying them to a stereoscopic dome display made up of 16 high definition projectors equipped with Infitec filters. On that display we achieved a significant improvement of the colorimetric properties compared to regular soft-edge blending. Our reference setup shows that the multi-projector approach combined with interference filters allows to build highly immersive stereoscopic surround displays fulfilling today's requirements on spatial resolution, frame rates and interaction latencies.

8288-67, Poster Session

Design of extended viewing zone at autostereoscopic 3D display based on diffusing optical element

M. Kim, Y. S. Hwang, E. Kim, Kwangwoon Univ. (Korea, Republic of)

In these days, 3D display industry changes currently a non-glasses type to a non-glasses type. This paper proposes that a viewing zone for multiple view which is designed for the 3D display using DOE (Diffusing Optical Element) on the panel shows larger than that of the parallax barrier method or lenticular method and a crosstalk can be improved. Through proposed method, it is shown to enable the expansion and adjustment of the area of viewing zone according to viewing distance. The DOE is made by recording angularly different multiple interference pattern according to Bragg diffraction using 532nm green laser. In summary, for lateral and depth direction viewing zone of autostereoscopic 3D display for a multiple view under the DOE

system can be freely designed according to a diffusing angle and a decline angle.

8288-68, Poster Session

Floating image device with auto-stereoscopic display and viewer-tracking technology

C. Chen, K. Tseng, C. Wang, C. Tsai, Industrial Technology Research Institute (Taiwan)

To realize a display system which can project the 3D image to a free space, we proposed a floating image device with auto-stereoscopic display and viewer tracking technology. The floating image device has 3 major parts: (1) an optical projection system, (2) a 2-view auto-stereoscopic display, and (3) a viewer tracking equipment. With these three parts, the viewer can watch 3D images floating in the air with both motion and binocular parallax from a wide view angle. The floating image is designed to be projected in a free space reachable by the hands of the viewer. Therefore, free-space touch and interaction technologies can be included to increase the functions of the system.

8288-69, Poster Session

A time-sequential, mutli-view, autostereoscopic display without resolution loss using a multi-directional backlight unit and an LCD panel

H. Kwon, H. Choi, Sejong Univ. (Korea, Republic of)

In this paper, a time-sequential multi-view autostereoscopic display without resolution loss is proposed. The system consists of an LCD panel and a multi-directional backlight which has several groups of point light sources (LEDs) and a lens array. The distance between the lens array and the point light source is set to be the focal length of the lens array so that the backlight tilts the direction of light rays to make the observers see different images in different viewpoints. By turning on and off the point light sources sequentially, the images on the LCD panel can be observed in different locations. Therefore, if the change the images on the LCD panel is synchronized with the operation of point light source array, the number of 3D viewpoints can be increased while the resolution of the 3D images is not reduced. The proposed method can be realized with the latest flat panel display technologies and it is expected to realize an 8-view autostereoscopic display without resolution loss. In addition, the 2D mode can be realized by simply turning on all point light sources in the backlight.

8288-70, Poster Session

Full-resolution autostereoscopic display with all-electronic tracking system

J. Gaudreau, PolarScreens, Inc. (Canada)

PolarScreens is developing a new 3D display technology capable of displaying full HD resolution in each eye without the need for glasses. The technology combines a full resolution polarized display, a vertical oversize Patterned Polarizer Film and a head tracking system. This concept can be applied to any full resolution polarized display like iZ3D, Perceiva, or coming up active retarder 3D display. The technology rely on a multi-pixels alternated pattern encoded in the stereo image to follow the head movement.

The end result is a full resolution autostereoscopic display with complete head movement freedom. There are no mechanical moving part (like lenticular) or extra active components (like slit LCD acting as parallel barrier) to follow user's eyes. The display has all 2D/3D information all the time so there is no need for full screen or windowed 2D/3D switchable apparatus.

8288-98, Poster Session

Design of crossed-mirror array to form floating 3D LED signs

H. Yamamoto, Univ. of Tokushima (Japan) and JST CREST (Japan); H. Bando, S. Suyama, Univ. of Tokushima (Japan)

We have designed a reflective optical device to form a floating image of an LED signage. The optical device is composed of crossed mirrors. Each grid in the crossed-mirror array acts as a dihedral corner cube and reflects incident light to the opposite side of the array with keeping the direction parallel to the array. When an LED sign is placed in front of the crossed-mirror array, a real image of the LED sign is formed at the plane symmetric position of the LED sign about the crossed mirror array. In order to implement three-dimensionally floating LED signage, the optical device are required wide (at least 45 deg.) field of angle. Because there are black areas between LED lamps, wide point spread function is suitable. Furthermore, the required size of the optical elements is larger than our body in future. Then, scalability of composition is needed. In this paper, we investigate the design issue on crossed mirrors. In order to provide wide point spread function, aperture ratio is 0.8 horizontally and vertically. Formation of floating image with a certain spread has been investigated with prototype devices. Furthermore, floating display of an LED sign has been demonstrated.

8288-71, Poster Session

Novel multi-view generation framework for 3D displays

K. Hwang, Y. Cho, H. Lee, D. Park, C. Kim, Samsung Advanced Institute of Technology (Korea, Republic of)

In this paper, we propose a novel multi-view generation framework that considers not only high-quality of each view but also spatio-temporal consistency of each synthesized multi-view. Rather than independently filling in the holes of the individual generated image, the proposed framework gathers the hole information in each synthesized multi-view image to reference viewpoint and constructs hole map and SVRL(Single View Reference Layer) at the reference viewpoint and restores the holes in the SVRL, which results in the spatio-temporally consistent view generation. The hole map is constructed using the depth information of the reference viewpoint and the ratio of input/output baseline length so that the holes in the SVRL can represent the holes in the other multi-view images. For temporally consistent hole-filling of the SVRL, the restoration of holes in the current SVRL is performed by propagating the pixel value of the previous SVRL. The remained hole-filling is performed by depth-based exemplar based inpainting method. The experimental results showed that the proposed method can generate high quality and spatio-temporal consistent multi-view images in various input/output environments. In addition, the proposed framework can decrease the complexity of hole-filling process by reducing the repetition of the hole filling.

8288-72, Poster Session

Calibration of multiview camera with parallel and decentered image sensors

M. Ali-Bey, S. Moughamir, N. Manamanni, Univ. de Reims Champagne-Ardenne (France)

This paper focuses on the calibration problem of a multi-view shooting system designed for the production of 3D content for auto-stereoscopic visualization. The considered multi-view camera is characterized by coplanar and decentered image sensors regarding to the corresponding optical axis. Based on the Faugéras and Toscani's calibration approach, a calibration method is herein proposed for the case of multi-view camera with parallel and decentered image sensors. At first, the geometrical model of the shooting system is recalled and some industrial prototypes with some shooting simulations are presented. Next, the development of the proposed calibration method

is detailed. Finally, some simulation results are presented before ending with some conclusions about this work.

8288-73, Poster Session

Inversion-free multiview subpixel rendering for natural 3D presentation

Y. Kim, G. Sung, J. Park, D. Nam, D. Park, Samsung Advanced Institute of Technology (Korea, Republic of)

The purpose of this paper is to remove the inversion problem for the wide viewing angle and increase the 3D resolution using the subpixel rendering algorithm. In this paper, an inversion-free subpixel rendering method that uses eye tracking in a multiview display is proposed. The multiview display causes an inversion problem when one eye of the user is focused on the main region and the other eye is focused on the side region. The proposed rendering algorithm uses one main lobe as well as two side lobe regions depending on the user's movement. This tracks the user's eye position, and the movement of the eye continuously changes the view image using the relation between the eye position and the main lobe and side lobe region. Also, to increase the 3D resolution without the color artifact, the subpixel rendering algorithm using subpixel area weighting is proposed. This method uses the superposition of the subpixel and the lens ray axes. The subpixel values are re-calculated with the ratio of the superposition. In the experiments, 36-view images were seen using active subpixel rendering with the eye tracking system in a 4-view 3D display.

8288-74, Poster Session

Wide-viewing angle three-dimensional display based on the ray reconstruction method using multiple micro-projectors

H. Takahashi, K. Hirooka, Osaka City Univ. (Japan); K. Yamada, Osaka Univ. (Japan)

We propose an omnidirectional 3D display system. This is a tool for communication around a 3D image between a small number of people. This 3D display system consists of multiple basic 3D display units. The basic unit consists of a micro-projector, a lenticular lens array sheet, and a cylindrical lens. In this basic unit, since a screen is not used, the light rays from a micro-projector pass through a lenticular lens array sheet and observed directly. Thus, the spatial density distribution of projected light rays is partial. To average the spatial density of projected light rays, we use a cylindrical lens. To increase the viewing angle, we aligned multiple basic units in a circle, and displayed 3D images at the center of a circle. To verify the effectiveness of the proposed 3D display, we constructed the prototype system. This prototype consists of 8 basic units. They are aligned 18-degree apart in a circle and the radius is 95 mm. The maximum size of displayed 3D images is 35 mm x 40 mm x 35 mm. The viewing angle of a 3D image is 124-degree. This paper describes the principle of proposed 3D display system, and also describes the experimental results.

8288-75, Poster Session

Multi-layered, see-through movie in diminished reality

Y. Uematsu, T. Hashimoto, T. Inoue, N. Shimizu, H. Saito, Keio Univ. (Japan)

This paper presents a 3D see-through movie for the auto-stereoscopic display. Our target scene is a baseball game. We create a catcher's viewpoint movie, which is usually impossible to be captured at the official game, based on Diminished Reality technique by view translation and obstacle removing with multiple cameras.

We locate three color cameras behind the catcher and umpire, and remove those obstacles to visualize the pitcher's appearance as if seeing the pitcher through the catcher and umpire. Since the pitcher is far enough away from the cameras, the pitcher's appearance at the center camera is recovered by planar translation of the left and right cameras. The regions of the obstacles are detected by using depth information. By overlaying the recovered pitcher onto the regions of obstacles, we can virtually see the catcher's viewpoint movie.

Our see-through movie is also applied to the auto-stereoscopic display. Each object is assigned to a layer which has each distance in 3D space. In contrast with seeing the 2D see-through movie by changing transparency of the obstacles, then, the background (ex. Pitcher) can be seen in front of the obstacles (ex. Catcher) by changing the 3D position of the layer of background. Therefore we can manipulate the position of real objects in 3D by using our system.

8288-76, Poster Session

An independent motion and disparity vector prediction method for multiview video coding

S. Ryu, J. Seo, Yonsei Univ. (Korea, Republic of); J. Y. Lee, Samsung Advanced Institute of Technology (Korea, Republic of); D. H. Kim, Yonsei Univ. (Korea, Republic of); H. Wey, Samsung Advanced Institute of Technology (Korea, Republic of); K. Sohn, Yonsei Univ. (Korea, Republic of)

The inter-view prediction is used in addition to the temporal prediction in order to exploit both the temporal and inter-view redundancies in Multiview video coding. Accordingly, the multiview video coding has two types of motion vectors that are called the temporal motion vector and the disparity vector, respectively. The disparity vector is generally uncorrelated with the temporal motion vector. However, they are used together to predict the motion vector regardless of their types, therefore an efficiency of the conventional predictive coding of multiview video coding is decreased. In order to increase the accuracy of the predicted motion vector, a new motion vector prediction method including virtual temporal motion vector and virtual disparity vector is proposed for both the multiview video and multiview video plus depth formats. The experimental results show that the proposed method can reduce the coding bitrates by 6.5% in average and 14.7% at maximum in terms of the Bjontegaard metric compared to the conventional method.

8288-77, Poster Session

Geometry compensation using depth and camera parameters for three-dimensional video coding

D. Kim, J. Seo, S. Ryu, Yonsei Univ. (Korea, Republic of); J. Y. Lee, H. Wey, Samsung Advanced Institute of Technology (Korea, Republic of); K. Sohn, Yonsei Univ. (Korea, Republic of)

One of the important issues for the next generation broadcasting system is how to compress a massive amount of three-dimension (3D) video efficiently. In this paper, a geometric compensation method is proposed to obtain higher coding efficiency using the information of the color videos, depth videos and camera parameters.

Considering the disocclusion properties, a hole-filling method for P view is also proposed to further increase the coding performance in high bit-rates.

The experimental results show that the proposed algorithm can increase PSNR over 0.22dB and 0.63dB for P and B pictures, respectively. Meanwhile, we can achieve up to 6.28% and 18.32% BD bit-rates gain for P and B pictures, respectively.

8288-78, Poster Session

Converting conventional stereo pairs to multi-view sequences using morphing

R. Olsson, V. K. Adhikarla, S. Schwarz, M. Sjöström, Mid Sweden Univ. (Sweden)

Conference 8288

Autostereoscopic multi view displays require multiple views of a scene to provide motion parallax. When an observer changes viewing angle different stereoscopic pairs are perceived. This allows new perspectives of the scene to be seen giving a more realistic 3D experience. However, capturing arbitrary number of views is at best cumbersome, and in some occasions impossible. Conventional stereo video (CSV) operates on two video signals captured using two cameras at two different perspectives. Generation and transmission of two views is more feasible than that of multiple views. It would be more efficient if multiple views required by an autostereoscopic display can be synthesized from these sparse set of views.

This paper addresses the conversion of stereoscopic video to multiview video using the video effect morphing. Different morphing algorithms are implemented and evaluated. Contrary to traditional conversion methods, these algorithms disregard the physical depth explicitly and instead generate intermediate views using sparse sets of correspondence features and image morphing. A novel morphing algorithm is also presented that uses scale invariant feature transform (SIFT) and segmentation to construct robust correspondence features and qualitative intermediate views. All algorithms are evaluated on a subjective and objective basis and the comparison results are presented.

8288-79, Poster Session

Multiview video and depth compression for free-view navigations

Y. Higuchi, M. P. Tehrani, T. Yendo, T. Fujii, M. Tanimoto, Nagoya Univ. (Japan)

In this paper, we discuss a multiview video and depth coding system for Free viewpoint TV. MPEG adhoc group is discussing a new coding system which assumes that input and output data consists of multiview video sequences captured by a camera array and depth information corresponding each view. After decoding the data, virtual view synthesis is used to generate novel synthesized views. For compression of multiview video sequences, Multiview Video Coding (MVC) is standardized by the Joint Video Team (JVT). However, MVC only takes the multiview video sequences, so depth information has to be transmitted by alternative coding algorithm in order to provide audiences with multiview video applications in low complexity. Additionally, we should adjust and select quantization parameters for each video and depth stream which maximize the quality of synthesized view, because transmittable volume is generally limited. In this paper, as a first step to propose an appropriate compression method for the new framework, we investigate the effect on virtual view synthesis quality given by difference of transmission rates between multiview video and depth sequences. Our experimental result show that bitrates of multiview depth stream has less effect on the view synthesis quality compared with the video stream.

8288-80, Poster Session

A layered inpainting method for virtual view synthesis

S. Kim, K. J. Lee, Seoul National Univ. (Korea, Republic of); I. D. Yun, Hankuk Univ. of Foreign Studies (Korea, Republic of); S. U. Lee, Seoul National Univ. (Korea, Republic of)

In this paper, we have presented a new approach in view synthesis based on the concept missing area being filled with the background parts. Indeed, it basically consisted of two part: layer separation and layer inpainting. In the layer separation, it extracts the background part from the images and become a big issue since layer inpainting inferred the result of layer separation. Then layer inpainting is implied to replace the occlusion region with the background layer. To find the most similar patch, priority is used. Therefore, it significantly improve the quality of inpainting in view synthesis. It, yet, should be noted that this research has only examined with image set from the Microsoft. It has clear boundary between object and background, so the separation is easy compared to other images. Therefore, better mechanism to separate object is also necessary in further research.

8288-81, Poster Session

Analysis on ray reconstruction characteristics of multi-view and integral imaging display

H. Kim, H. Kim, K. Jeong, J. Park, Chungbuk National Univ. (Korea, Republic of)

We performed a comparative analysis for multi-view autostereoscopic display and horizontal parallax only integral imaging display. The spatio-angular ray distribution reconstructed by two technologies is calculated and used as a metric to evaluate the three-dimensional image reconstruction quality. Based on the ray reconstruction characteristics, we also propose a method to convert a set of perspective images for multi-view displays to a set of orthographic images for integral imaging display.

8288-82, Poster Session

A content-based method for perceptually driven joint color/depth compression

E. Bosc, L. Morin, M. Pressigout, Institut National des Sciences Appliquées de Rennes (France)

Multi-view Video plus Depth (MVD) data refer to a set of conventional color video sequences and an associated set of depth video sequences, all acquired at slightly different viewpoints. This huge amount of data necessitates a reliable compression method. However, there is no standardized compression method for MVD sequences. H.264/MVC compression method, which was standardized for Multi-View-Video representation (MVV), has been the subject of many adaptations to MVD. However, it has been shown that MVC is not well adapted to encode multi-view depth data.

We propose a novel option as for compression of MVD data. Its main purpose is to preserve joint color/depth consistency. The originality of the proposed method relies on the use of the decoded color data as a prior for the associated depth compression. This is meant to ensure consistency in both types of data after decoding. Our strategy is motivated by previous studies of artifacts occurring in synthesized views: most annoying distortions are located around strong depth discontinuities and these distortions are due to misalignment of depth and color edges in decoded images. Thus the method is meant to preserve edges and to ensure consistent localization of color edges and depth edges.

To ensure compatibility, colored sequences are encoded with H.264. Depth maps compression is based on a 2D still image codec, namely LAR (Locally adapted Resolution). It consists in a quad-tree representation of the images. The quad-tree representation contributes in the preservation of edges in both color and depth data.

The adopted strategy is meant to be more perceptually driven than state-of-the-art methods. The proposed approach is compared to H.264 encoding of depth images. Objective metrics scores are similar with H.264 and with the proposed method, and visual quality of synthesized views is improved with the proposed approach.

8288-84, Poster Session

Efficient panoramic sampling of real-world environments for image-based stereoscopic telepresence

L. E. Gurrieri, E. Dubois, Univ. of Ottawa (Canada)

A key problem in telepresence systems is how to effectively emulate the subjective experience of 'being there' delivered by our visual system. A step towards visual realism can be obtained by using high-quality panoramic snapshots instead of computer generated models of the scene. Unarguably, an even better immersive experience can be achieved by introducing monoscopic depth cues or by using stereoscopic-panoramic images. The latter is a daunting task

considering the large number of panoramic sources required to produce a comfortable stereoscopic navigation. In this scenario, constraints such as the acquisition time, rendering complexity and storage capacity become important. These design limitations motivated the development of an alternative method to facilitate a large scale stereoscopic survey of a scene. Our approach is based on the sampling of a target real-word scene using an arrangement of panoramic clusters. The mosaicking of specific sections of the panoramic images within each cluster helps to generate one complete binocular panoramic pair per cluster. This novel omnistereoscopic technique enables the fast acquisition and generation of high-quality stereoscopic images, using off-the-shelf panoramic cameras, and without requiring additional depth information.

8288-85, Poster Session

Partial 2D-to-S3D conversion and the cognitive characteristics

Y. Koido, T. Kawai, Waseda Univ. (Japan)

2D to stereoscopic 3D (S3D) conversion methods, one of the approaches to create S3D content, are divided into automatic "on-line" and manual "off-line". Off-line conversion of 2D to S3D conversion has cost issues in spite of higher S3D image quality than on-line one. In addition, although off-line conversion can control parallax amount more flexible than stereo filming, 2D material images are converted corresponding to the monocular information in most cases. Therefore, the authors propose a new method of S3D representation using 2D to S3D conversion which gives uncrossed parallax to whole 2D images and crossed parallax to the only particular area. The authors conducted subjective and objective evaluations to examine the cognitive characteristics of partial 2D to S3D conversion. In this paper, the details of the proposed method and the evaluation results are described.

8288-86, Poster Session

Disparity-compensated view synthesis for S3D content correction

R. Philippe, C. Thébault, V. Drazic, P. Conze, Technicolor S.A. (France)

The production of stereoscopic 3D content is considerably increasing and experience in 2-view acquisition is in progress. High quality material to the audience is required but not always ensured, and correction of the stereo views may be required. This is done via disparity-compensated view synthesis. A robust method has been developed dealing with these acquisition problems that introduce discomfort (e.g hyperdivergence and hyperconvergence,...) as well as those ones that may disrupt the correction itself (vertical disparity, color difference between views,...). The method has three phases : a preprocessing in order to correct the stereo images and estimate features (e.g. disparity range,...) over the sequence. The second (main) phase proceeds then to disparity estimation and view synthesis. Dual disparity estimation based on robust block-matching, discontinuity-preserving filtering, confidence and occlusion handling has been developed. Accurate view synthesis is carried out through disparity compensation. Disparity assessment has been introduced in order to detect and quantify errors. A post-processing deals with these errors as a fallback mode. The effectiveness of our framework has been demonstrated with complex video sequences. Quality assessment has shown the high quality of our DIBR method as well as the significant advantage of introducing pre- and post-processing in the very critical cases.

8288-87, Poster Session

Virtual view interpolation at arbitrary view points for mixed-resolution 3D videos

S. Lee, S. Lee, H. Wey, D. Park, Samsung Advanced Institute of Technology (Korea, Republic of)

For full motion parallax 3D display, it is necessary to supply multiple views obtained from a series of different locations. However, it is impractical to deliver all of the required views because they will result in a huge size of bit stream. In the previous work, authors proposed a mixed resolution 3D video format composed of heterogeneous resolution video and depth information pairs, and also suggested a view synthesis algorithm for the mixed resolution videos. This paper reports more refined view interpolation method and improved results.

The mixed resolution image view synthesis algorithm is to exploit high frequency components from the high resolution image input. The proposed algorithm is composed of the processes of 1) view warping 2) high frequency component detection and weight computation, and 3) virtual pixel blending.

On the contrary of the conventional view synthesis approach utilizing the same resolution inputs, the proposed synthesis method employs an approach that uses the high frequency components from the high resolution images. The experimental results show that the proposed view synthesis algorithm can improve the PSNR as well as the perceptual image quality.

8288-88, Poster Session

Efficient dense blur map estimation for automatic 2D-to-3D conversion

L. Vosters, Technische Univ. Eindhoven (Netherlands) and Axon Digital Design (Netherlands); G. de Haan, Philips Research Nederland B.V. (Netherlands) and Technische Univ. Eindhoven (Netherlands)

Focus is an important depth cue for 2D-to-3D conversion of low-depth of field images and video. However, focus can be only reliably estimated on edges. Therefore, Bea et al. [1] first proposed an optimization based approach to propagate focus to non-edge image portions, for single image focus editing. While their approach produces accurate dense blur maps, the computational complexity and memory requirements for solving the resulting sparse linear system with standard multigrid or (multilevel) preconditioning techniques, are infeasible within the stringent requirements of the consumer electronics and broadcast industry. In this paper we propose fast efficient, low latency, line scanning based focus propagation, which mitigates the need for complex multigrid or (multilevel) preconditioning techniques. In addition we propose facial blur compensation to compensate for false shading edges that cause incorrect blur estimates in people's faces. In general shading leads to incorrect focus estimates, which may lead to unnatural 3D and visual discomfort. Since visual attention mostly tends to faces, our solution solves the most distracting errors. A subjective assessment by paired comparison on a set of challenging low-depth-of-field images, shows that the proposed approach achieves equal 3D image quality as optimization based approaches, and that facial blur compensation results in a significant improvement.

8288-89, Poster Session

ROI-based transmission method for stereoscopic video to maximize rendered 3D video quality

C. T. E. R. Hewage, M. G. Martini, Kingston Univ. (United Kingdom)

A technique to improve the rendering quality of novel views for colour plus depth based 3D video is proposed. Most depth discontinuities occur around the edges of depth map objects. If information around edges of both colour and depth map images are lost during transmission, this will affect the quality of the rendered views. Therefore this work proposes a technique to categorize edge and surrounding areas into two different regions (i.e., Region Of Interests (ROIs)) and later protect them separately to provide Unequal Error Protection during transmission. In this way the most important edge areas (vital for novel view rendering) will be protected ahead of other surrounding areas. This method is tested over a H.264/AVC based simulcast encoding and transmission setup. The results show improved rendered

quality with the proposed ROI base Unequal Error Protection (UEP) method compared to Equal Error Protection (EEP) method.

8288-90, Poster Session

Stereoscopic image-inpainting-based, view-synthesis algorithm for glasses-based and glasses-free 3D displays

C. Yuan, Sharp Labs. of America, Inc. (United States); C. H. H. Liao, The Univ. of Southern California (United States); H. Pan, Sharp Labs. of America, Inc. (United States)

Recent trends in 3D displays clearly show that the glasses-based 3D displays have become main stream, while the glasses-free 3D displays are emerging. The existing stereoscopic content, however, does not always provide the optimal viewing experiences on the 3D displays. As for glasses-based displays, the original stereoscopic 3D content may not have the desired strong depth or may cause visual discomfort due to excessive depth and distortion. New stereo images with tuned 3D depth need to be synthesized for viewers' preferences and comfort. As for the glasses-free displays, more than two views are often desired and therefore the original two-view content needs to be converted into multi-view format. A view synthesis algorithm is needed for generating high-quality stereoscopic content for both types of 3D displays. The goal of our work is to develop a novel view synthesis algorithm for generating high-quality stereoscopic 3D content and providing optimal viewing experiences on both glasses-based and glasses-free 3D displays. The proposed approach consists of the following steps: (1) Disparity (depth) scaling; (2) Stereo inpainting; (3) Removal of grid quantization artifact. We have applied the view synthesis algorithm to various real-life image sequences. The method is able to generate high-quality virtual views needed by depth tuning on glasses-based 3D displays (2-view content) and virtual composite 8-view imagery needed by glasses-free 3D displays.

8288-91, Poster Session

Towards a real-time, high-definition depth sensor with hardware-efficient stereo matching

K. Zhang, IMEC (Belgium) and Katholieke Univ. Leuven (Belgium); G. Yi, C. Liao, C. Lin, H. Yeh, IMEC (Belgium); R. Lauwereins, IMEC (Belgium) and Katholieke Univ. Leuven (Belgium); L. J. Van Gool, Katholieke Univ. Leuven (Belgium) and ETH Zurich (Switzerland); G. Lafruit, IMEC (Belgium)

In this paper, we present an efficient hardware design of stereo matching for depth extraction. It is capable of matching high-definition images at real-time speed, producing high-quality disparity maps. The stereo algorithm comprises three major components, i.e. census transform, dynamic programming and local-adaptive voting. The census transform makes our method robust to radiometric differences. The dynamic programming method propagates information along scanlines by imposing a smoothness criterion. The local-adaptive voting further improves the matching quality, utilizing the correlation between intensity images and disparity maps. The whole algorithm is prototyped on a FPGA platform, effectively harnessing the power of parallel computing. With a working frequency of 65MHz, our design produces high-quality 1024x768 disparity maps at 60fps. Thanks to the hardware-oriented optimizations, our design is cost-effective and suitable to a variety of embedded applications.

8288-92, Poster Session

Interactive floating windows: a new technique for stereoscopic video games

A. Hogue, B. Kapralos, C. Zerebecki, M. Tawadrous, B. Stanfield, Univ. of Ontario Institute of Technology (Canada)

The film industry has a long history of creating compelling experiences in stereoscopic 3D. Recently, the video game as an artistic medium has matured into an effective way to tell engaging and immersive stories. Given the current push to bring stereoscopic 3D technology into the consumer market there is considerable interest to develop stereoscopic 3D video games. Game developers have largely ignored the need to design their games specifically to use stereoscopic technology and have thus relied on automatic conversion technology. Game developers need to evaluate solutions used in other media, such as film, to correct perceptual problems such as window violations, and modify or create new solutions to work within an interactive framework. We extend the dynamic floating window technique into the interactive domain enabling the player to position a virtual window in space. Interactively changing the position, size, and the 3D rotation of the virtual window, objects can be made to 'break the mask' dramatically enhancing the stereoscopic effect. By demonstrating that solutions from the film industry can be extended into the interactive space, it is our hope that this initiates further discussion in the game development community to strengthen their story-telling mechanisms in stereoscopic 3D games.

8288-93, Poster Session

Stereoscopic reconfiguration for 3D displays

J. Houde, P. Jodoin, Univ. de Sherbrooke (Canada); F. Deschênes, Univ. du Québec à Rimouski (Canada)

In this paper, we present a method to reconfigure 3D movies in order to minimize distortion when seen on a different display than the one it has been configured for. By their very nature, 3D broadcasts come with a stereoscopic pair to be seen by the left and right eyes. However, according to reasons that we ought to explain in the paper, the cameras used to shoot a movie are calibrated according to specific viewing parameters such as the screen size, the viewing distance and the eye separation. As a consequence, a 3D broadcast seen on a different display (say a home theater or a PC screen) than the one it has been configured for (say an IMAX® screen) will suffer from noticeable distortions. In this paper, we describe the relationship between the size of the 3D display, the position of the observer, and the intrinsic and extrinsic parameters of the cameras. With this information, we propose a method to reorganize the stereoscopic pair in order to minimize distortion when seen on an arbitrary display. In addition to the raw video pair, our method uses the viewing distance, a rough estimate of the 3D scene, and some basic information on the 3D display. An inpainting technique is used to fill disoccluded areas.

8288-94, Poster Session

Single lens dual-aperture 3D imaging system: color modeling to reduce the difference between two viewpoints

S. Y. Bae, Jet Propulsion Lab. (United States) and The Univ. of Southern California (United States); A. Ream, Montana State Univ. (United States); R. Korniski, M. Shearn, Jet Propulsion Lab. (United States); H. Shahinian, Skull Base Institute (United States); H. Manohara, Jet Propulsion Lab. (United States)

In an effort to miniaturize a 3D imaging system, we created two viewpoints in a single objective lens camera. This works by placing a pair of Complementary Multi-band Bandpass Filters (CMBFs) in one half and the other half of the aperture area. Two key characteristics about the CMBFs are that (1) the passbands that are staggered or interdigitated, so only one viewpoint is opened at a time when a light band matched to the passband is illuminated, and (2) the passbands are positioned throughout the visible spectrum, so each viewpoint can render color by taking RGB spectral images. However, because the passbands are staggered, each CMBF does not only cover the full visible spectrum but also omits regions of the visible spectrum. Thus, each viewpoint takes a different spectral image from the other viewpoint hence yielding a different color image. It raises concern for mismatched color from the two viewpoints because it can lead to color rivalry, where viewers cannot resolve the two different colors into one in their brain. This can be resolved by using a Chromatic Adaptation

Transform (CAT), which maps the colors seen through the CMBFs to colors that would appear under D65 illumination. The mapping is more complete as the number of passbands increases. However, the cost and technique to fabricate such constrains the number of passbands. A simulation was put together to evaluate the color differences under CMBFs with dual-, triple-, and quadruple-band bandpasses. In this report, results of the color simulation will be explained in CIE Lab space.

8288-15, Poster Session

A stereoscopic movie player with real-time content adaptation to the display geometry

S. Duchene, INRIA Rhône-Alpes (France); M. Lambers, Univ. Siegen (Germany); F. Devernay, INRIA Rhône-Alpes (France)

3D shape perception in a stereoscopic movie depends on several depth cues, including stereopsis. For a given stereoscopic content, the stereopsis depth cue highly depends on the camera setup, as well as on the display size and distance. But such problems can be resolved by performing view synthesis to reduce these perceived depth distortions. Many media such as TV broadcasts , blu-ray discs, and video-on-demand are now available in 3D but when the same stereoscopic movie has to be displayed on a different display size and distance than what it was produced for, severe depth distortions or eye divergence may happen. For this reason, we propose a real time method able to adapt any stereoscopic movie to any display, based on shooting and viewing geometries, which can be easily controlled by the viewer.

8288-95, Poster Session

A simultaneous 2D/3D autostereo workstation

D. Chau, B. McGinnis, J. Talandis, J. Leigh, Univ. of Illinois at Chicago (United States); T. Peterka, A. Knoll, A. Sumer, M. Papka, J. Jellinek, Argonne National Lab. (United States)

The use of Immersive Virtual Reality Environments (IVREs) in science is impeded by steep learning curves, high setup cost, and lack of interoperable software. We lower these barriers in this work with a new version of Dynallax, a dynamic barrier autostereo workstation, and DVC, a software library for driving Dynallax and coupling it to other visualization applications. We evaluate our system by deploying it as the primary work environment of a materials scientist to support her daily work flow.

Dynallax is an autostereoscopic immersive workstation with head-tracking and other novel 3D interactions that allows scientists to seamlessly interact with full resolution 2D content such as web browsers alongside autostereo 3D content such as volume renderings and geometric models. DVC is the software library that enables scientists to easily render their data on Dynallax, even with existing visualization tools.

Our solution aided scientists from the Chemical Sciences and Engineering division at Argonne National Laboratory to render and model geometry for density functional theory computation for use in catalysis research. Our research demonstrated that combining an IVRE with a traditional desktop results in a hybrid 2D/3D workstation for scientific discovery.

8288-96, Poster Session

Interactive stereoscopic visualization of large-scale astrophysical simulations

R. Kaehler, SLAC National Accelerator Lab. (United States)

This presentation describes the experiences of rendering high quality HD stereoscopic animations from large-scale astrophysical simulations. In particular it presents the C++ software framework we developed and used for our interactive stereoscopic visualizations as well as the production of animations for world-wide planetarium shows like "Life: A cosmic story", California Academy of Sciences (narrated by Jodie Foster) and "The Big Bang", American Museum of National History (narrated by Liam Neeson) and "Journey to the Stars", American Museum of National History (narrated by Whoopi Goldberg).

We will give several detailed descriptions of application examples of visualizations from large-scale state-of-the-art simulations. In particular we will shows how the stereoscopic visualization helped the scientist to better understand and analyze their data. The examples include 3D stereoscopic animations of the large-scale structure formation simulation with the highest dynamic range that was ever carried out, simulations of the evolution of the first stars in the universe, the highest resolved galaxy merger simulation as well as an large-scale simulation of the cosmological re-ionization era.

8288-97, Poster Session

Designing stereoscopic information visualization for 3D-TV: what can we can learn from S3D gaming?

J. Schild, M. Masuch, Univ. Duisburg-Essen (Germany)

The paper explores graphical design and spatial alignment of visual information elements into stereoscopic content, e.g. captions, subtitles, and especially more complex elements in 3D-TV productions. The method used is a descriptive analysis of existing computer- and video games that have been adapted for stereoscopic display using semi-automatic rendering techniques (e.g.Nvidia 3D Vision) or games which have been specifically designed for stereoscopic vision. Digital games often feature compelling visual interfaces that combine high usability with creative visual design. We explore these game interfaces in stereoscopic vision regarding their stereoscopic characteristics, how they draw attention and where they fail. We present an overview structured according to different levels of content relation, ranging from information that does not refer at all to the scene, to visual elements which are closely related to a particular part of the scene. The resulting list of design considerations is used to discuss possible consequences, opportunities and challenges for integrating visual information elements into 3D-TV content. This work shall help to improve current post-production editing systems and identifies a need for future editing systems for 3D-TV, e.g. live editing and real-time alignment of visual information into 3D footage.

8288-38, Session 10

Case study: the introduction of stereoscopic games on the Sony PlayStation 3

I. H. Bickerstaff, Sony Computer Entertainment Europe Ltd. (United Kingdom)

A free stereoscopic firmware update on Sony Computer Entertainment's PlayStation 3 console provides the potential to increase enormously the popularity of stereoscopic 3D in the home. For this to succeed though, a large selection of content has to become available that exploits 3D in the best way possible.

In addition to the existing challenges found in creating 3D movies and television programmes, the stereography must compensate for the dynamic and unpredictable environments found in games. Automatically, the software must map the depth range of the scene into the display's comfort zone, while minimising depth compression.

This paper presents a range of techniques developed to solve this problem and the challenge of creating twice as many images as the 2D version without excessively compromising the frame rate or image quality.

Over fifty stereoscopic PlayStation 3 games have now been released and notable titles are used as examples to illustrate how the techniques have been adapted for different game genres.

Since the firmware's introduction in 2010, the industry has matured with a large number of developers now producing increasingly sophisticated

3D content. New technologies such as viewer head tracking and head-mounted displays should increase the appeal of 3D in the home still further.

8288-39, Session 10

Stereoscopic 3D video games and their effects on engagement

A. Hogue, B. Kapralos, C. Zerebecki, M. Tawadrous, B. Stanfield, Univ. of Ontario Institute Of Technology (Canada)

With the film industry embracing stereoscopic 3D, affordable 3D capable televisions are poised to enter the home market en masse. While many industry proponents have treated 3D movies as the primary market, consumers see video games as the "killer application" of 3D technology. This presents an opportunity for game developers to create even more interesting and immersive experiences by integrating stereoscopic 3D viewing modes into their game development workflow. While stereoscopic 3D (S3D) game development is in its early stages, it is poised to have significant impact on the consumer entertainment market. Although game developers understand the basics of S3D and its impact in terms of hardware performance, they are currently ill-equipped to understand the many factors that S3D has on the overall user experience. Currently, there is no standard mechanism that defines "user experience" in video games. However there are many contributing factors that one can investigate to determine how users respond. We describe results of a study to determine whether stereoscopic 3D viewing in a game setting has an effect on game engagement. A user-based experiment shows the effects stereoscopic 3D displays have on the user's level of immersion and flow within video games.

8288-40, Session 10

Stereoscopic display in a slot machine

M. Laakso, Finland's Slot Machine Association (RAY) (Finland)

This paper reports the results of a user trial with a slot machine equipped with a stereoscopic display. The main research question was to find out what kind of added value does S-3D bring to slot games? Work was started with a thorough literature survey, after which design and implementation of a novel gaming platform was started. RAY's multi-game slot machine "Nova" was converted to "3DNova" by replacing the monitor and converting six original games to S-3D format. Five 3DNova machines were put available into RAY arcades for six months. Both qualitative and quantitative analysis was carried out from statistical values, questionnaires and observations. According to the results, people find the S-3D concept interesting but the technology is not optimal yet. No statistical difference was found between Nova and 3DNova. Twentysomething males were excited, older people more cautious. Especially the glasses provide a challenge; ultimate system would probably use autostereoscopic technology. Also the games should be designed to utilize its full power. The main contributions are lessons learned from creating an S-3D slot machine platform, novel information about human factors related to S-3D gaming and analysis why certain type of games and effects suit better to S-3D gaming than others.

8288-42, Session 11

YouDash3D: exploring stereoscopic 3D gaming for 3D movie theaters

J. Schild, Univ. Duisburg-Essen (Germany); S. Seele, Hochschule Bonn-Rhein-Sieg (Germany); M. Masuch, Univ. Duisburg-Essen (Germany)

Along with the success of the digitally revived stereoscopic cinema, other events beyond 3D movies become attractive for movie theater operators, i.e. interactive 3D games. In this paper, we present a case that explores possible challenges and solutions for interactive 3D games to be played by a movie theater audience. We analyze the setting and showcase current issues related to lighting and interaction. Our second focus is to provide gameplay mechanics that make special use of stereoscopy, especially depth-based game design. Based on these results, we present YouDash3D, a game prototype that explores public stereoscopic gameplay in a reduced kiosk setup. It features live 3D HD video stream of a professional stereo camera rig rendered in a real-time game scene. We use the effect to place the stereoscopic effigies of players into the digital game. The game showcases how stereoscopic vision can provide for a novel depth-based game mechanic. Projected trigger zones and distributed clusters of the audience video allow for easy adaptation to larger audiences and 3D movie theater gaming.

8288-43, Session 11

3D storytelling: a case study

K. Lehto, 4th Wall Productions (Finland)

No abstract available

8288-44, Session 11

Thinking in z-space: flatness and spatial narrativity

R. Zone, The 3-D Zone (United States)

Now that digital technology has accessed the Z-space in cinema, narrative artistry is at a loss. Motion picture professionals no longer can readily resort to familiar tools. A new language and new linguistics for Z-axis storytelling are necessary. After first examining the roots of monocular thinking in painting, prior modes of visual narrative in two-dimensional cinema obviating the need for true binocular stereopsis can then be explored. Particular attention can be paid to montage, camera motion and depth of field through historic examples. Special consideration can be given to the manner in which monocular cues for depth have been exploited to infer depth on a planar screen. Both the artistic potential and visual limitations of actual stereoscopic depth as a filmmaking language must be interrogated. After an examination of the historic basis of monocular thinking in visual culture, a context for artistic exploration of the use of the z-axis as a heightened means of creating dramatic and emotional impact upon the viewer can be illustrated.

8288-45, Session 12

Floating three-dimensional display with 360-degree viewing angle

D. Miyazaki, K. Okoda, Y. Maeda, T. Mukai, Osaka City Univ. (Japan)

The aim of this research is to develop a full-parallax auto-stereoscopic display system, which can generate a floating three-dimensional (3-D) image viewable form the surrounding area. A 3-D display method based on the combination of integral imaging, 360-degree scanning with a rotating mirror, and imaging in the air with a concave mirror is proposed. A scanning system is composed of a hemisphere concave mirror and a mirror scanner, which is located around the center of the concave mirror. Integral imaging is implemented by a two-dimensional display device and a lens array to achieve full-parallax auto-stereoscopic imaging. By putting an image generated by the integral imaging system into the scanning system, a floating stereoscopic image can be generated around the center of the concave mirror. When the mirror scanner rotates and the image on the integral imaging system is switched according to the mirror angle, each directional image can be observed from each viewing angle. A digital micro-mirror device is used to switch directional images. The feasibility of the proposed method was examined by preliminary experiments. The abilities of generation of a floating full-parallax auto-stereoscopic image and a floating auto-stereoscopic image with 360-degree viewing angle

were demonstrated.

8288-46, Session 12

A novel 3D display system using combined integral imaging and Fresnel hologram

J. Chen, N. Collings, D. Chu, Univ. of Cambridge (United Kingdom)

A 3D display system which combines integral imaging and holography is proposed. It is capable of combining merits of these two techniques and delivering a better performance than that by any of them individually. It is predicted that the viewing angle would be improved in comparison with that of Fresnel hologram on a phase-only spatial light modulator and the image depth would be enhanced in comparison with that of conventional integral imaging. By analysing the properties of the images generated by this method, the feasibility of this approach is confirmed. The obtained experimental results demonstrate clearly the two predicted benefits. The work here not only shows the potential of this approach for future 3D image displays but also provides an insight into two important topics in 3D display: integral imaging and holography.

8288-47, Session 12

Analysis and management of geometric distortions on multi-view displays with only horizontal parallax

A. Said, B. Culbertson, Hewlett-Packard Labs. (United States)

Many light field (multi-view) displays are designed to support horizontal parallax only (HPO), since this represents a large reduction in complexity, compared to full parallax. It is commonly assumed that there are no important practical disadvantage on using HPO, except small losses in 3D perceptual quality. The reality is that, because they mix different perspective views in the horizontal and vertical directions, all HPO displays present some anamorphic distortions, which were first observed in holographic stereograms. These distortions depend of observer's position, and can only be eliminated in a pre-defined viewing distance. In this paper we extend previous work on the theoretical analysis of the problem to create tools that can be used by creators of multi-view 3D content to manage the problem, and keep the distortion within acceptable ranges for all objects in a 3D scene, and all expected viewing position. We present many examples of simulated views of HPO displays, which demonstrate how the distortions can affect visual appearance, and how they are managed.

8288-48, Session 12

Mixed-resolution view synthesis using non-local means refined image merging

T. Richter, M. Schöberl, J. Seiler, T. Tröger, A. Kaup, Friedrich-Alexander-Univ. Erlangen-Nürnberg (Germany)

Synthesizing novel views from original available camera perspectives is a key issue in the 3D video domain. Both, a left-side and a right-side reference view are warped into the desired intermediate camera position via depth maps. Afterwards, the two warped images are merged to synthesize the intermediate camera perspective. Due to different reasons like wrong depth values or disocclusion, there exist some areas inside the synthesized view for which the corresponding information can only be found in one reference view. If the two reference cameras have different spatial resolutions, the insertion of upsampled and hence blurred regions from the low-resolution view produces annoying visible artifacts in the merged image. We propose a refinement step, based on the non-local means (NLM) algorithm, to efficiently adapt those blurred regions to their surrounding and thus increase the local image sharpness. During this refinement step, every pixel inserted from the low-resolution view gets refined by comparing its neighborhood with the neighborhoods of pixels within a specific search area. Subsequently the pixel gets replaced by a weighted sum of the pixels within this search area. The experimental results show, that the NLM-refined image merging achieves a PSNR gain of up to 0.9 dB and also a better subjective image quality compared to an unrefined mixed-resolution setup.

8288-49, Session 13

Investigating the cross-compatibility of IR-controlled active shutter glasses

A. J. Woods, J. Helliwell, Curtin Univ. (Australia)

Active Shutter Glasses (also known as Liquid Crystal Shutter (LCS) 3D glasses or just Shutter Glasses) are a commonly used selection device used to view stereoscopic 3D content on time sequential stereoscopic displays.

Regrettably most of the IR controlled active shutter glasses released to date by various manufacturers have used a variety of different IR communication protocols which means that active shutter glasses from one manufacturer are generally not cross compatible with another manufacturer's emitter.

The reason for the lack of cross-compatibility between different makes of active shutter glasses mostly relates to differences between the actual IR communication protocol used for each brand of glasses.

We have characterised eleven different 3D Sync IR communications protocols in order to understand the possibility of cross-compatibility between different brands of glasses.

This paper contains a summary of the eleven different 3D Sync IR protocols as used by a selection of emitters and glasses.

The paper provides a discussion of the similarities and differences between the different protocols, the limitations for creating a common 3D Sync protocol, and the possibility of driving multiple brands of glasses at the same time.

8288-50, Session 13

New high-brightness interference filter developments

H. Jorke, A. Simon, Infitec GmbH (Germany)

A new interference filter design was developed as a three and four band pass design. This design offers well balanced colors at a very high transmission rate.

Glasses based stereoscopic technologies offer a cheap, robust and reliable approach to provide 3D content for large venues. Besides polarization and active shutter technologies, the wavelength multiplexing approach is an upcoming major technology branch.

The existing triple band interference filter system required an electronic color correction which also implied a reduction of brightness. The goal of this work was the improvement of brightness and colors performance for stereoscopic interference filter systems.

We have collected a dataset of different types of projector and lamp configurations. The variations of imagers (LCD, LCoS, 1-Chip DLP and 3-Chip DLP), of light sources (high pressure mercury and xenon based) and dichroic filters have been considered for a optimization for a new standard interference filter design. A multitude of filter designs using three, five or more band pass filters has been simulated. Results show that the 3-4 band filter design exhibits the highest transmission and while maintaining balanced colors.

8288-51, Session 13

Real-world stereoscopic performance in multiple-focal-planes displays: how far apart should the image planes be?

S. J. Watt, K. J. MacKenzie, L. C. Ryan, Bangor Univ. (United Kingdom)

Conventional stereoscopic displays present conflicting stimuli to vergence and accommodation, causing fatigue, discomfort, and poor stereo depth perception. One promising solution is 'depth filtering', in which continuous variations in focal distance are simulated by distributing image intensity across multiple focal planes. The required image-plane spacing is a critical parameter, because there are constraints on the total number that can be used. Depth-filtered images have been shown to support continuous and reasonably accurate accommodation responses with 1.1 diopter (D) image-plane spacings. However, retinal contrast is increasingly attenuated with increasing image-plane separation. Thus, while such stimuli may eliminate the vergence-accommodation conflict, they may also unacceptably degrade stereoscopic depth perception. Here we measured stereoacuity, and the time needed for stereoscopic fusion, for real targets and depth-filtered approximations to the same stimuli (image-plane spacings of 0.6, 0.9 and 1.2 D). Stereo fusion time was reasonably consistent across conditions. Stereoacuity for depth-filtered stimuli was only slightly poorer than for real targets with 0.6 D image-plane separation, but deteriorated rapidly thereafter. Our results suggest that stereoscopic depth perception, not accommodation and vergence responses, is the limiting factor in determining acceptable image-plane spacing for depth-filtered images. We suggest that image-plane spacing should not exceed ~0.6 D.

8288-52, Session 13

An eyeglass-like, eye-tracked, optical see-through, head-mounted display using freeform optics

H. Hua, College of Optical Sciences, The Univ. of Arizona (United States)

In this paper, we present an innovative optical approach to the design of an optical see-through eyetracked head-mounted display (ET-HMD) system based on freeform optical technology and an innovative optical scheme that uniquely combines the display optics with the eye imaging optics. Our approach allows creating a lightweight, compact, and robust ET-HMD solution with a non-obtrusive form factor. The optical design of the see-through HMD path is based on a freeform wedge-shaped prism cemented with a freeform lens. The wedge-shaped prism freeform prism, composed of three non-rotationally symmetric freeform surfaces, serves as a near-eye imaging system that magnifies the image displayed on a mircodisplay. To enable low-distortion see-through capability, the attached freeform lens provides correction to the distortion introduced by the freeform prism to the see-through view of a real-world scene. The dimensions for the cemented prism and lens are about 30mm by 30mm by 15mm with a weight of about 15grams. The same freeform prism along with only one additional optical element serves the eye imaging function for a video-based eyetracking system. A preliminary design of the described ET-HMD system will be presented.

8288-53, Session 13

Optimization of a multi-view system based on pulsed LED-LCD projectors

A. Pross, R. Blach, M. Bues, R. Reichel, O. Stefani, Fraunhofer-Institut für Arbeitswirtschaft und Organisation (Germany)

Recent advancement in stereoscopic displays and 3D projection technologies has been shown, driven by the current dissemination of 3D technologies at cinemas and consumer products. Building Virtual Reality systems of these products with head tracking and a perspective-correct view provide only a single user view. Other viewers share the same image from a different position. Providing perspective-correct stereoscopic views for multiple, individually tracked users on a shared surface are necessary. We present a brightness and crosstalk optimized multi-view stereo display based on pulsed LED light sources of a set of multiple LCD projectors. Pulsed LED projectors allow high frequency switching between the displayed images, almost eliminating a major source of crosstalk between the individual views. We evaluated the power consumption, projection brightness and crosstalk of our multi-view system and present a solution which alleviates the inherent problems of shutter-based multi-view systems, which are limited energy inefficiency, brightness degradation due to the projector shutter, and relatively high crosstalk, which moreover increases with the number of views.

Conference 8289 · The Engineering Reality of Virtual Reality 2012

Tuesday 24 January 2012

Part of Proceedings of SPIE Vol. 8289 The Engineering Reality of Virtual Reality 2012

8289-01, Session 1

Foreign language learning in immersive virtual environments

B. Chang, L. Sheldon, M. Si, Rensselaer Polytechnic Institute (United States)

Virtual reality has long been used for training simulations in fields from medicine to welding to vehicular operation, but simulations involving more complex cognitive skills present new design challenges. Foreign language learning, for example, is increasingly vital in the global economy, but computer-assisted education is still in its early stages. Immersive virtual reality is a promising avenue for language learning as a way of dyamically creating believable scenes for conversational training and role-play simulation. Visual immersion alone, however, only provides a starting point. We suggest that the addition of social interactions and motivated engagement through narrative gameplay can lead to truly effective language learning in virtual environments. In this paper, we describe the development of a novel application for teaching Mandarin using CAVE-like VR, physical props, human actors and intelligent virtual agents, all within a semester-long multiplayer mystery game. Students travel (virtually) to China on a class field trip, which soon becomes complicated with intrigue and mystery surrounding the lost manuscript of an early Chinese literary classic. Virtual reality environents such as the Forbidden City and a Beijing teahouse provide the setting for learning language, cultural traditions, and social customs, as well as the discovery of clues through conversation in Mandarin with characters in the game.

8289-02, Session 1

Virtual reality technology prevents accidents in extreme situations

Y. Badihi, Jerusalem College of Technology (Israel); M. Reiff, Jerusalem College of Technology (Korea, Republic of)

A Summary:

1. Technology goals & hypotheses:

a. Creating a training environment for acquisition of early cues in a dynamic "World Picture" for expected driving situations, especially extreme situations.

b. Creating training conditions that enable greater longevity of skills acquired.

2. Basic Findings:

a. The more "immersion" features are incorporated into VR extreme-situations training, the faster and the better the reactions of the trained driver.

b. The more "immersion" features are incorporated into VR extreme-situations training, the greater the longevity of skills acquired in training.

3. Methods & research tools:

a. Creation of a training facility using VR technology, based on a VR computerized model, incorporating interactive activity. The "world picture" display (driving behavior within the VR) is received through tangible methods.

b. Building varied levels of immersion: going from the basic simulation system, with a low level of immersion, to the richest simulation experience with deep immersion, including high-level audio features.

c. Building a scenario bank with varied levels of complexity. The scenarios represent various high workload and extreme situations which characterize the trainee's world picture. The subject sample includes approximately 70 randomly chosen drivers for the various training profiles.

4. Scientific and practical contributions:

a. This research focuses on instances where the driver needs to act quickly, absorb more information, and make an exact decision under pressure.

b. The training experience with VR, as expressed in "immersion", is a traumatic experience meant to leave a lasting impression.

8289-03, Session 1

Motion parallax in immersive cylindrical display systems

N. Filliard, G. Reymond, Renault Technocentre (France); A. Kemeny, Renault Technocentre (France) and Arts et Métiers ParisTech (France); A. Berthoz, Collège de France (France)

Motion parallax is a crucial visual cue produced by translations of the observer for the perception of depth and self-motion. Therefore, tracking the observer viewpoint has become inevitable in immersive virtual (VR) reality systems used e.g. in automotive industry or in scientific studies of visual perception.

The perception of a stable and rigid world requires that this visual cue be coherent with other extra-retinal (e.g. vestibular, kinesthetic) cues signaling ego-motion. Although world stability is never questioned in real world, rendering head coupled viewpoint in VR can lead to the perception of an illusory perception of unstable environments, unless a non-unity scale factor is applied on recorded head movements. Besides, cylindrical screens are usually used with static observers due to image distortions when rendering image for viewpoints different from a sweet spot.

We developed a technique to compensate in real-time these non-linear visual distortions, in an industrial VR setup, based on a cylindrical screen projection system.

Additionally, a "motion parallax gain" between the velocity of the observer's head and that of the virtual camera was introduced in this system. Results indicate that, below unity, gains significantly alter postural control. Conversely, the influence of higher gains remains limited, suggesting a certain tolerance of observers to these conditions.

8289-04, Session 2

Meta!Blast computer game: a pipeline from science to 3D art to education

W. Schneller, P. J. Campbell, E. S. Wurtele, D. Bassham, Iowa State Univ. (United States)

Meta!Blast (http://www.metablast.org) is designed to address the challenges students may encounter in understanding the spatial and time sensitive components of the mechanisms of bioenergy. Meta!Blast, developed by faculty and students in biology, biochemistry, computer science, game design, pedagogy, art and story, is being modeled/created in MAYA and a Unity game engine, for MACs/PCs in classrooms; it has also been exhibited in an immersive environment.

Here, we describe the pipeline from protein structural data and holographic information to art to the 3D environment to the game engine, by which we provide a publicly-available interactive three-dimensional cellular world that mimics a photosynthetic plant cell.

8289-05, Session 2

Learning immersion without getting wet

J. C. Aguilera, Adler Planetarium & Astronomy Museum (United States)

This paper describes the teaching of an immersive environments class on the Spring of 2011. The class had students from undergraduate as well as graduate art related majors. Their digital background and interests were also diverse. These variables were channeled as different approaches throughout the semester. Class components included fundamentals of stereoscopic computer graphics to explore spatial depth, 3D modeling and skeleton animation to explore presence, exposure to formats like a stereo projection wall and dome environments to compare field of view across devices, and finally, interaction and tracking to explore issues of embodiment. All these components were supported by theoretical readings discussed in class. Guests artists presented their work in Virtual Reality, Dome Environments and other immersive formats. Museum professionals also introduced students to space science visualizations which utilize immersive formats. We will present the assignments and their outcome, together with insights as to how the creation of immersive environments can be learned through constraints that expose students to situations of embodied cognition.

8289-06, Session 2

Byte me

D. Gromala, M. Song, A. Zeffiro, Simon Fraser Univ. (Canada)

Within the last few years there has been a feminist turn within HCI (Human-Computer Interaction). In 2010, Interacting with Computers issued a call for a special issue on feminism and HCI, a paper on feminism and HCI won Best Paper at CHI 2010 and, at CHI 2011, a feminist interactive design workshop was held. These events demonstrate the manner in which feminism continues to influence numerous disciplines, from human-robotic interaction (HRI) and CSCW (Computer-Supported Cooperative Work) to HCI and Social Media. Indeed, just as women were among the first directors and producers in the nascent film industry, women have and continue to produce within and around VR.

What then is a feminist perspective within VR? In addressing this question, we explore a diverse range of virtual environments and the assumptions underlying their design, examining and articulating factors that are all but invisible. We also examine the methodological and theoretical frameworks that undergird VR development and practice, and their practical consequences in the design and testing of virtual environments (VEs), especially those that ostensibly draw strict dichotomies between human and machine, and attribute quasi-autonomy to technological systems.

8289-07, Session 2

Female artists and the VR crucible: expanding the aesthetic vocabulary

J. F. Morie, The Univ. of Southern California (United States)

A survey done in 2007 of VR Artworks showed that women have created the majority of these immersive works. While this seems counter to popular ideas that the field is dominated by men, it seems rather, that the truly unique works appear to emerge from a feminine approach. Such an approach seems well suited to immersive environments as it incorporates aspects of inclusion, wholeness, and a blending of the body and the spirit. Female attention to holistic concerns fits the gestalt approach needed to create in a fully functional yet open-ended virtual world. It focuses not so much on producing a finished object (like a text or a sculpture) but rather on creating a possibility for becoming, like bringing a child into the world. Immersive VEs are not objective works of art to be hung on a wall and critiqued. They are vehicles for experience, vessels to live within for a piece of time.

8289-08, Session 3

Markerless 3D facial motion capture system

Y. Hwang, J. Kim, W. Bang, J. D. Kim, C. Kim, Samsung Advanced Institute of Technology (Korea, Republic of)

We propose a novel markerless 3D facial motion capture system using only one common camera. This system is simple and easy to transfer facial expressions of a user's into virtual world. It has robustly tracking facial feature points associated with head movements. In addition, it estimates high accurate 3D points' locations. We designed novel approaches to the followings; Firstly, for precisely 3D head motion tracking, we applied 3D constraints using a 3D face model on conventional 2D feature points tracking approach, called Active Appearance Model (AAM). Secondly, for dealing with various expressions of a user's, we designed 2D/3D face generic models from around 5000 images data and 3D shape data including symmetric and asymmetric facial expressions. Lastly, for accurately facial expression cloning, we invented a manifold space to successfully transfer 2D low dimensional feature points to 3D high dimensional points. The manifold space is defined by eleven facial expression bases. Consequently, our system has 4.8mm of average error on 5186 test images which are including various persons, various facial expressions associated with head movements, various illumination, etc.

8289-09, Session 3

Capturing geometry in real-time using a tracked Microsoft Kinect

D. Tenedorio, M. Fecho, J. Schwartzhaupt, R. Pardridge, J. Lue, J. P. Schulze, Univ. of California, San Diego (United States)

We investigate the suitability of the Microsoft Kinect device for capturing real-world objects and places. Our new geometry scanning system permits the user to obtain detailed triangle models of non-moving objects with a tracked Kinect. The system generates a texture map for the triangle mesh using video frames from the Kinect's color camera and displays a continually-updated preview of the textured model in real-time, allowing the user to re-scan the scene from any direction to fill holes or increase the texture resolution. We also present filtering methods to maintain a high-quality model of reasonable size by removing overlapping or low-precision range scans. Our approach works well in the presence of degenerate geometry or when closing loops about the scanned subject. We demonstrate the ability of our system to acquire 3D models at human scale with a prototype implementation in the StarCAVE, a virtual reality environment at the University of California, San Diego. We designed the capturing algorithm to support the scanning of large areas, provided that accurate tracking is available.

8289-10, Session 3

An example-based face relighting

H. Shim, Samsung Advanced Institute of Technology (Korea, Republic of)

We propose a new face relighting algorithm powered by a large database of face images captured under various known lighting conditions (i.g. a Multi-PIE database). The motivation of our algorithm is that a face can be represented by the assemble of patches from many other faces. The algorithm finds the most similar face patches in the database in terms of the lighting and the appearance. By assembling the matched patches, we can visualize the input face under various lighting conditions. Unlike existing face relighting algorithms, we neither use any kinds of face model nor make a physical assumption. Instead, our algorithm is a data-driven approach, synthesizing the appearance of the image patch using the appearance of the example patch. Using a data-driven approach, we can account for various intrinsic facial features including the non-Lambertian skin properties as well as the hair. Also, our algorithm is insensitive to the face misalignment. We demonstrate the performance of our algorithm by face relighting and face recognition experiments. Especially, the synthesized results show

that the proposed algorithm can successfully handle various intrinsic features of an input face. Also, from the face recognition experiment, we show that our method is comparable to the most recent face relighting work.

8289-11, Session 3

Heteronyms and avatars: a self-reflective system for artistic activity

E. E. Ayiter, Sabanci Univ. (Turkey)

This text will examine how avatars and the socially interactive, online virtual realities in which they become embodied may lead to an understanding regarding identity and self-perception, how such shifts in awareness may relate to the notion of the undividedly holistic 'self' and whether such perceptual shifts may be instrumental in bringing forth novel experiential states of artistic activity. The basis of this study is a self-observational social system, of which the agents are a coterie of avatars of both sexes, endowed with distinct physical attributes, both human and non-human; with uniquely emergent personalities which have progressed towards further idiosyncrasy over a period of three years. I, their creator am also the observer of their undertakings, their interactions, and their creative output, all of which manifest as disparate facets of my own persona and my artistic activity.

8289-12, Session 4

ManifestAR: an augmented reality manifesto

J. C. Freeman, Emerson College (United States)

ManifestAR is an international artists' collective working with emergent forms of augmented reality as interventionist public art. The group sees this medium as a way of transforming public space and institutions by installing virtual objects, which respond to and overlay the configuration of located physical meaning. This paper will describe the ManifestAR vision, which is outlined in the groups manifesto.

8289-13, Session 4

Invisible waves and hidden realms: augmented reality and experimental art

S. P. Ruzanka, Rensselaer Polytechnic Institute (United States)

Augmented reality is way of both altering the visible and revealing the invisible. It offers new opportunities for artistic exploration through virtual interventions in real space. In this paper, the author describes the implementation of two art installations using different AR technologies, one using optical marker tracking on mobile devices and one integrating stereoscopic projections into the physical environment. The first artwork, De Ondas y Abejas (The Waves and the Bees), is based on the widely publicized (but unproven) hypothesis of a link between cellphone radiation and the phenomenon of bee colony collapse disorder. Using an Android tablet, viewers search out small fiducial markers in the shape of electromagnetic waves hidden throughout the gallery, which reveal swarms of bees scattered on the floor. The piece also creates a generative soundscape based on electromagnetic fields. The second artwork, Urban Fauna, is a series of animations in which features of the urban landscape become plants and animals. Surveillance cameras become flocks of birds while miniature cellphone towers, lamp posts, and telephone poles grow like small seedlings in time-lapse animation. The animations are presented as small stereoscopic projections, integrated into the physical space of the gallery. These two pieces explore the relationship between nature and technology through the visualization of invisible forces and hidden alternate realities.

8289-14, Session 4

Immersive realities: articulating the shift from VR to AR through artistic practice

T. Cornish, T. Margolis, Univ. of California, San Diego (United States)

Our contemporary imaginings of technological engagement with space has transitioned from flying through Virtual Reality environments to mobile interactions with the physical world through personal media devices. Social media experiences within physical environments are now being preferenced over isolated environments such as CAVEs or HMDs. Examples of this trend can be seen in early tele-collaborative artworks which strove to use advanced networking to join multiple participants in shared virtual environments. Game consoles and powerful desktop graphics brought highly interactive content into our home while sacrificing visual and auditory immersion. Recent developments in mobile AR allow untethered access to such shared realities in places far removed from labs and home entertainment environments, and without the bulky and expensive technologies attached to our bodies that accompany most VR. This paper addresses the emerging trend favoring socially immersive artworks via mobile Augmented Reality rather than sensorially immersive Virtual Reality installations.

8289-15, Session 4

MetaTree: augmented reality narrative explorations of urban forests

R. G. West, T. Margolis, Univ. of California, San Diego (United States); J. O'Neil-Dunne, The Univ. of Vermont (United States); E. Mendelowitz, Smith College (United States)

As major cities world-wide adopt and implement reforestation initiatives to plant millions of trees in urban areas, they are engaging in what is essentially a massive ecological and social experiment. Existing air-borne, space-borne and field-based imaging and analysis mechanisms fail to provide key information on urban tree ecology crucial to informing management, policy, and supporting citizen initiatives for planting and stewardship of trees. Due to lack of spatio-temporal resolution, poor vantage point, cost constraints and biological metric limitations, they are ineffective as real-time inventory and monitoring tools. Novel methods for imaging and monitoring the status of these emerging urban forests and encouraging their ongoing stewardship by the public are required to ensure their success. Our art-science collaboration proposes to re-envision our relationship with urban spaces by foregrounding urban trees in relation to local architectural features and simultaneously creating new methods for urban forest monitoring. We explore creating a shift from overhead imaging or field-based tree survey data acquisition methods to continuous, ongoing monitoring as part of a mobile augmented reality experience. We consider the possibilities of this experience as a medium for interacting with and visualizing urban forestry data, and for creating cultural engagement with urban ecology.

8289-23, Poster Session

Prediction of visually perceived location using reaching action and effect of reaching distance on it

M. Suzuki, K. Takazawa, K. Uehira, Kanagawa Institute of Technology (Japan)

We examine the effect of the reaching distance on the prediction of the visually perceived location using the reaching action. For direct interaction between an observer's body and a virtual object, the system presenting the virtual object must execute the process of the interaction when the body is just on the visually perceived location of the virtual object. In contemporary techniques, it is assumed that the visually perceived location is the same as the location defined by

binocular disparity. However, both locations are often different. In our previous studies, we proposed a new technique for the prediction of the visually perceived location using an observer's action. Also, we demonstrated the prediction using the action that an observer reaches out for a virtual object. This study is an examination on the applicable range of our technique. In the experiment, an observer reaches out for a virtual object, and the reaching distance is the experimental variable. The results do not support the effect of the reaching distance on the prediction. We demonstrate that our technique is applicable to a wide range of reaching distance.

8289-24, Poster Session

An innovative multimodal virtual platform for communication with devices in a natural way

C. R. Kinkar, R. Golash, A. R. Upadhyay, Sagar Institute of Research, Technology & Science (India)

As technology grows people are diverted and are more interested in communicating with machine or computer naturally .This will make machine more compact and portable by avoiding remote, keyboard etc, also it will help them to live in an environment free from electromagnetic waves .This thought has made 'Recognition of natural modality in human computer interaction' a most appealing and promising research field. Simultaneously it has been observed that using single mode of interaction limit the complete utilization of commands as well as data flow. In this paper a multimodal platform, where out of many natural modalities like eye gaze, speech ,voice , face etc. human gestures are combined with human voice is proposed which will minimize the mean square error .This will loosen the strict environment needed for accurate and robust interaction while using single mode. Gesture complement Speech, gestures are ideal for direct object manipulation and natural language is used for descriptive tasks. Human computer interaction basically requires two broad sections recognition and interpretation. Recognition and interpretation of natural modality in complex binary instruction is a tough task as it integrate real world to virtual environment .The main idea of the paper is to develop a efficient model for data fusion coming from heterogeneous sensors, camera and microphone. Through this paper we have analyzed that the efficiency is increased if heterogeneous data (image & voice) is combined at feature level using artificial intelligence. The long term goal of this paper is to design a robust system for physically not able or having less technical knowledge

8289-25, Poster Session

Composite lay-up process with application of elements of augmented reality

J. Novak-Marcincin, J. Barna, M. Janak, L. Novakova-Marcincinova, V. Fecova, Technical Univ. of Kosice (Slovakia)

Article investigates the application possibilities of elements of augmented reality in area of composite lay-up technology with focus to utilization of principles of open source philosophy. It describes the software and hardware tools and aspects of development of environment consisting of combination of real and virtual items. In detail it describes the process of exploitation and creation working environment with use of augmented reality. All the theoretical aspects are explained and demonstrated on real example of composite lay-up process with use of special device for positioning of working table and total controlling of the process. Connections and relations are also described between hardware and software elements of the solution (sensors, logical loops, scripting). Described application provides its user with visual demonstration and guidance of entire lay-up process by displaying the final positions and boundaries of individual composite layers. Final part of the article informs about application possibilities of new solution and approaches further improvements and updates of proposed application.

8289-16, Session 5

Reordering virtual reality: recording and recreating real-time experiences

M. Dolinsky, W. R. Sherman, C. Eller, E. A. Wernert, Y. Chi, Indiana Univ. (United States)

With the proliferation of technological devices and artistic strategies, there is an urgent and justifiable need to capture site-specific time-based virtual reality experiences. These art events are dependent on the orchestration of multiple inputs including hardware, software, site specific location, user inputs and 3D stereo and sensory interactions. Although a photograph or video may illustrate a particular component of the work, such as an illustration of the artwork or a sample of the sound, they only represent a fraction of the overall experience. This paper seeks to discuss documentation strategies that combines multiple approaches and captures the interactions between art projection, acting, stage design, sight movement, dialogue and audio design.

8289-18, Session 5

The cognitive implications of virtual locomotion with a restricted field of view

W. E. Marsh, J. W. Kelly, V. J. Dark, J. H. Oliver, Iowa State Univ. (United States)

Virtual reality (VR) systems do not support fully natural locomotion through an infinite virtual environment. While this is partly due to constrained physical boundaries and tracking ranges, VR systems also provide low-fidelity sensory feedback, particularly visual. Visual feedback, such as optic flow, is used naturally to guide locomotion through the real world. In the presence of restricted visual feedback, users may need to employ cognitively demanding strategies to guide locomotion. This may lead to competition with other concurrent tasks that a user wishes to perform in the environment.

Results showed that users had more trouble remembering spatial or verbal items when concurrently performing locomotion with a low field-of-view (FOV) than with a high FOV. This memory decrement may indicate that performing locomotion movements with a restricted FOV requires more general cognitive resources than are needed with a high FOV. Movement start and completion times were also measured in order to examine resource requirements of specific aspects of individual movements. The movement data suggest that the starting/planning phase of movement requires more general cognitive resources when there is a restricted FOV. Understanding the specific performance problems that result from concurrent task performance may provide insight for the design of future systems.

8289-19, Session 6

Use of a simulator based on virtual reality to assess the application of economic driving techniques by truck drivers

T. N. Bogoni, Univ. Estadual do Mato Grosso Sul (Brazil); M. S. Pinho, PUCRS-IPCT (Brazil)

This paper presents the project and the development of a prototype for a truck simulator aimed at assessing the use of Economic Driving Techniques. We describe the techniques for economic driving and the way they are monitored, as well as the process of modeling and creating a virtual environment and the interaction devices used in the simulator. The simulation is accomplished by using a virtual desktop environment with hardware configuration similar to that in a real truck, in order to provide the user with a higher level of immersion. By testing drivers and economic driving instructors, it was possible to observe that the prototype can be used as a tool for assessing drivers and that the system is able to perceive a great part of the violations in the use of Economic Driving Techniques during the simulation similarly to a human expert.

8289-20, Session 6

LVC interaction within a mixed-reality training system

B. B. Pollock, Iowa State Univ. (United States) and Virtual Reality Application Ctr. (United States); E. Winer, S. Gilbert, Iowa State Univ. (United States); J. de la Cruz, U.S. Army RDECOM/STTC (United States)

The United States military is increasingly pursuing advanced live, virtual, and constructive (LVC) training systems for reduced cost, greater training flexibility, and decreased training times. Combining the advantages of realistic training environments and virtual worlds, mixed reality LVC training systems can enable live and virtual trainee interaction as if co-located. This paper discusses a system that overcomes tracking, latency, and hardware challenges to empower LVC interaction in a reconfigurable, mixed reality environment.

Iowa State University (ISU) developed electronics augment real-time tracking data of physical trainees, props, and vehicles to update virtual avatars. Virtual Battlespace 2 (VBS2) is utilized as a scenario authoring game engine, providing a first-person-shooter interface for virtual trainees and behaviors for constructive entities. Trainees additionally interact with the system through a high-resolution six sided CAVE commanding semi-autonomous unmanned aerial vehicles and through tracked teleoperated vehicles at other physical locations at ISU.

This system was developed and tested in an immersive, reconfigurable, and mixed reality LVC training system for the dismounted warfighter at ISU known as the Veldt. The Veldt achieves rare LVC interaction within multiple physical and virtual immersive environments for training in real-time across many distributed systems.

8289-21, Session 6

Immersive volume rendering of blood vessels

G. Long, H. S. Kim, A. Marsden, Y. Bazilevs, J. Schulze, Univ. of California, San Diego (United States)

In this paper, we present a novel method of visualizing flow in blood vessels. Our approach reads unstructured tetrahedral data, resamples it, and uses slice based 3D texture volume rendering. Due to the sparse structure of blood vessels, we utilize an octree to efficiently store the resampled data by discarding empty regions of the volume. We use animation to convey time series data, wireframe surface to give structure, and utilize the StarCAVE, a 3D virtual reality environment, to add a fully immersive element to the visualization.

Our tool has great value in interdisciplinary work, helping scientists collaborate with clinicians, by improving the understanding of blood flow simulations. Full immersion in the flow field allows for a more intuitive understanding of the flow phenomena, and can be a great help to medical experts for treatment planning.

8289-22, Session 6

A virtual reality interface for pre-planning of surgical operations based on a customized model of the patient

M. Witkowski, R. Sitnik, J. Lenar, Warsaw Univ. of Technology (Poland)

We present a human-computer interface that enables the operator to perform a virtual surgical procedure on the musculo-skeletal (M-S) model of the patient's lower limbs, send the modified model to the bio-mechanical analysis module, and export the scenario's parameters to the surgical navigation system. The interface provides the operator with tools for: bone cutting, manipulating and removal, repositioning muscle insertion points and placing implants.

The operator may export surgery parameters for bio-mechanical analysis of its functional outcome. If the result is satisfactory the scenario data may be used during the actual surgery.

The hardware configurations proposed to be used with the interface are: (a) a standard computer keyboard and mouse, and a 2-D display, (b) a touch screen as a single device for both input and output, or (c) a 3-D display and a haptic device for natural manipulation of 3-D objects.

The interface may be utilized in two main fields. Experienced surgeons may use it to simulate their operative plans and prepare input data for a surgical navigation system while student or novice surgeons can use it for training.

Conference 8290 · 3D Image Processing (3DIP) and Applications 2012

Tuesday-Thursday 24-26 January 2012

Part of Proceedings of SPIE Vol. 8290 Three-Dimensional Image Processing (3DIP) and Applications II

8290-01, Session 1

Edge-aware stereo matching with O(1) complexity

C. Cigla, Middle East Technical Univ. (Turkey) and Vestek (Turkey); A. A. Alatan, Middle East Technical Univ. (Turkey)

A computationally efficient stereo matching algorithm is introduced providing high precision dense disparity maps via local aggregation approach. The proposed algorithm exploits a novel paradigm, namely separable successive weighted summation (SWS) among horizontal and vertical directions with constant operational complexity, providing effective connected 2D support regions based on local color similarities. The intensity adaptive aggregation enables crisp disparity maps which preserve object boundaries and depth discontinuities. The same procedure is also utilized to diffuse information through overlapped pixels during occlusion handling. According to the experimental results on Middlebury online stereo benchmark, the proposed method is one of the most effective local stereo algorithm providing high quality disparity models by unifying constant time filtering and weighted aggregation. Hence, the proposed algorithm provides a competitive alternative for various local methods in terms of achieving precise disparity maps from stereo video within fast execution time.

8290-02, Session 1

Establishing eye contact for home video communication using stereo analysis and free viewpoint synthesis

C. Weigel, Technische Univ. Ilmenau (Germany); N. Treutner, Humboldt-Univ. zu Berlin (Germany)

Eye contact has been proven to be an important cue for video communication. We present a method to re-establish eye contact in a home video communication scenario caused by misalignment between the camera and the communication window the participant looks at. Our method covers the complete algorithm chain from acquisition to rendering and uses a pixel-based 3D analysis and rendering approach to create a virtual view of a camera placed at the position of the communication window. The outcomes of a large scale subjective study identified the crucial problems of such an approach. Based on the significant observations during the study in this paper we address the most important problems. We propose a method that produces spatially and temporally consistent depth maps using cross-check based filling and motion compensated filtering. We address aliasing artifacts during point rendering and present a method to f enhance the virtual view by image inpainting based on robust contour warping.

8290-03, Session 1

Depth adaptive hierarchical hole filling for DIBR-based 3D videos

M. Solh, G. Al-Regib, Georgia Institute of Technology (United States)

In this paper we introduce a depth adaptive approach for disocclusion removal in depth image-based rendering (DIBR). This approach extends the hierarchical hole-filling (HHF) presented in an earlier work. Similar to HHF, the depth adaptive approach results in synthesized 3D videos that are free of geometric distortions. Furthermore, the edges and texture around the disoccluded areas can be sharpened and enhanced through adding the depth adaptive pre-processing step before applying the hierarchical hole-filling. The subjective and objective results show a significant improvement in quality for the synthesized views using the depth adaptive approach.

8290-04, Session 1

Space carving MVD sequences for modelling natural 3D scenes

Y. Alj, Technicolor S.A. (France) and Institut National des Sciences Appliquées de Rennes (France); G. Boisson, Technicolor S.A. (France); L. Morin, M. Pressigout, Institut National des Sciences Appliquées de Rennes (France); P. Bordes, Technicolor S.A. (France)

In a classical multi-view video framework, from video acquisition to view synthesis, removing redundancy remains a challenging task. We aim to build a compact representation of the multiview-video material to be transmitted on a communication channel with high fidelity at decoder side. To this end, because of the efficiency of the rendering engines towards geometric primitives, we focus our effort on geometric-based representations of MVD sequences. More precisely, we generate a high resolution triangular mesh per view and propose to merge the generated meshes into a single volumetric framework. Hereby we take advantage of the exact per-view triangular mesh construction and the robustness of the volumetric representation. Our volumetric framework uses a variant of the space carving algorithm by introducing a new criterion for classify voxels into opaque or transparent. The merged space-carved volumetric model is turned back into a final mesh model with the marching cubes algorithm. Last that final mesh is simplified in order to fit realistic broadcast constraints, dropping vertices and edges that do not impact too much final rendered views quality. It turns out that our new formulation based on geometric consistency gives more accurate results than the traditional photo-consistency-based space carving.

8290-05, Session 3

A locally content-dependent filter for inter-perspective anti-aliasing

M. Sjöström, S. Tourancheau, X. Wang, R. Olsson, Mid Sweden Univ. (Sweden)

Presentations of contents on multiview and lightfield displays projecting a number of views into a viewing zone have become increasingly popular. The restricted number of views implies an unsmooth motion between views if objects with sharp edges are far from the display plane. The phenomenon is explained by inter-perspective aliasing. This is undesirable in applications where a correct perception of the scene is required, such as science and medicine, where depth compression also should be avoided. Anti-aliasing filters have been proposed in the literature, and are defined according to the minimum and maximum depth present in the scene. We suggest a method that subdivides the ray-space and adjusts the anti-aliasing filter to the scene contents locally. We further propose new filter kernels based on the ray space frequency domain that assures no aliasing, yet keeping maximum information unaltered. The proposed method outperforms filters of earlier works. Different filter kernels are compared. Details of the output are sharper using a proposed filter kernel, which also preserves the most information.

8290-06, Session 3

Photometric and geometric rectification for stereoscopic images

S. Han, J. Min, T. Park, Y. Kim, Samsung Electronics Co., Ltd. (Korea, Republic of)

Stereo image pairs are captured by two cameras at different positions. In general these two images have distortions such as vertical misalignment, rotation, and keystone. Even in case of a parallel stereo setup, the capturing system produces the distor-tions that cause uncomfortable 3D experiences to users. In this paper, we propose an algorithm to adjust the image pairs to give better stereoscopic experience to users by correcting the distortions. Our algorithm compensates photometrical and ge-ometrical distortions in image pairs efficiently. Moreover, we care the optimal horizontal parallax so that algorithm maximizes 3-D experiences based on user's preference.

8290-40, Poster Session

Novel time- and depth-stamped imaging for 3D-PIV (particle image velocimetry) using correlation image sensor

K. Komiya, T. Kurihara, S. Ando, The Univ. of Tokyo (Japan)

In this paper, we deal with a novel principle of three-dimensional particle image velocimetry(3D-PIV) of fluid motion using correlation image sensor (CIS) and a structured illumination. The correlation image sensor is the device which outputs the temporal correlation between incident light intensity and reference signal. CIS captures the loci of tracer particles with the passing time information in the form of phase distribution relating to a reference sinusoidal signal. Using temporal-spatial modulated light, it also records depth-directional component of the particle velocity. In other words, we can obtain the 3D shape of particle trajectory in one frame with monocular camera. Therefore, the proposed 3D-PIV system enables us to acquire the position and time information of numbers of tracers at the same time. The principle and implementation method in detail and several results of basic experiments are shown.

8290-41, Poster Session

3D imaging for ballistics analysis using chromatic white-light sensor

A. Makrushin, M. Hildebrandt, J. Dittmann, E. Clausing, Otto-von-Guericke-Univ. Magdeburg (Germany); R. Fischer, C. Vielhauer, Univ. of Applied Sciences Brandenburg (Germany)

The novel application of sensing technology, based on chromatic white light (CWL), gives a new insight into ballistic analysis of cartridge cases. The CWL sensor uses a beam of white light to acquire highly detailed topography and 2D-grayscale data simultaneously. The proposed 3D imaging system combines advantages of 3D and 2D image processing algorithms in order to automate the extraction of firearm specific toolmarks shaped on fired specimens. The most important characteristics of a fired cartridge case are the type of the breech-block marking as well as size, shape and location of extractor, ejector and firing pin marks. The feature extraction algorithm normalizes the casing surface and consistently searches for the appropriate distortions on the rim and on the primer. The location of the firing pin marking in relation to the lateral scratches on the rim provides unique rotation invariant characteristics of the firearm mechanisms. Additional characteristics are the volume and shape of the firing pin marking. The experimental evaluation relies on the data set of 15 cartridge cases fired from three 9mm firearms of different manufactures. The results show very high potential of 3D imaging systems for casing-based computer-aided firearm identification, which is prospectively going to support human expertise.

8290-42, Poster Session

Computer-aided 3D-shape construction of hearts from CT images for rapid prototyping

Y. Kato, M. Fukuzawa, N. Nakamori, Kyoto Institute of Technology (Japan); S. Ozawa, Kyoto Prefectural Univ. of Medicine (Japan); I. Shiraishi, National Cardiovascular Ctr. Research Institute (Japan)

By developing a computer-aided modeling system, the 3D shapes of infant's heart have been constructed interactively from quality-limited CT images for rapid prototyping of biomodels. The 3D model was obtained by following interactive steps: (1) rough region cropping, (2) outline extraction in each slice with locally-optimized threshold, (3) verification and correction of outline overlap, (4) 3D surface generation of inside wall, (5) connection of inside walls, (6) 3D surface generation of outside wall, (7) synthesis of self-consistent 3D surface. The manufactured biomodels revealed characteristic 3D shapes of heart such as left atrium and ventricle, aortic arch, subclavian artery and right auricle. Their real shape of cavity and vessel is suitable for surgery planning and simulation. It is a clear advantage over so-called "blood-pool" model which is massive and often found in 3D visualization of CT images as volume rendering perspective. The developed system contributed both to quality improvement and to modeling-time reduction, which may suggest a practical approach to establish a routine process for manufacturing heart biomodels. Further study on the system performance is now still in progress.

8290-43, Poster Session

Operation-free 3D visualization of pulsatile tissues in freehand ultrasonic diagnosis

K. Kawata, M. Fukuzawa, N. Nakamori, Kyoto Institute of Technology (Japan); Y. Kitsunezuka, Saiseikai Hyogo-ken Hospital (Japan)

The pulsatile tissues due to local blood flow of arteries, which is a significant diagnostic feature of ischemic diseases, have been detected from the freehand 2D ultrasonic (US) movies of several sections and visualized their 3D structure as re-sampled scatter diagram without any adjusting operation. It was realized by implementing automatic parameter-determination functionalities in the PC-based 3D visualization system designed to be used at the bedside by connecting directly to the conventional US apparatus (ATL HDI5000) and a compact tilt-sensor attached to the US probes of 5-7 MHz suitable for neonatal cranium. The major parameters to be adjusted were viewpoint, field of view (FOV), re-sampling interval Q and threshold of pulsatile strength ITH. The viewpoint and FOV were determined based on the probe orientation, while the Q and ITH were determined based on statistics of pulsatile strength. By using the 2D US movies between the frontal and the posterior coronal sections, the 3D structure of pulsatile tissue was described well along middle and posterior cerebral arteries, Willis ring and cerebellar arteries in neonatal cranium. Since pediatricians are relieved from any adjusting operation, it is demonstrated that our technique is practical to assist freehand US diagnosis at the bedside.

8290-44, Poster Session

Semiautomatic generation of semantic building models from image series

S. Wirtz, P. Decker, D. Weiß, D. W. Paulus, Univ. Koblenz-Landau (Germany)

In the recent years semantic based modeling, reconstruction of buildings and building recognition became more and more important. Semantic building models provide more information than just the geometry, thus making them more suitable for recognition or simulation tasks. The time consuming creation of such models and annotations makes an automatism desirable. We present an approach embedded in the photostitching tool Hugin, which is able to create a 3D model of a building with semantic annotations from a series of images. The

following steps are necessary: image acquisition, feature detection, pose estimation, 3D reconstruction, semantic interpretation/annotation and export in a suitable format. An entire automatic processing which delivers reliable models is not possible yet. Therefore, we present a semiautomatic approach towards semantic model generation. Our approach reduces the interaction with the system to a minimum on the supposition that reliable and accurate models are needed. The resulting model contains semantic, geometric and appearance information and is represented in Geography Markup Language (GML).

8290-45, Poster Session

Complex virtual urban environment modeling from CityGML data and OGC web services: application to the SIMFOR project

J. Chambealland, G. Gesquière, Lab. des Sciences de l'Information et des Systèmes (France)

Due to the advances in computer graphics and network speed it is possible to navigate in 3D virtual world in real time. This technology proposed for example in computer games, has been adapted for training systems. In this context, a collaborative serious game for urban crisis management called SIMFOR is born in France. This project has been designed for intensive realistic training and consequently must allow the players to design new urban operational theater. In this goal, importing, structuring, processing and exchanging 3D urban data remains an important underlying problem. This communication will focus on the design of the 3D Environment Editor (EE) and the related data processes needed to prepare the data flow to be exploitable by the runtime environment of SIMFOR. We will use solutions proposed by the Open Geospatial Consortium (OGC) to aggregate and share data. A presentation of the proposed architecture will be proposed. The overall design of the EE and some strategies for efficiently analyzing, displaying and exporting large amount of urban CityGML information will be presented in this part. The last part will be devoted to an example illustrating the potentiality of the EE and the reliability of the proposed data processing.

8290-46, Poster Session

Liquid crystal materials and structures for image processing and 3D shape acquisition

K. Garbat, Military Univ. of Technology (Poland); P. Garbat, Warsaw Univ. of Technology (Poland)

Liquid crystals have been extensively studied and are massively used in display and optical processing technology. Their recent use to provide 3D shape acquisition has focused the attention on new specific requirements. Recent improvements in the structure light measurements system based on digital light projection supported by processing allow rapid acquisition of data about 3D real objects. We present here the main relevant parameters of liquid crystal for image processing and 3D shape acquisition and we compare the main liquid crystal options which can be used with their respective advantages. We propose here to compare performance of several types of liquid crystal materials: nematic mixtures with high and medium optical and dielectrical anisotropies and relatively low rotational viscosities, nematic materials which may operate in TN mode twisted, materials for dual frequency addressing systems and V-shaped switching smectics liquid crystal.

8290-47, Poster Session

Piece-wise linear function estimation for platelet-based depth maps coding using edge detection

D. Sebai, The Univ. of Manouba (Tunisia)

Many research works on depth maps coding issues have been carried out such as those based on segmentation. The main idea is to subdivide the depth map into homogenous regions on which pixels disparities are more or less similar. In this context, the platelet-based coding approach is based on a recursive quadtree segmentation of the depth map. However, the estimation of the wedgelet and platelet functions is a computationally expensive task that has an exponentially increasing complexity in step with the depth map resolution. The main idea of the proposed method is based on wedgelet and platelet functions' definition: two constant or linear functions separated by a straight subdivision line. The constant or linear functions are used to approximate the two sub-regions of a node while the straight line is used to approximate a grayscale discontinuity so an object boundary. Therefore, we operate an edge detection for each quadtree node to look for the subdivision line.

8290-48, Poster Session

Source modeling for effective 3D video delivery

C. T. Hewage, M. G. Martini, Kingston Univ. (United Kingdom)

The rendered 3D video quality at the receiver-side will be affected by the rendering artifacts as well as by the concealment errors which occur in the process of recovering missing 3D video packets. Therefore it is vital to have an understanding of the artifacts prior to transmitting data. Therefore this work proposes a model to quantify rendering and concealment errors at the sender-side and use the information generated through the model to effectively deliver 3D video content.

8290-49, Poster Session

New technique for capturing images containing invisible depth information on object using brightness modulated right

S. Isaka, K. Uehira, Kanagawa Institute of Technology (Japan)

This paper presents a new technique for capturing images where depth information on the object is invisibly and simultaneously embedded in its 2-D image when the object image is taken with a camera. An object is illuminated by light that contains invisible information whose characteristics change depending on depth; therefore, the images of objects captured with a camera also contain such information invisibly. This invisible information on depth can be extracted by appropriate image processing from the captured image of the object. Images taken with this technique can be treated as conventional 2-D images because the image format is for conventional 2-D images. 3-D images can also be constructed by abstracting depth information embedded in the image. We carried out experiments using line and space pattern and confirmed the feasibility of the technology by demonstrating that the projected pattern could be embedded in the captured image invisibly and its frequency component, which depended on the depth of the object, could be read out from the captured image.

8290-50, Poster Session

Interactive 3D segmentation by tubular envelope model for thoracic aorta treatment

P. J. Lubniewski, Univ. d'Auvergne Clermont-Ferrand I (France) and Univ. Kardynala Stefana Wyszynskiego w Warszawie (Poland); B. Miguel, V. Sauvage, C. Lohou, Univ. d'Auvergne Clermont-Ferrand I (France)

We propose a novel interactive 3D segmentation approach and geometric model definition called a tubular envelope model. It is conceived to express tubular objects shape, specially aortas, also including pathological aortic dissection cases. The main challenges we have achieved are the speed and interactivity of the construction. The computer program designed for this task gives the user full control of the shape and precision, with no significant computational errors. The low complexity of the model and ease of interactive design makes the

tubular envelope very effective in comparison to other segmentation methods. The model accuracy is adjustable by the user, according to his requirements, but the time of construction is always significantly smaller. The explicit geometric form of the envelope could be directly used to other proceedings like shape approximation, region of interest for more precise segmentation or particular feature extraction, parametric deformable model definition, rapid 2D projection, and others.

8290-51, Poster Session

A parallel stereo reconstruction algorithm with applications in entomology (APSRA)

R. Bhasin, Univ. of Illinois at Urbana Champaign (United States); W. J. Jang, J. C. Hart, Univ. of Illinois at Urbana-Champaign (United States)

We propose a fast parallel algorithm for reconstruction of 3-Dimensional point clouds of insects from binocular stereo image pairs using a hierarchical approach for disparity estimation. Entomologists study various features of insects to classify them, build their distribution maps, and discover genetic links between specimens among various other essential tasks.This information is important to the pesticide and the pharmaceutical industries among others. When considering the large collections of insects Entomologists analyze, it becomes difficult to physically handle the entire collection and share the data with researchers across the world. With the method presented in our work, Entomologists can create an image database for their collections and use the 3D models for studying the shape and structure of the insects thus making it easier to maintain and share. Initial feedback shows that the reconstructed 3D models are good and preserve the shape and size of the specimen. We further optimize our results to incorporate multi-view stereo which produces better overall structure of the insects. Our main contribution is applying stereoscopic vision techniques to Entomology to solve the problems faced by Entomologists.

8290-07, Session 4

Depth map upscaling through edge-weighted optimization

S. Schwarz, M. Sjöström, R. Olsson, Mid Sweden Univ. (Sweden)

Dense and precise depth information is a pre-requisite in auto-stereoscopic three-dimensional television (3DTV). Only accurate depth maps assure a high quality view synthesis, but this information is not always easily obtained. Depth gained by stereo correspondence matching suffers from disocclusions and low-texturized regions, leading to erroneous depth maps. These errors can be avoided by using depth from dedicated range sensors, e.g. time-of-flight sensors. Because these sensors only have restricted resolution, the resulting depth data need to be adjusted to the resolution of the appropriate texture frame. Standard upscaling methods such as nearest neighbor or bicubic filtering provide only limited quality results. This paper proposes a solution for upscaling low resolution depth data to match high resolution texture data. We introduce the Edge Weighted Optimization Concept (EWOC) for fusing low resolution depth maps with corresponding high resolution video frames by solving an over-determined linear equation system. Similar to other approaches, EWOC takes information from the high resolution texture, but additionally validates this information with the low resolution depth to accentuate correlated data. Objective tests show an improvement in depth map quality in comparison to other upscaling approaches. This improvement is subjectively confirmed in the resulting view syntheses.

8290-08, Session 4

Adaptive switching filter for noise removal in highly corrupted depth maps from time-of-flight image sensors

S. Lee, K. Bae, K. Kyung, T. Kim, SAMSUNG Electronics Co., Ltd. (Korea, Republic of)

In this work, we present an adaptive switching filter for noise reduction and sharpness preservation in depth maps provided by Time-of-Flight (ToF) image sensors. Median filter and bilateral filter are commonly used in cost-sensitive applications where low computational complexity is needed. However, median filter blurs fine details and edges in depth map while bilateral filter works poorly with impulse noise present in the image. Since the variance of depth is inversely proportional to intensity, we suggest an adaptive filter that switches between median filter and bilateral filter based on the level of intensity. If a region of interest has low intensity indicating low confidence level of measured depth data, then median filter is applied on the depth at the position while regions with high level of intensity is processed with bilateral filter using Gaussian kernel with adaptive weights. Results show that the suggested algorithm performs surface smoothing and detail preservation as well as median filter and bilateral filter, respectively. By using the suggested algorithm, significant gain in visual quality is obtained in depth maps while low computational cost is maintained.

8290-09, Session 4

Parametric model-based noise reduction for ToF depth sensors

Y. S. Kim, B. Kang, H. Lim, O. Choi, K. Lee, J. D. K. Kim, C. Kim, Samsung Advanced Institute of Technology (Korea, Republic of)

This paper presents a novel Time-of-Flight (ToF) depth denoising algorithm based on parametric noise modeling. ToF depth image includes space varying noise which is related to IR intensity value at each pixel. By assuming ToF depth noise as additive white Gaussian noise, ToF depth noise (or standard deviation) can be modeled by using a power function of IR intensity. Meanwhile, nonlocal means filter is popularly used as an edge-preserving denoising method for removing additive Gaussian noise. To remove space varying depth noise, we propose an adaptive nonlocal means filtering. According to the estimated noise, the search window and weighting coefficient are adaptively determined at each pixel so that pixels with large noise variance are strongly filtered and pixels with small noise variance are weakly filtered. Experimental results demonstrate that the proposed algorithm provides good denoising performance while preserving details or edges compared to the typical nonlocal means filtering.

8290-10, Session 4

Silhouette extraction using color and depth information

E. V. Tolstaya, V. Bucha, Samsung Electronics Co., Ltd. (Russian Federation)

Recently applications involving capture of scenes with object of interest among surroundings gained high popularity. Such applications include video surveillance, human motion capture, human-computer interaction, etc. For proper analysis of object of interest a necessary step is to separate object of interest from surroundings, i. e. perform background subtraction (or silhouette extraction). This is a challenging task because of several problems: slight changes in background; shadows, caused by the object of interest; and similarly colored objects. In this work we propose a new method for extracting silhouette of an object of interest, based upon the joint use of both depth (range) and color data. Depth-based data is independent of color image data, and hence not affected by the limitations associated with color-based segmentation, such as shadows and similarly colored objects. At the initial moment image of background (not containing object of interest) is present, and it is updated in every frame using extracted silhouette using "running average". Silhouette extraction method is based on k-means clustering of depth data and color difference data, and per-pixel silhouette mask computation, using clusters centroids. Proposed solution is very fast, allows real-time processing of video. Developed algorithm was successfully applied in human recognition application and provided good results for modeling human figure.

8290-11, Session 4

Discrete and continuous optimizations for depth image super-resolution

O. Choi, H. Lim, Y. S. Kim, B. Kang, K. Lee, J. D. K. Kim, C. Kim, Samsung Advanced Institute of Technology (Korea, Republic of)

Recently, a Time-of-Flight 2D/3D image sensor has been developed, which is able to capture a perfectly aligned pair of a color and a depth image. To increase the sensitivity of infrared light, the sensor electrically combines multiple adjacent pixels into a depth pixel at the expense of depth image resolution.

To restore the resolution, we present a depth image super-resolution method that uses a high-resolution color image aligned with the input depth image. The input depth image is interpolated into the size of the color image, and our discrete optimization converts the interpolated depth image into a high-resolution disparity image, whose discontinuities precisely coincide with object boundaries. Subsequently, the interpolated depth image is iteratively filtered, preserving the disparity discontinuities. Meanwhile, our unique way of enforcing the depth reconstruction constraint gives a high-resolution depth image that is able to perfectly reconstruct the original input depth image.

We show the effectiveness of the proposed method both quantitatively and qualitatively, comparing the proposed method with two existing methods. The experiment reveals that the proposed method gives sharp high-resolution depth images with less error than the two methods, for scaling factors of 2, 4, and 8.

8290-12, Session 4

Superpixel-based depth image superresolution

Y. Soh, Seoul National Univ. (Korea, Republic of); J. Sim, Ulsan National Institute of Science and Technology (Korea, Republic of); C. Kim, Korea Univ. (Korea, Republic of); S. U. Lee, Seoul National Univ. (Korea, Republic of)

In recent years, depth acquisition from a scene has become easier, due to the development of Time of Flight (TOF) depth sensors. However, the depth images derived from current TOF sensors suffer from low resolution and noise. The state of the art TOF sensor retrieves the depth images at the size of 320 by 240 pixels which a relatively low resolution compared to the conventional image sensors. Because of the resolution difference between TOF sensor and CCD image sensor, there have been attempts to match the resolutions between the two differently obtained images, by increasing the resolution of depth image to that of the color image, which is called super-resolution. Using the modified bilateral filter is a popular approach to tackle this problem, due to its suitability for real time implementations [1][2]. The main idea of the bilateral filter is to apply an edge preserving smoothing filter [3]. The following assumptions are made to use such filter: the occurrence of edges or intensity gradients between depth image and color image are highly correlated. An early work using the bilateral filter proposes a fusion method called joint bilateral upsampling [2]. Using the depth image and color image both as inputs, the depth map is adaptively smoothened depending on the color similarity. It is shown that, the results suffer from artifacts such as texture copying or edge smoothing in cases where the assumption does not hold. In more recent works by F. Garcia et al. [1], attempts were made to reduce the artifacts of edge smoothing and texture copying. Compared to the work in [2] where the range term is always determined by the color similarity term, [1] extends its work by introducing the credibility map as an additional factor. The credibility map assigns low credibility to the pixels along the edge in the depth image considering them unreliable. This enables the range term to be adaptively weighted depending on the credibility. This approach significantly reduces the texture copying artifact but still suffers from the edge smoothing artifact. Using MRF to solve the problem was first presented by Diebel et al [4] and was extended by Gloud et al [5]. By defining the posterior probability of the depth value for each pixel in the reconstructed depth image domain as a Markov Random Field, both papers solve the problem using the maximum-

a-posteriori (MAP) solution. However, these approaches also share the assumption that the occurrences of edges or intensity gradients between a depth image and a color image are highly correlated, and thus they suffer from the same artifacts previously mentioned. Also, the computation load compared to other approaches is significantly high which makes it not applicable to certain applications. In this work, we point out that the joint bilateral filtering approach is inherited to suffer from an edge smoothing artifact due to its weighting-based operations. This can be explained by examining the range term used in these approaches. The range term is the key factor which determines the weight of smoothing term according to the strength of edges. In all the joint bilateral filtering approaches, a continuous Gaussian kernel is used as the range term, which is the reason a certain amount of edge smoothing is inevitable. When the edges are strong enough, the range term is a negligible value and the weight of smoothing term will be zero, which preserves the edges. On the contrary, when the edges are weak, a non-negligible amount of smoothing will occur which results in edge smoothing. Smoothed edges may not be noticeable in the results shown in previous works. However, when we reconstruct 3D scenes using the color image and the corresponding depth image, a few incorrectly assigned pixels may cause a great degree of perceived awkwardness in human visual system. Therefore, we propose a superpixel based super-resolution algorithm that enhances the edges instead of preserving them. By using a superpixel based approach, we exploit the fact that a superpixel preserves all the boundary edges and the depth image is piecewise planar. The proposed algorithm is composed of three stages. In the first stage, the size of the low resolution depth images is increased to the target resolution using bi-cubic interpolation. The increased depth image is then divided into superpixels using oversegmentation. During the oversegmentation process, the boundary edges are preserved. Because a depth image can be assumed to be piecewise planar, for each superpixel, we can fit the data points to a plane function. In this process the smooth edges which occurred due to bi-cubic interpolation are exterminated and the edges become discrete along the boundary edges since the plane parameters are determined by the non-smoothed regions. At the second stage, using the oversegmentation results of the high resolution color image, the high frequency edge information is transferred to the superpixel map used in the first stage. This process is essential in cases when the resolution difference between the low resolution depth image and the high resolution color image is large since the edges in the low resolution depth image and the edges in the high resolution color image are prone to be misaligned. At the last stage, using the refined superpixel map and the plane functions for each superpixel, we reconstruct the high resolution depth map. During the plane fitting process, there may be discontinuities along regions where there are no edges. By adopting a MAP-MRF framework, we handle such outliers and render a high quality depth image. By using the proposed superpixel based approach, we can prevent the artifacts that inevitably occur in the previous approaches, as well as enhance the depth image by sharpening the edges. Moreover, by dividing the image into several superpixels, we have the benefit of performing this algorithm in a parallel manner which allows reduction of computation time.

8290-13, Session 4

Efficient spatio-temporal hole filling strategy for Kinect depth maps

M. Camplani, L. Salgado, Univ. Politécnica de Madrid (Spain)

The market of active depth cameras and the computer vision community have been revolutionized by the Microsoft Kinect thanks to its low price and good resolution. At the moment several human-computer interaction applications have been developed. However, the Kinect depth maps present noise-related problems that have to be solved in order to improve their accuracy and to broaden its future applicability. One of the most important noise effects on the accuracy of the Kinect depth maps is due to the presence of regions for which the camera is not able to correctly estimate the depth. Simple smoothing filtering procedures lead to poor results introducing blurring effect. The presented approach is an efficient hole filling strategy, based on a joint bilateral filter, that aims at improving the Kinect depth map accuracy. The proposed approach analyzes the visual information in order to reduce the blurring effect; the depth gradient information is

used to consider the local depth variation and, the temporal evolution of the neighbor pixels of the hole regions is included in the filtering process through a temporal-consistency map. Finally, the obtained new depth values are used in the filtering process in successive frames improving the depth map accuracy.

8290-14, Session 5

Experimental results of bispectral invariants discriminative power

K. Kubicki, Jagiellonian Univ. (Poland); R. Kakarala, Nanyang Technological Univ. (Singapore)

One of the main tools in shape matching and pattern recognition are invariants. For three-dimensional data, rotation invariants comprise of two main kinds: moments and spherical harmonic magnitudes. Both are well examined and both suffer from certain limitations. In search for better performance, a new kind of spherical-harmonic invariants have been proposed recently, called bispectral invariants. They are well-established from theoretical point of view. They posses numerous beneficial properties and advantages over other invariants, include the ability to distinguish rotation from reflection, and the sensitivity to phase. However, insufficient research has been conducted to check their behavior in practice. In this paper, results are presented pertaining to the discriminative power of bispectral invariants. Objects from Princeton Shape Benchmark database are used for evaluation. It is shown that the bispectral invariants outperform power spectral invariants, but perform worse than other descriptors proposed in the literature such as SHELLS and SHD. The difference in performance is attributable to the implicit filtering used to compute the invariants.

8290-15, Session 5

Evaluation of surface reconstruction from point clouds by shape parameters

L. Cao, F. J. Verbeek, Leiden Univ. (Netherlands)

3D surface reconstruction from a point cloud is an important research area for computer graphics and visualization; it can be potentially important for application in bio-imaging. In the past decade, numbers of algorithms for surface reconstruction have been proposed. Generally speaking, these methods can be separated into two categories: i.e., explicit representation and implicit approximation.

Although most of the methods are firmly based in theory, no analytical evaluation exists between the methods. The most straightforward way of evaluation is still by visual inspection. We search for a method that can precisely preserve the surface characteristics and is robust in the presence of noise, in order to improve the surface reconstruction method for biological models. Therefore, we use an analytical approach by selecting several features as surface descriptors and measuring them. We selected surface distance, surface area and surface curvature as three major features to compare quality of the surface created by the different algorithms. For ground truth values we used analytical shapes such as the sphere and the ellipsoid.

Here we present four classical surface reconstruction methods from the two categories mentioned above. These are the Power Crust, the Robust Cocone, the Fourier-based method and the Poisson reconstruction method. Our experimental results indicate that Poisson reconstruction method performs the best in the presence of noise.

8290-16, Session 5

3D mesh Reeb graph computation using commute-time and diffusion distances

R. EL Khoury, J. P. Vandeborre, M. Daoudi, Institut TELECOM (France) and TELECOM Lille 1 (France)

3D-model analysis plays an important role in numerous applications. In this paper, we present an approach for Reeb graph extraction using a novel mapping function. Our mapping function computes a real value for each vertex which provides interesting insights to describe topology structure of the 3D-model. We perform discrete contour for each vertex according to our mapping function. Topology changes can be detected by discrete contours analysis to construct the Reeb graph. Our mapping function has some important properties. It is invariant to rigid and non rigid transformations, it is insensitive to noise, it is robust to small topology changes, and it does not depend on parameters. From the extracted skeleton, these properties show the signicant parts of a 3D-model. We retain the evaluation criteria to the properties of the mapping function, and compared them to those used in the state of the art. In the end, we present extracted Reeb graph on various models with dierent positions.

8290-17, Session 5

Geometric modeling of pelvic organs with thickness

T. Bay, Lab. des Sciences de l'Information et des Systèmes (France); Z. W. Chen, Univ. d'Evry-Val d'Essonne (France); R. Raffin, M. Daniel, Lab. des Sciences de l'Information et des Systèmes (France); P. Joli, Univ. d'Evry-Val d'Essonne (France); Z. Q. Feng, M. Bellemare, Lab. des Sciences de l'Information et des Systèmes (France)

Physiological changes in the spatial configuration of the internal organs in the abdomen can induce different disorders that need surgery. Following the complexity of the surgical procedure, mechanical simulations are necessary. But the in vivo factor makes complicate the study of pelvic organs. To determine a realistic behavior of the organs, a physical modeling associated with an accurate geometric model is therefore required. Our approach is integrated in the partnership between a geometric and physical module.

The Geometric Modeling seeks to build a continuous geometric model. From a dataset of 3D points provided by a Segmentation step, surfaces are created through a B-spline fitting process. An energy function is built to measure the bidirectional distance between surface and data. This energy is minimized with an alternate iterative Hoschek-like method. A thickness is added with an offset formulation, and the geometric model is finally exported in a hexahedral mesh.

Afterwards, the Physical Modeling tries to calculate the properties of the soft tissues to simulate the organs displacements. The physical parameters attached to the data are determined with a feedback loop between finite-elements deformations and ground-truth acquisition (dynamic MRI).

8290-18, Session 5

Refined facial disparity maps for automatic creation of 3D avatars

R. Pagés, F. Morán, L. Salgado, D. Berjón, Univ. Politécnica de Madrid (Spain)

We propose a new method to automatically refine a facial disparity map obtained with standard cameras and lighting by using a smart combination of traditional computer vision and 3D graphics techniques. Our system inputs two stereo images acquired with normal (calibrated) cameras and uses dense disparity estimation to obtain a coarse initial disparity map, and SIFT to detect several feature points in the subject's face. We then use these points as anchors to modify the disparity in the facial area by building a Delaunay triangulation of their convex hull and interpolating their disparity values inside each triangle. We thus obtain a refined disparity map representing much better the subject's facial features. This refined facial disparity map may be easily transformed, through the camera calibration parameters, into a depth map to be used, also automatically, to improve the facial mesh of a 3D avatar to match the subject's real human features.

Conference 8290

IS&T / SPIE

8290-19, Session 5

Fast human pose estimation using 3D Zernike descriptors

D. Berjón, F. Morán, Univ. Politécnica de Madrid (Spain)

Abstract Markerless video-based human pose estimation algorithms face a high-dimensional problem that is frequently broken down into several lower-dimensional problems by estimating the pose of each limb separately. However, to do so they need to reliably locate the torso, for which they typically rely on time coherence and tracking algorithms that when lose track result in catastrophic failure of the process, requiring human intervention and thus precluding their usage in real-time applications. We propose a very fast rough pose estimation scheme based on global shape descriptors built on 3D Zernike moments. Using an articulated model that we configure in many poses, a large database of descriptor/pose pairs can be computed off-line. Thus, the only steps that must be done on-line are the extraction of the descriptors for each input volume and a search against the database to get the most likely poses. While the result of such process is not a fine pose estimation, it can be useful to help more sophisticated algorithms to regain track or make more educated guesses when creating new particles in particle-based schemes. We have achieved a performance of about ten fps on a single computer using a database of about one million entries.

8290-20, Session 5

Analysis of binning of normals for spherical harmonic cross-correlation

R. Larkins, M. J. Cree, A. A. Dorrington, The Univ. of Waikato (New Zealand)

The registration of overlapping point clouds has recently received increased attention due to the growing availability of 3D sensors. Spherical harmonic cross-correlation is a robust registration technique that uses the normals of two overlapping point clouds to bring them into coarse rotational alignment. This registration technique however has a high computational cost as spherical harmonics need to be calculated for every normal. By binning the normals, the computational efficiency is improved as the spherical harmonics can be pre-computed and cached at each bin location. In this paper we evaluate the accuracy and efficiency of two previously reported binning methods and a third that we propose. It is found that of the two methods used in previous literature, one method of binning decreases the achieved accuracy for certain registrations, and the other does not allow an exact number of bins to be specified a prior. The new proposed method of binning overcomes these disadvantages. The primary goal of the analysis is to identify a binning method that performs as well as the original cross-correlation algorithm without binning, while also providing a significant improvement in computational efficiency.

8290-21, Session 5

Topology reconstruction for B-Rep modeling from 3D mesh in reverse engineering applications

R. Bénière, Lab. d'Informatique de Robotique et de Microelectronique de Montpellier (France) and C4W (France); G. Subsol, Lab. d'Informatique de Robotique et de Microelectronique de Montpellier (France); G. Gesquière, Lab. des Sciences de l'Information et des Systèmes (France); F. Le Breton, C4W (France); W. Puech, Lab. d'Informatique de Robotique et de Microelectronique de Montpellier (France)

In an industrial context, recovering a B-Rep model from a 3D mesh is necessary to make modifications or to exchange data with a format including continuous representation of objects like STEP. But for many reasons, the initial continuous object can be lost after a discretization process to display or exchange data. The mesh can also be deformed after a numerical computation. It is then important to have a method to create a new continuous model of the object from a mesh. For CAD object, the first step is to detect simple primitives like planes, spheres, cones and cylinders from a 3D mesh.

In previous works, we use curvature features to recover these primitives. Segmentation in point areas is based on the curvature features computed for each vertex. Each one is associated with a primitive type. Final parameters of the primitives are found with a fitting process according to the curvature features.

The novelty of this paper is a method to determine the topology of this set of 3D primitive extracted previously. The topology definition is based on both continuous informations from primitives and discrete informations from point areas. It leads to construct adjacency relations, then the wires and to finish the B-Rep model.

8290-22, Session 5

An evaluation of local shape descriptors for 3D shape retrieval

S. Y. Tang, Princeton Univ. (United States); A. Godil, National Institute of Standards and Technology (United States)

As the usage of 3D models increases, so does the importance of developing accurate 3D shape retrieval algorithms. Many local descriptors have been proposed to approach this problem; however, they are evaluated on different datasets or used within different algorithms, making them difficult to compare. Using the SHREC 2011 Shape Retrieval Contest of Non-rigid 3D Watertight Meshes dataset, we compare a set of local shape descriptors by applying each to the bag-of-words paradigm. We go on to assess possible benefits of using sample points returned by a salient point detection algorithm rather than finding a random selection. Finally, two local descriptors are combined and applied together. In this paper, we present results and observations from these experiments.

8290-23, Session 6

Fractal geometry and multimedia retrieval: a theoretical framework

E. Paquet, National Research Council Canada (Canada); H. L. Viktor, Univ. of Ottawa (Canada)

This paper presents a theoretical analysis of the occurrence of fractal geometry within index spaces and discusses the impact for multimedia retrieval. Firstly, we explain how to detect the presence of such a fractal geometry. Then, with the fractal hypothesis in hand, we analyze the impact of this geometry when calculating the distance between indexes and searching for similar multimedia objects. We demonstrate that it is possible to define probability distributions in fractal index space and we present an approach to calculate them. Practical consequences are discussed, with particular emphasis to multimedia retrieval methods based on Bayesian analysis.

8290-24, Session 6

Spatial modeling of bone microarchitecture

H. Li, K. Li, T. Kim, A. Zhang, M. Ramanathan, Univ. at Buffalo (United States)

In this research, we focus on developing the spatial component of such a modeling framework and demonstrate how bone microarchitecture, bone composition and fracture risk can be incorporate to obtain mechanistic assessments of fracture risk from that can be obtained in clinical settings. Our model for bone mineral is developed and its parameters were estimated from imaging data obtained with dual energy x-ray absorptiometry and x-ray imaging methods. Using these parameters, we develop a proper 3D Microstructure Bone Model.

We develop and evaluate a novel three-dimensional computational bone framework capable of providing: 1) structurally derived

quantitative measures of bone microarchitecture, 2) analysis of bone mineral density (BMD)and fracture risk, 3)a mechanistic understanding of the relationship of bone microarchitecture to fracture risk, and 4) insights into other clinical problems, such as neurological disorders and leukemia, by identifying the disease progression and prevention. From the bone microscopic structures, a three-dimensional rod-like structure network is generated as an abstracted bone network. Attributes of bone microscopic structures are calculated and applied to bone network.

8290-25, Session 6

A new affine invariant method for image matching

P. Montesinos, Ecole des Mines d'Alès (France)

This paper describes a new approach in color or grey-scale image matching by points of interest. As many point matching methods, this method is based on two main steps: computation of points and descriptors, followed by a matching process. This paper presents the new descriptor defined and a matching process dealing with the data obtained from the descriptor.

8290-26, Session 6

2D-3D feature association via projective transform invariants for model-based 3D pose estimation

O. S. Gedik, A. A. Alatan, Middle East Technical Univ. (Turkey)

The three dimensional (3D) tracking of rigid objects is required in many applications, such as 3D television (3DTV) and augmented reality. Accurate and robust pose estimates enable improved structure reconstructions for 3DTV and reduce jitter in augmented reality scenarios. On the other hand, reliable 2D-3D feature association is one of the most crucial requirements for obtaining high quality 3D pose estimates. In this paper, a 2D-3D registration method which is based on projective transform invariants is proposed. Due to the fact that projective transform invariants are highly dependent on 2D and 3D coordinates, the proposed method relies on pose consistencies in order to increase robustness of 2D-3D association. The reliability of the approach is shown by comparisons with RANSAC, perspective factorization and SoftPOSIT based methods on real and artificial data.

8290-27, Session 6

Reprocessing anaglyph images

H. G. Dietz, Univ. of Kentucky (United States)

In related work, we have shown that conventional digital cameras easily can be modified to directly capture anaglyphs. Anaglyph images have commonly been used to encode stereo image pairs for viewing, but anaglyphs also can be treated as an efficient encoding of two-view image data for reprocessing. Each of the two views encoded within an anaglyph has only partial color information, but our preliminary results demonstrate that the "lost" information can be approximately recovered with any of a variety of reasonably efficient algorithms. This not only allows credible full-color stereo pairs be computationally extracted, but also enables more sophisticated computational photography transformations such as creation of depthmaps and various types of point-spread-function (PSF) substitutions.

8290-28, Session 7

X-ray stereo imaging for micro 3D motions within non-transparent objects

W. H. M. Salih, J. A. N. Buytaert, J. J. J. Dirckx, Univ. Antwerpen (Belgium)

We propose a new technique to measure the 3D motion of marker points along a straight path within an object using x-ray stereo projections. From recordings of two x-ray projections with 90 degree separation angle, the 3D coordinates of marker points can be determined. By synchronizing the x-ray exposure time to the motion event, a moving marker leaves a trace in the image of which the gray scale is linearly proportional to the marker velocity. From the gray scale along the motion path, the 3D motion (velocity) is obtained. The path of motion was reconstructed and compared with the applied waveform. The results showed that the accuracy is in order of 5%. The difference of displacement amplitude between the new method and laser vibrometry was less than 5μm.

We demonstrated the method on the malleus ossicle motion in the gerbil middle ear as a function of pressure applied on the eardrum. The new method has the advantage over existing methods such as laser vibrometry that the structures under study do not need to be visually exposed. Due to the short measurement time and the high resolution, the method can be useful in the field of biomechanics for a variety of applications.

8290-29, Session 7

A stereoscopic imaging system for laser back scatter-based trajectory measurement in ballistics: part 2

U. Chalupka, H. Rothe, Helmut-Schmidt-Univ. (Germany)

The progress on a laser- and stereo-camera-based trajectory measurement system that we already proposed and described in recent publications is given. The system design was extended from one to two more powerful, DSP-controllable LASER systems. Experimental results of the extended system using different projectile-/weapon combinations will be shown and discussed. Automatic processing of acquired images using common 3DIP imaging techniques was realized. Processing steps to extract trajectory segments from images as representative for the current application will be presented. Used algorithms for backward-calculation of the projectile trajectory will be shown. Verification of produced results is done against simulated trajectories, once in terms of detection robustness and once in terms of detection accuracy. Fields of use for the current system are within the ballistic domain. The first purpose is for trajectory measurement of small and middle caliber projectiles on a shooting range. Extension to big caliber projectiles as well as an application for sniper detection is imaginable, but would require further work. Beside classical RADAR, acoustic and optical projectile detection methods, the current system represents a further projectile location method under the new class of electro-optical methods that have been evolved in recent decades and that uses 3D imaging acquisition and processing techniques.

8290-30, Session 7

Single frame coaxial 3D measurement using depth from defocus of projection system

T. Kurihara, S. Ando, The Univ. of Tokyo (Japan)

We propose coaxial 3D measurement system based on depth from defocus of projection system. In this system, the stripe pattern is projected on the object, and it moves on object's surface generating temporal variation of the light intensity. The projected stripe pattern is defocused depending on its distance from the focal plane. By moving the stripe pattern, defocused spatial frequency component is captured by temporal frequency analysis. DLP LightCommander is used for moving the stripe pattern fast. We use correlation image sensor(CIS), which is developed by us, to capture the temporal frequency component in a single frame. The CIS outputs the Fourier coefficients of incident light in each pixel for every frames, and therefore it enables single frame 3D measurement. Evaluation experiments show that projection defocus depends on the distance from the focal plane, and it can be used for 3D measurement.

8290-31, Session 7

Multidirectional four-dimensional shape measurement system

J. Lenar, R. Sitnik, M. Witkowski, Warsaw Univ. of Technology (Poland)

Currently, a lot of different scanning techniques are used for 3D imaging of human body. Most of existing systems are based on static registration of internal structures using MRI or CT techniques as well as 3D scanning of outer surface of human body by laser triangulation or structured light methods. On the other hand there is an existing mature 4D method based on tracking in time the position of retro-reflective markers attached to human body. There are two main drawbacks of this solution: markers are attached to skin (no real skeleton movement is registered) and it gives (x, y, z, t) coordinates only in those points (not for the whole surface). In this paper we present a novel multidirectional structured light measurement system that is capable of measuring 3D shape of human body surface with frequency reaching 60Hz. The developed system consists of two spectrally separated and hardware-synchronized 4D measurement heads. The principle of the measurement is based on single frame analysis. Projected frame is composed from sine-modulated intensity pattern and a special stripe allowing absolute phase measurement. Several different geometrical set-ups will be proposed depending on type of movements that are to be registered. Initial measurement uncertainty assessment will be presented together with the calibration procedure.

8290-32, Session 8

Estimation of surface normal vectors based on 3D scanning from heating approach

O. Aubreton, Univ. de Bourgogne (France); G. Eren, Galatasaray Üniv. (Turkey); F. Truchetet, Univ. de Bourgogne (France)

The Scanning From Heating is a 3D scanning approach initially developed to realise 3D acquisition of transparent or specular surfaces. A laser source is used to create a local heating point. An infrared camera is used to observe the IR radiation emitted by the scene. The 2D coordinates of the heated point are computed in the 2D image of the camera. Knowing the parameters of the system (which are obtained by a previous calibration), the 3D coordinates of the point are computed using triangulation method. In this article we will present an extension of this technique. We propose here to analyse the shape of the hot spot observed by the IR camera, and, from the analysis to determine information on the local orientation of the surface at each measured point.

8290-33, Session 8

First topographical features for the age determination of latent biometric fingerprint traces using a 3D optical and contact-less chromatic white light (CWL) sensor

R. Merkel, Otto-von-Guericke-Univ. Magdeburg (Germany); S. Gruhn, Fachhochschule Brandenburg (Germany); J. Dittmann, Otto-von-Guericke-Univ. Magdeburg (Germany); C. Vielhauer, Fachhochschule Brandenburg (Germany)

Determining the age of a latent fingerprint trace found at a crime scene is an unresolved research issue since many decades. In this paper, we describe a novel approach lifting such traces with a powerful 3D image acquisition system and transferring them into a 3D aging feature. In particular, we use a contact-less, high-resolution Chromatic White Light (CWL) sensor to observe the acquired 3D shapes and their transformations over time. We also suggest examining different preprocessing techniques to deal with noise and other measurement artifacts, to transform the topographical fingerprint data into a form which can be used for segmentation and pattern recognition. Based on prior work, we introduce an aging feature called binary pixel and extend it to 3D-topography images. For this purpose, we use a test set of 10 test series, each scanning a fingerprint sample from a hard disk platter surface over 100 hours every 30 minutes, leading to a total of 2000 three-dimensional fingerprint images. We show that with adequate preprocessing techniques (Difference of Gaussian, fingerprint pixel masking, local variance), a characteristic logarithmic aging property is produced for the 3D-topography images, which can be considered for the age determination. Future work will be the fusion of such 3D aging feature with our prior developed 2D features, to improve the accuracy of the overall age determination approach.

8290-34, Session 8

A single-imager, single-lens video camera prototype for 3D imaging

L. Christopher, W. Li, Indiana Univ.-Purdue Univ. Indianapolis (United States)

A new method for capturing 3D video from a single imager and lens is introduced. The benefit of this method is that it does not have the calibration and alignment issues associated with binocular 3D video cameras. It also does not require special ranging transmitters and sensors. Because it is a single lens/imager system, it is also less expensive than either the binocular or ranging cameras. Our system outputs a 2D image and associated depth image using the combination of microfluidic lens and Depth from Defocus (DfD) algorithm. The lens is capable of changing the focus to obtain two images at the normal video frame rate. The Depth from Defocus algorithm uses the in focus and out of focus images to infer depth. The basic principles of DfD are described in Pentland [1] and Chaudhuri [2]. We performed our experiments on synthetic and on the real aperture CMOS imager with microfluidic lens. On synthetic images, we found a two times reduction in percent ranging error. On camera images, our research showed that DfD combined with edge detection and segmentation provided subjective improvements in the images.

8290-35, Session 8

3D multimodal data fusion system

P. Garbat, Warsaw Univ. of Technology (Poland)

Recently one of the biggest challenges in computer vision is the 3D analysis and reconstruction of objects and scenes. This requires gathering real data about static or dynamic 3D objects and scenes and delivering them to a remote visualization system. We propose a hybrid, multimodal data capture system based on three band, near infrared, fringe projection system with fast camera, hi-resolutin camera for intensity and state of polarization acquisition unit.

This paper describes a fast algorithmic approach to combine Full-HD RGB images with 3D data (X,Y,Z), and information about state of polarization on object surface POL. We propose method based on ICP algorithms with hierarchical, multiscale 3D data representation. The method of merging points sets based on matching feature points (regions) in curvature scale space CSS. The feature vectors are determined using geometry, texture, polarization and other local parameters of matched 3D surface/volume. The matching process can be used in weak-calibration procedure of camera set. It makes the 3D video system more universal in comparison with methods based on fully calibrated setup.

8290-36, Session 8

Fully automatic 3D digitization of unknown objects using progressive data bounding box

A. Aigueperse, S. Khalfaoui, Lab. d'Electronique, Informatique et Image (France); R. Seulin, Y. Fougerolle, D. Fofi, Univ. de Bourgogne (France)

The goal of this work is to develop a complete system capable of

scanning objects automatically with minimum a priori information of its shape. We aim to establish a methodology for the automation of the 3D digitization chain. The paper presents a method based on the evolution of the Bounding Box of the object during the acquisition steps. The registration of the different acquired data is done using the modeling of the positioning system. The obtained models are analyzed and inspected in order to evaluate the robustness of our method. Tests with real objects have been performed and results of digitization are provided.

8290-37, Session 9

3D video compression with the H.264 codec

N. L. Karpinsky, S. Zhang, Iowa State Univ. (United States)

Advances in 3D scanning have enabled the real-time capture of high-resolution 3D video. With these advances comes the challenge of streaming and storing this 3D video in a manner that they can be quickly and effectively used. To do this different approaches have been taken, a popular one being image based encoding, which projects from 3D into 2D, uses 2D compression techniques, and then decodes from 2D back to 3D. One such technique that does this is the Holovideo technique, which we have shown to yield great compression ratios. However, the technique was originally designed for the RGB color space and until recently could not be used with codecs that use the YUV color space such as the H.264 codec. This paper addresses this issue, generalizing Holovideo to the YUV color space, allowing it to leverage the H.264 codec making it a viable solution for 3D video compression.

8290-38, Session 9

3D multiresolutions synchronization scheme based on feature point selection

N. Tournier, Lab. d'Informatique de Robotique et de Microelectronique de Montpellier (France) and Stratégies S.A. (France); W. Puech, G. Subsol, Lab. d'Informatique de Robotique et de Microelectronique de Montpellier (France); J. Pedeboy, Stratégies S.A. (France)

Multimedia protection is one of the main research challenges in computer sciences. We can encrypt the media in order to make the content unreadable without a secret key of decryption, protect the file with Digital Right Management (DRM), or embed an hidden message in the file (watermarking and steganography). We are interested in data hiding applications for 3D mesh. In this domain, there a mainly problem: the synchronization. It is the operation that permits to scan a mesh with a unique path and by selecting the same areas (vertices, triangles, quadrangles, for example) before and after the embedding.

In this paper, we propose a new synchronization technique based on feature point selection in a low resolution of the object. The building of the low resolution is made by decimation and the feature point selection is based on the discrete curvature computing. We evaluate the robustness of the synchronization in the low resolution and in the heigh resolution.

8290-39, Session 9

A content-adaptive scheme for reduced-complexity, multiview video coding

A. Avci, J. De Cock, R. Beernaert, J. De Smet, Univ. Gent (Belgium); Y. Meuret, Vrije Univ. Brussel (Belgium); P. Lambert, H. De Smet, Univ. Gent (Belgium)

Disparity estimation is a highly complex and time consuming process of the encoder. Since multiple views taken from a 2D camera array need to be coded at every time instance, the complexity of the encoder plays an important role besides the rate-distortion performance of the encoder. Because of the strong geometrical correspondence between views, a new frame type called D frame has been proposed in our

previous work to minimize the complexity of the encoder. By employing D frames in place of some of the P frames in the prediction structure, significant complexity gains can be achieved if the threshold value, which is a keystone element to adjust the complexity at the cost of quality and/or bit-rate, is selected wisely. In this work, a new adaptive method to calculate the threshold value automatically from existing information during the encoding process is presented. In this method, the threshold values are generated for each block of each D frame to increase the accuracy. The algorithm is applied to several image sets and 20.6% complexity gain is achieved by using the automatically generated threshold values without compromising the quality and bit-rate.

Conference 8291: Human Vision and Electronic Imaging XVII

Monday-Thursday 23-26 January 2012

Part of Proceedings of SPIE Vol. 8291 Human Vision and Electronic Imaging XVII

8291-01, Session 1

Image representations learned for natural images and a non-linear maximum synthesis rule with applications in computational photography

J. Hocke, E. Barth, T. Martinetz, Univ. zu Lübeck (Germany)

With sparse coding the learning of a basis is non-linear, but the basis elements are still linearly combined to form an image. We here propose to use a nonlinear synthesis rule, which is defined such that at each location in the image the point-wise maximum over all basis elements is used to generate the image. Only recently an attempt was made to learn a basis that is optimal given this synthesis rule (Lücke et al. who coined this Maximal Causes Analysis - MCA). One can argue that such a non-linear synthesis rule is more realistic for natural images because it would generate the image from its main features. The challenge then is to properly learn these main features given a particular set of images.

We here present two important extensions. First, we propose a novel and more efficient learning scheme based the Sparse Coding Neural Gas algorithm, which was designed to learn sparse bases for linear image synthesis and is adapted here to the non-linear MCA synthesis rule.

Second, we apply the novel algorithm to classical problems in Computational Photography such as image deblurring and image completion.

8291-02, Session 1

How to make a small phone camera shoot like a big DSLR: creating and fusing multi-modal exposure series

T. Binder, F. Kriener, C. Wichner, M. Wille, Nik Software GmbH (Germany); M. Wellner, T. Kaester, E. Barth, Pattern Recognition Lab. (Germany)

In the first part we present the design of an imaging system consisting of two cameras, one being a high-quality DSLR, and the other a cellphone camera (the Nokia N900 Frankencamera). The cameras are connected via USB. The Frankencamera may take many exposures with varying parameters, while the DSLR takes a reference image of the scene.

Since the system is designed to take many exposures of the same scene, a stable mechanical coupling of the cameras and the use of a tripod are required.

The second part deals with the processing of the resulting cellphone exposure stack, which is first aligned and then fused into a single image. Therefore the whole process of producing the single image is controlled by two parameter sets: the acquisition parameters of the exposure stack (number of exposures, shutter speed and focus setting of each exposure) and the fusion parameters.

The fusion algorithm is based on a dual-tree complex wavelet transform (DCWT) and involves both linear summation of the DCWT coefficients and nonlinear transforms of the coefficients. The nonlinearities enforce sparseness on the coefficients and implement divisive normalization.

Finally, the fused image is compared to the reference image taken with the DSLR. A perceptual quality index is defined, which expresses the perceived similarity of the fused and the reference image. We then use the index to optimize both the image acquisition and the fusion parameters.

The novelty of our approach is due to the joint multidimensional fusion of images with different acquisition parameters, the fact that we use a high-quality reference image and a perceptual quality index, and the joint multidimensional optimization of the acquisition and fusion parameters. The ultimate goal is to improve the image quality of cheap and small cameras by using perceptually optimized acquisition and fusion algorithms.

8291-03, Session 1

Metric image mosaics with bounded error

D. Ghosh, J. Tumblin, Northwestern Univ. (United States)

This paper describes an open-source multi-image stitching system that constructs metric planar mosaics suitable for building visual archives of mostly-flat museum artifacts such as paintings, frescoes, and floors. The system outperforms existing panorama-making software such as Microsoft's ICE, Brown and Lowe's 'AutoStitch', or even Canon Utilties 'PhotoStitch' that lack strict error bounds and rely on homographies applied to single-viewpoint image sets. Its advantages include: a) The new 'mosaic-making' system accepts input images from unknown, widely varying viewpoints including low-angle (or 'raking angle') viewpoints and lighting, making it suitable for planar surfaces far too large or too complex for a single viewpoint (e.g. floor of a cathedral; a large painting or mural); b) The mosaic result ensures pixel distances are directly proportional to flat-surface distances. The system imposes strict error-bounds on mosaic-result image error vs. point-correspondence mismatches measured in the source images. c) Its all-at-once optimizer/solver includes a 7-term parametric lens-distortion model that corrects for low-cost uncalibrated zoom lenses with unknown settings, and d) written within the SourceForge HUGIN/PanoTools open-source panorama-maker system, the new 'mosaic-maker' leaves the GUI unchanged and permits extensive scripting capabilities suitable for extremely large image sets. Unlike most panorama-makers, this mosaic-maker removes all foreshortening from the source images by a chain of physically-meaningful 3D parametric transformations (roll, pitch, yaw, tilt, spin) instead of conventional homographies. This more-elaborate but comprehensible process provides greater flexibility--users can individually enable each degree-of-freedom for each source image if desired. Further 2D parametric transforms (rotate, scale, translate) arrange the mosaic, and the system optimizes all of the source image's lens-correction, 2D and 3D parameters at once. After extensive tests with synthetic source images to test system robustness, we assembled hundreds of metric mosaics of one large painting at the Art Institute of Chicago. Built from a vast photo-archive (65,000 images) of tiled, variously-lit, raking-angle photos, the of assembled mosaics revealed several previously unknown overpainted features ('pentimenti') that helped clarify the history of this important artwork.

8291-04, Session 1

Single-lens 3D camera with extended depth-of-field

C. Perwass, L. Wietzke, Raytrix GmbH (Germany)

A camera design that has regained much popularity in the last couple of years is that of the plenoptic camera where a microlens array is placed in front of the image sensor. Taking pictures with a plenoptic camera offers a whole new way to perceive images. Such images are not static, but allow the user to modify the image after it has been recorded. The user can interactively change the focus, the point of view and the perceived depth of field.

While a standard plenoptic camera's effective resolution equals the

Conference 8291

number of microlenses, we present an advanced technology with an effective resolution of up to one quarter of the sensor resolution and an actually extended depth of field. Note that a standard plenoptic camera does not extend the depth of field as compared to a standard camera using the same lens and the same aperture. The microlens technology developed by Raytrix, on the other hand, increases the depth of field by up to to a factor of 6. This is achieved through specially designed microlens arrays that consist of microlenses of different focal lengths.

8291-05, Session 1

3D holoscopic video imaging system

J. H. Steurer, Arri Cine Technik (Germany)

The EU funded project 3D VIVANT is aiming to capture events automatically in three-dimensional (3D) space and deliver them to home viewers for realistic, interactive, fatigue-free and immersive playback without glasses. The central idea of this project is a new form of multimedia content (3D video and audio). 3D holoscopic imaging (also referred to as Integral Imaging) is a technique for creating full color 3D optical models that exist in space independently of the viewer.

The 3D VIVANT project is developing new technologies for capture and display of 3D content. For the capture, 3D VIVANT is taking full advantage of 3D holoscopic imaging technology, where a single camera is required. Hence in this project a world first single aperture ultra-high definition 3D holoscopic imaging camera is being constructed which will permit live capture of 3D content. For the display, the project is taking advantage of the Holovisio system to provide an immersive, ultra-high resolution presentation of 3D content.

3D VIVANT will make a significant number of technological advances, which demonstrate the capabilities of the 3D holoscopic video technology as an alternative 3D vision system especially in the field of entertainment.

8291-06, Session 2

Predictive rendering for accurate material perception

K. Bala, Cornell Univ. (United States)

Computer graphics rendering is used to simulate the appearance of objects and materials in a wide range of applications. Designers and manufacturers need to rely entirely on rendered images to correctly recognize and differentiate between different types of fabrics, paint finishes, plastics, and metals. Thus, the rendering algorithms producing these images must be predictive of the appearance of materials.

While progress has been made on predictive rendering for metals and plastic, predictive rendering of cloth and fabrics remains particularly challenging. The structure of cloth complicates the reflectance of light. Further, irregularities in the structure of cloth are difficult to model, but are visually important. For example, the structure of the pile in velvet affects its fuzzy appearance, and its characteristic highlights at grazing angles.

For cloth, predictive rendering requires the acquisition of detailed models representing the macro-scale and meso-scale structure of fibers and yarns, better optical models, and volumetric light transport algorithms that simulate the scattering of light within the volume of structured fibers and yarns. We have developed new scattering models and light transport algorithms that create highly realistic images of cloth. But they require detailed 3D structural models as input.

We introduce an entirely different approach to building volume appearance models for cloth, by using X-ray computed tomography (CT) to acquire detailed geometric structure. For example, for woven cloth, these scans automatically capture the structure of the cloth with a detailed view of the interlaced yarns and their component fibers, showing exactly how the fibers are oriented and how the yarns are positioned. The density and orientation information from these CT scans is augmented with optical properties (albedo and gloss) inferred from a reference photograph. This powerful approach can easily produce volume appearance models with extreme detail in closeup, and at larger scales, with the distinctive textures and highlights of a range of very different fabrics like satin, velvet, and wool.

Coupling detailed structural information with volumetric light transport algorithms is a first step towards predictive rendering for cloth.

8291-07, Session 2

From color to appearance in the real world

F. Lamy, X-Rite, Inc. (United States)

Xrite's declared ambition is to create a digital ecosystem for appearance.; a daunting challenge that has many dimensions and has proven so massive that all previous attempts efforts have failed so far. After having invested 3 years in exploring the problem, we can now deliver the first elements of answers and the practical path to tackle this massive undertaking.

We will explore the practical implications of the intermediation of 2 stages between color and full appearance : extended color and augmented color, and how these steps, rooted in the realities of the ecosystem they serve are constituting vectors and enablers of a more effective transition.

We will survey the road map implication in the design for measurement and capture instrumentations , packaging digital formats and delivery infrastructure as well as rendering and display devices that will enable true value creation built on appearance attributes.

8291-08, Session 2

Towards material literacy: making perception and knowledge of materials explicit through visual and verbal documentation of material engagement

A. Lehmann, Utrecht Univ. (Netherlands)

No abstract available

8291-09, Session 2

Mixing material modes

S. C. Pont, Technische Univ. Delft (Netherlands)

No abstract available

8291-10, Session 2

Tangible display systems: bringing virtual objects into the real world

J. A. Ferwerda, Rochester Institute of Technology (United States)

No abstract available

8291-11, Session 3

Quality estimation for images and video with different spatial resolutions

A. M. Demirtas, Univ. of California, Irvine (United States); A. R. Reibman, AT&T Labs. Research (United States)

Full-reference (FR) quality estimators (QEs) for images and video are typically designed assuming that the displayed, degraded image has the same spatial resolution as the original, reference image. No-reference (NR) QEs use no knowledge about the reference image to assess quality of the displayed image. However, in many practical systems, a reference image may be available that has a different spatial resolution than the displayed image.

In this paper, we explore objective quality estimation when the displayed image to be evaluated has a different spatial resolution than the reference image. We begin by identifying a range of potential weaknesses that might be present in a QE designed for this situation. Then, we create pairs of images with potential False Ties, in which a QE estimates the two images have equal quality while viewers disagree. Armed with the knowledge of potential flaws, we then present effective design strategies of QEs for this situation.

8291-12, Session 3

Automatic parameter prediction for image denoising algorithms using perceptual quality features

A. Mittal, A. K. Moorthy, A. C. Bovik, The Univ. of Texas at Austin (United States)

We propose a perception based approach to create a blind image denoising algorithm, by combining blind parameter estimation with a state-of-the-art denoising algorithm. Our approach to blind parameter estimation is learning-based, where input parameter is estimated using statistical properties of natural scenes.The features are based on marginal statistics of local normalized luminance signals and distribution of pairwise products of adjacent normalized luminance signals. We show that our blind parameter estimation procedure "statistically" gives higher quality of denoised images than the baseline on an exhaustive set of 3000 images with a wide range of noise variance values.

8291-13, Session 3

Viewer preferences for classes of noise removal algorithms for high-definition content

S. G. Deshpande, Sharp Labs. of America, Inc. (United States)

Perceived video quality studies were performed on a number of key classes of noise removal algorithms to determine viewer preference. The noise removal algorithm classes represent increase in complexity from linear filter to nonlinear filter to adaptive filter to spatio-temporal filter. The subjective results quantify the perceived quality improvements that can be obtained with increasing complexity. The specific algorithm classes tested include: linear spatial one channel filter, nonlinear spatial two-channel filter, adaptive nonlinear spatial filter, multi-frame spatio-temporal adaptive filter. All algorithms were applied on full HD (1080P) content. Our subjective results show that spatio-temporal (multi-frame) noise removal algorithm performs best amongst the various algorithm classes. The spatio-temporal algorithm improvement compared to original video sequences is statistically significant. On the average, noise-removed video sequences are preferred over original (noisy) video sequences. The Adaptive bilateral and non-adaptive bilateral two channel noise removal algorithms perform similarly on the average thus suggesting that a non-adaptive parameter tuned algorithm may be adequate.

8291-14, Session 3

Image quality assessment in the low-quality regime

G. O. Pinto, S. S. Hemami, Cornell Univ. (United States)

Image Quality Assessment (IQA) algorithms are traditionally designed to operate over the entire quality range, from very low quality images to visually lossless images. However, many, if not all, imaging applications are not designed to operate over the entire range; for example, digital camera image outputs do not produce extremely compressed images, and are generally only of extremely low quality when operator error has occurred. As such, designing and even optimizing a quality estimator for the entire quality range may unreasonably limit the estimator's

performance in more realistic ranges of operation. As this paper will show, the performances of traditional quality estimators over the entire quality range differ greatly from the performances on low quality images. For example, in the LIVE database, the SSIM estimator has a Spearman rank-order correlation (SROC) coefficient of 0.91 for the entire database, but it has a coefficient of 0.50 in the low quality regime.

8291-15, Session 4

The question of simultaneity in multisensory integration

M. E. McCourt, L. Leone, North Dakota State Univ. (United States)

Early reports of audiovisual (AV) multisensory integration (MI) indicated that unisensory stimuli must cause simultaneous physiological responses to produce decreases in mean reaction time (RT). When unisensory RTs were unequal the onset of the stimulus eliciting the faster RT had to be delayed relative to the onset of the stimulus eliciting the slower RT, for MI to occur. The "temporal rule" states that MI depends on the temporal proximity of the unisensory stimuli, the neural responses to which must fall within a window of integration. Ecological validity demands that MI should occur only for simultaneous events (which may give rise to non-simultaneous neural activations). On the other hand, instances of neural simultaneity which are unrelated to unitary multisensory events must be rejected as spurious. We measured AV MI under fully dark-adapted conditions for visual stimuli that were either weak (scotopic; long RT) or strong (photopic; short RT). Auditory stimulus intensity was constant. Despite a 190 ms increase in mean RT to the scotopic stimulus, facilitative MI in both conditions occurred exclusively at an AV SOA of 0 ms. The mechanisms whereby large variations in response latency arising from changes in stimulus intensity are discounted are considered.

8291-16, Session 4

Multisensory integration deficits in children with autism spectrum disorders: a behavioral and ERP investigation

C. D. Saron, Y. Takarae, M. Beransky, D. M. Horton, A. Stark, A. Schneider, F. Vieira, S. M. Rivera, Univ. of California, Davis (United States)

Multisensory integration (MSI), the combination of various senses to form a single integrated experience of the world, is essential to everyday life. It is now widely claimed by clinicians and researchers alike that deficits in sensory processing, including MSI, are part of the phenotype of autism spectrum disorders (ASD). This study examined MSI in children (10-14 yrs.) with ASD and age-matched typically developing (TD) children using of dense-channel array event-related potentials (ERPs) elicited in response to visual (V), auditory (A), and somatosensory (S) stimuli delivered alone or in simultaneous combination. The task was the detection of all stimulus events (simple reaction time (RT)). Differences between multisensory (MS) and the summed responses of unisensory stimuli (US) were used to estimate MSI. Results demonstrated a lack of RT facilitation for auditory/tactile multisensory stimuli for the ASD group (N=9/group). Unisensory ERPs were generally similar between the TD and ASD groups. ERPs to MS stimuli differed between groups. The TD pattern typically showed decreased early activations in MS conditions compared with summed US data. These data suggest specific deficits in MSI of children with ASD and are consistent with the view that low-level sensory processing and integration contribute to the complex phenotype of ASD.

8291-17, Session 4

The spatiotopic 'visual' cortex of the blind

L. T. Likova, The Smith-Kettlewell Eye Research Institute (United States)

Conference 8291

Visual cortex activity in the blind has been shown in sensory tasks. Can it be activated in memory tasks? Our recent results in short-term blindfolded subjects imply that human primary visual cortex (V1) may operate as a modality-independent 'sketchpad' for working memory (Likova, 2010). We now raise the question of whether under long-term visual deprivation V1 is also employed in non-visual memory, in particular in congenitally blind subjects, who have never had visual stimulation to guide the development of the visual area organization. The outcome has implications for an emerging reconceptualization of the principles of brain architecture and its reorganization under sensory deprivation. Methods: We used a novel fMRI drawing paradigm in congenitally and late blind, compared with sighted and blindfolded subjects in three conditions of 20s duration, separated by 20s rest-intervals, (i) Tactile Exploration: raised-line images explored and memorized; (ii) Tactile Memory Drawing: drawing the explored image from memory; (iii) Scribble: mindless drawing movements with no memory component. Results and Conclusions: V1 was strongly activated for Tactile Memory Drawing and Tactile Exploration in these totally blind subjects. Remarkably, even in the memory task, the mapping of V1 activation largely corresponded to the angular projection of the tactile stimuli relative to the ego-center (i.e., the effective visual angle); beyond this projective boundary, peripheral V1 signals were dramatically reduced or suppressed. The matching extent of the activation in the congenitally blind rules out vision-based explanatory mechanisms, and supports the more radical idea of V1 as an modality-independent 'sketchpad', whose mapping scales to the projective dimensions of objects explored in the peripersonal space. In higher occipital areas the differential pattern of engagement across the subject categories suggests progressive neuroplastic reorganization as visual deprivation is prolonged, in accord with an evolving cross-modal utilization of the occipital cortex to maximize its inherent functional potential.

8291-18, Session 4

Acoustic-tactile rendering of visual information

P. M. Silva, T. N. Pappas, Northwestern Univ. (United States); J. Atkins, J. E. West, The Johns Hopkins Univ. (United States)

No abstract available

8291-19, Session 5A

Apparent stereo: the Cornsweet illusion can enhance perceived depth

P. Didyk, Max-Planck-Institut für Informatik (Germany); T. Ritschel, E. Eisemann, Telecom ParisTech (France); K. Myszkowski, H. Seidel, Max-Planck-Institut für Informatik (Germany)

It is both a technical and an artistic challenge to depict three-dimensional space using a flat two-dimensional medium. Fitting within the limits of a given display technology and at the same time achieving comfortable viewing can require modifications to stereo content, e. g., to flatten or even remove binocular disparity. However, in such scenarios crucial depth details might get lost. Recently, a backward-compatible stereo technique was proposed, which builds upon the Craik-O'Brien-Cornsweet effect - a visual illusion, which was applied to brightness in CG before. The approach replaces disparity by Cornsweet profiles in disparity where depth discontinuities occur. This allows for the reduction of the overall disparity range while still conveying a comfortable stereo experience. The present work extends the previous idea by showing that Cornsweet profiles can also be used to enhance the 3D impression in regions where the disparity range was compressed. A user study measures the performance of backward-compatible stereo as well as our disparity enhancement.

8291-20, Session 5A

Perceived depth of multi parallel, overlapping, transparent, stereoscopic surfaces

S. Aida, K. Shimono, Tokyo Univ. of Marine Science and Technology (Japan); W. J. Tam, Communications Research Ctr. Canada (Canada)

The geometric relational expression of horizontal disparity, viewing distance, and depth magnitude between objects in stereopsis suggests that, for a given viewing distance, the magnitude of perceived depth of objects would be the same as long as the disparity magnitudes are the same. However, we found that this is not necessarily the case for random dot stereograms depicting two or more parallel-transparent-overlapping-stereoscopic-surfaces (POTS). Data from two experiments indicated that, when the stimulus size is relatively large (e.g., 13 x 20 arc deg), the magnitude of reproduced depth between two stereo-surfaces of a two-POTS configuration can be larger than that between an identical pair of stereo-surfaces but with an additional stereo-surface located between the pair. In follow-up experiments, dot density, the size of the random dot stereograms and the number of stereo-surfaces lying between the two outermost stereo-surfaces were manipulated. Overall, the results indicate a difference of approximately 10% in the reproduced depth, depending on whether there is an additional stereo-surface between the two surfaces of a two-POTS configuration or not. The results are discussed in terms of global stereopsis which "operates" for relatively large stimulus sizes (such as 13 x 20 arc deg).

8291-21, Session 5A

Diagnosing perceptual distortion present in group stereoscope viewing

M. M. Burton, B. B. Pollock, J. W. Kelly, S. B. Gilbert, E. Winer, Iowa State Univ. (United States); J. de la Cruz, U.S. Army RDECOM/STTC (United States)

Stereoscopic virtual environments are frequently employed to immerse viewers in virtual reality visualizations, simulations, and training systems. Stereo systems render the scene graph from the perspective of a single person or static viewpoint which can be defined as a leader, while others see inaccurate and distorted views from a follower position. Greater understanding of distortion from the follower perspective can lead to more effective multi-view solutions and configurations of leader-follower systems.

An experiment at Iowa State University was conducted to gain insight into the perceptual distortion present in leader-follower systems where multiple users view the same stereoscopic, projection-based virtual environment. The experiment was comprised of tasks reporting quantities of angles in degrees from positions parallel to the screen and reporting the depth dimension of a rectangle at positions centered and perpendicular to the screen. Results from this study indicated a relationship of angular distortion with leader-follower offsets parallel to the screen and distance to the screen affects the quantity of depth distortion in addition to perpendicular offsets from the rendered viewpoint. Further work will attempt to understand the effects of this distortion on group collaboration and further develop relationships of distortion and leader-follower offsets.

8291-34, Session 5A

Visual discomfort and the timing of vergence-accommodation conflicts

J. Kim, D. Kane, M. S. Banks, Univ. of California, Berkeley (United States)

Stereo 3D (S3D) displays cause visual discomfort in many viewers, so it is very important to understand the causes of the discomfort so one can minimize it. We know that the vergence-accommodation conflict

causes visual discomfort and eventually visual fatigue. Vergence and accommodation both have phasic and tonic components that adapt at different rates. To determine if the different adaptation rates in phasic and tonic components affect visual discomfort, we varied the temporal properties of the vergence-accommodation conflict and assessed discomfort.

Observers viewed random-dot stereograms in a novel volumetric display that allows us to independently manipulate the stimulus to vergence and the stimulus to accommodation. There were two main conditions: natural viewing (cues-consistent) and S3D viewing (cues-inconsistent). The stimuli to vergence and accommodation were varied at different three different rates, presented in different sessions.

We found that observers experienced worse symptoms in the S3D viewing than in the natural viewing. We also found that more frequent changes in the stimulus to vergence yielded worse symptoms. Thus, the time-average magnitude of the vergence-accommodation conflict is not the sole predictor of the amount of discomfort. It also matters how the conflict is distributed over time.

8291-22, Session 5B

Measuring 3D discomfort from vertical and torsional disparities in natural images

C. W. Tyler, L. T. Likova, The Smith-Kettlewell Eye Research Institute (United States); K. Atanassov, V. Ramachandra, S. Goma, Qualcomm (United States)

The two major aspects of camera misalignment that cause visual discomfort when viewing images on a 3D display are vertical and torsional disparities. The goal of the study was to determine the discomfort ranges for the kinds of natural image that people are likely to take with 3D cameras. The results showed that, as a rule of thumb, vertical disparity between the two eyes needs to be held below about 15 arcmin, or a quarter of a degree, to be within an acceptable range of discomfort, while human vision is relatively insensitive to torsional disparities below about 30 deg.

8291-35, Session 5B

Visual fatigue versus eye-movements

C. Vienne, L. Blondé, D. Doyen, Technicolor S.A. (France)

Observing 3D content on a cinema or TV screen potentially generates fatigue. In research, experience of visual symptoms following the observation of stereo-content is usually assessed thanks to questionnaires and subjective reports. We attempted to explore the occurrence of visual fatigue using more objective methods, namely by using binocular eye-tracking and psychophysics. A main objective was to study the emergence of visual fatigue in relation with eye-movement knowing the stimulation of the oculomotor system and its response. We designed an experiment in which participants were asked to perform a repeated vergence effort task, just followed by a 3D space perception task. Participants' eye movements were recorded during the whole session using an eye-tracking system. The analysis revealed that the perception of 3D shape was gradually affected by the intensity of the vergence effort task. The effect on stereo-estimation was actually due to the occurrence of visual fatigue, 3D objects are perceived flatter. Results on the subjective reports of SSQ revealed that oculomotor factors were predominant in the visual symptoms. In addition, some effects and correlations on the micro-saccadic rate were obtained. It offers a perspective to characterize objectively visual fatigue when watching 3D content on stereoscopic displays.

8291-36, Session 5B

Visual comfort: stereoscopic objects moving in the horizontal and mid-sagittal planes

W. J. Tam, F. Speranza, C. Vázquez, R. Renaud, Communications Research Ctr. Canada (Canada); N. Hur, Electronics and Telecommunications Research Institute (Korea, Republic of)

The present study examines the effect of motion of stereoscopic objects on visual comfort using computer-generated video sequences based on a parallel-camera setting. In the first experiment, the visual comfort of a set of stereoscopic objects moving in the z-plane with different disparities and velocities is examined. Three types of movement in the z-plane were included: both uncrossed and crossed disparities; crossed only; and uncrossed only. In a second experiment, the visual comfort of a set of stereoscopic objects moving at a given depth in the horizontal plane is examined. Also in this case, three types of movement in the z-plane were included: both uncrossed and crossed disparities; crossed only; and uncrossed only. In both experiments, the movements (along the z-plane for Exp. 1 and along the x-plane for Exp. 2) involved three levels of velocity and five levels of disparity. Preliminary results appear to confirm that visual comfort is influenced more by object velocity than by disparity magnitude (for the range of magnitudes used); an interaction of velocity and disparity is also apparent. Therefore, the preliminary results appear to confirm that changes in disparity over time are a significant contributor to visual comfort of stereoscopic video images; the effect appears to be present also when the changes are confined within the generally accepted comfort zone.

8291-37, Session 5B

Visual discomfort with stereo 3D displays when the head is not upright

D. Kane, R. Held, M. Banks, Univ. of California, Berkeley (United States)

Stereo 3D (S3D) displays can cause visual discomfort. In part, discomfort is due to the well-known vergence-accommodation conflict. Additionally, discomfort may be caused by viewing S3D displays with the head rotated to the side (roll rotation). Head roll converts horizontal disparity on the screen into vertical disparity on the retina that in turn should elicit vertical vergence eye movements to minimize vertical disparity. Because such movements are small in natural viewing, the larger ones required with head roll are likely to cause discomfort. We investigated this by simulating head roll in a S3D display. We kept the observers' head upright, but rotated the stimulus. The stimulus was a random-dot stereogram depicting sinusoidal depth corrugations. It contained either no variation in overall disparity or a variation in overall disparity over time. Observers indicated the amount of visual discomfort experienced when viewing these stimuli. They reported worse symptoms in the simulated head-roll conditions and when the disparity pedestal changed by a greater amount. We attribute the increased symptoms to the vertical vergence eye movements elicited by head roll.

8291-50, Poster Session

Tracking of aging process by multiple 3D scans analysis

E. Bunsch, The Wilanów Palace Museum (Poland); R. Sitnik, Warsaw Univ. of Technology (Poland)

Currently, a lot of different 3D scanning devices are used for 3D acquisition of art artifact surface shape and color. Each of them has different technical parameters starting from measurement principle (structured light, laser triangulation, interferometry, holography) and ending on parameters like measurement volume size, spatial resolution and precision of output data and color information. Some of the 3D scanners can grab additional information like surface normal vectors, BRDF distribution, multispectral color. In this paper, we plan to present results of the measurements with selected sampling densities together with discussion of the problem of recognition and assessment of the aging process. We focus our interest on features that are important for the art conservators to define state of preservation of the object as well as to assess changes on the surface from last and previous measurement. Also different materials and finishing techniques requires different algorithms for detection and localization of aging changes.

Conference 8291

In this paper we consider exemplary stone samples to visualize what object features can be detected and tracked during aging process. The changes in sandstone surface shape, affected by salt weathering, will be presented as well as possibilities of identification of surface degradation on real object (garden relief made in sandstone).

8291-51, Poster Session

Aesthetics and entropy: optimization of brightness distribution

M. R. V. Sahyun, Consultant (United States)

The purpose of this work is to suggest directions for utilizing image statistics to guide optimization of brightness distributions, towards a goal of complete systematization of image processing to achieve a purely aesthetic objective, whether by application of human or machine intelligence, with entropy as a response metric. We start with a survey of classic pictorial photographs, proceed to a heuristic theoretical treatment of the brightness distribution function, and follow with several pictorial illustrations of the proposed approach to entropy-guided brightness optimization.

8291-52, Poster Session

A novel visualization tool for art history: a method for automated colorization of black and white archival photographs of works of art

S. Tsaftaris, Institute for Advanced Studies Lucca (Italy) and Northwestern Univ. (United States); F. Casadio, K. Lister, I. Fiedler, G. Gautier, The Art Institute of Chicago (United States); J. Andral, Musée Picasso (France); S. D'Alessandro, The Art Institute of Chicago (United States); A. Katsaggelos, Northwestern Univ. (United States)

Colorizing archival B&W photographs of works of art, helps conservators and art historians visualize artists' intention as expressed in their working process. While conservators typically use commercial software to colorize images by applying flat areas of color in a somewhat subjective way, the proposed method allows spreading objective color information in a computer-controlled fashion, preserving brushwork and shading. This new colorization methodology uses predefined hints recovered from remnants of previous colors still visible in the present state, or derived from observations of paint cross sections. By registering the B&W photograph and the present state and local correlation analysis we automatically recover color hints. These hints are complemented by expert input. We assume that pixels with similar luminance values will have similar chrominance values. We pose and solve a linear optimization problem with respect to the pixels of unknown color in a neighborhood and pixels of known color. Our methodology has been used to colorize historical photographs of Bathers by a River, 1909-1917, by Henri Matisse (Art Institute of Chicago 1953.158), and La Joie De Vivre, 1946, by Pablo Picasso (Musée Picasso, Antibes, 1946.1.4), providing new tools to art historians to unravel the alchemy of creation of these masterpieces.

8291-53, Poster Session

PHOG analysis of self-similarity in aesthetic images

S. A. Amirshahi, M. Koch, J. Denzler, C. Redies, Friedrich-Schiller-Univ. Jena (Germany)

In recent years, there have been efforts in defining the statistical properties of aesthetic photographs and artworks using computer vision techniques. However, it is still an open question how to distinguish aesthetic from non-aesthetic images with a high recognition rate. This is possibly because aesthetic perception is influenced also by a large number of cultural variables. Nevertheless, the search

for statistical properties of aesthetic images has not been futile. For example, we have shown that the radially averaged power spectrum of monochrome artworks of Western and Eastern provenance falls off according to a power law with increasing spatial frequency (1/f2 characteristics). This finding implies that this particular subset of artworks possesses a Fourier power spectrum that is self-similar across different scales of spatial resolution. Other types of aesthetic images, such as cartoons, comics and mangas also display this type of self-similarity, as do photographs of complex natural scenes. Since the human visual system is adapted to encode images of natural scenes in a particular efficient way, we have argued that artists imitate these statistics in their artworks. In support of this notion, we presented results that artists portrait human faces with the self-similar Fourier statistics of complex natural scenes although real-world photographs of faces are not self-similar. In view of these previous findings, we investigated other statistical measures of self-similarity to characterize aesthetic and non-aesthetic images. In the present work, we propose a novel measure of self-similarity that is based on the Pyramid Histogram of Oriented Gradients (PHOG). For every image, we first calculate PHOG up to pyramid level 3. The similarity between the histograms of each section at a particular level is then calculated to the parent section at the previous level (or to the histogram at the ground level). The proposed approach is tested on datasets of aesthetic and non-aesthetic categories of monochrome images. The aesthetic image datasets comprise a large variety of artworks of Western provenance. Other man-made aesthetically pleasing images, such as comics, cartoons and mangas, were also studied. For comparison, a database of natural scene photographs is used, as well as datasets of photographs of plants, simple objects and faces that are in general of low aesthetic value. As expected, natural scenes exhibit the highest degree of PHOG self-similarity. Images of artworks also show high selfsimilarity values, followed by cartoons, comics and mangas. On average, other (non-aesthetic) image categories are less self-similar in the PHOG analysis. A measure of scale-invariant self-similarity (PHOG) allows a good separation of the different aesthetic and non-aesthetic image categories. Our results provide further support for the notion that, like complex natural scenes, images of artworks display a higher degree of self-similarity across different scales of resolution than other image categories. Whether the high degree of self-similarity is the basis for the perception of beauty in both complex natural scenery and artworks remains to be investigated.

8291-54, Poster Session

Influence of the source contact and encoding configuration on the preceived quality for scalable video coding

Y. Pitrey, Univ. de Nantes (France); M. Barkowsky, Polytech' Nantes (France); R. Pepion, Univ. de Nantes (France); P. Le Callet, Polytech' Nantes (France)

In video coding, it is commonly accepted that the encoding paramaters such as the quantization step-size have an influence on the perceived quality. When dealing with Scalable Video Coding (SVC), the parameters used to encode each layer logically have an influence on the overall perceived quality. It is also commonly accepted that using given encoding parameters, the perceived quality does not change significantly according to the encoded source content. In this paper, we evaluate the impact of both SVC coding artifacts and source contents on the quality perceived by human observers. We exploit the outcomes of two subjective experiments designed and conducted under standard conditions in order to provide reliable results. The two experiments are aligned on a common scale using a set of shared processed video sequences, resulting in a database containing the subjective scores for 60 different sources combined with 20 SVC scenarios. We analyse the performance of several source descriptors in modeling the relative behaviour of a given source content when compared to the average of other source contents.

Conference 8291

8291-55, Poster Session

Evaluation of desktop interface displays for 360-degree video

S. B. Gilbert, W. Boonsuk, J. W. Kelly, Iowa State Univ. (United States)

A 360-degree video becomes necessary in applications ranging from surveillance to virtual reality. This research focuses on developing an interface for a system such as mobile surveillance that integrates 360-degree video feeds for remote navigation and observation in unfamiliar environments. An experiment evaluated the effectiveness of three 360-degree view user interfaces to identify the necessary display characteristics that allow observers to correctly interpret 360-degree video images displayed on a desktop screen. Video feeds were simulated using a game engine. Interfaces were compared, based on spatial cognition and participants' performance in finding target objects. Results suggest that 1) correct perception of direction within a 360-degree display is not correlated with a correct understanding of spatial relationships within the observed environment, 2) visual boundaries in the interface may increase spatial understanding, and 3) increased video gaming experience may be correlated with better spatial understanding of an environment observed in 360-degrees. This research will assist designers of 360-degree video systems to design optimal user interface for navigation and observation of remote environments.

8291-56, Poster Session

An evaluation of different setups for simulating lighting characteristics

B. A. Salters, M. J. Murdoch, D. Sekulovksi, P. Seuntiens, S. Chen, Philips Research Nederland B.V. (Netherlands)

The advance of LED technology enables a new range of luminaire designs, which previously were not possible. Evaluating such new designs has traditionally been done using actual prototypes, in a real environment. The iterations needed to build, verify, and improve luminaire designs incur substantial costs, and slow down the design process. A more attractive way is to evaluate designs using simulations, as they can be made cheaper and quicker, for a wider variety of prototypes. However, the value of such simulations is determined by how closely they predict the outcome of actual perception experiments.

In this paper, we discuss an actual perception experiment on several lighting settings, in a normal office environment. The same office environment also has been modeled at various detail levels, and photo-realistic renderings have been created of these models. These renderings were subsequently processed using various tone-mapping operators, and shown on different displays. The total chain of software and hardware can be considered a simulation setup, and we have executed several perception experiments on different setups. Finally, a comparison is made between the results obtained in reality, and the results obtained via simulation. Two possible approaches are discussed to rank the quality of different simulation setups for the purpose of simulating lighting characteristics.

8291-57, Poster Session

Biological visual attention guided automatic image segmentation with application in satellite imaging

M. I. Sina, A. Cretu, P. Payeur, Univ. of Ottawa (Canada)

Taking inspiration from the significantly superior performance of humans to extract and interpret visual information, the exploitation of biological visual mechanisms can contribute to the improvement of the performance of computational image processing systems. Computational models of visual attention have already been shown to significantly improve the speed of scene understanding by attending only the regions of interest, while distributing the resources where they are required. However, there are only few attention-based computational systems that have been used in practical applications dealing with real data and up to now, none of the computational attention models was demonstrated to work under a wide range of image content, characteristics and scales such as those encountered in satellite imaging. This paper outlines some of the difficulties that the current generation of visual attention-inspired models encounter when dealing with satellite images. It then proposes a novel algorithm for automatic image segmentation and regions of interest search that combines elements of human visual attention with Legendre moments applied on the probability density function of color histograms. The experimental results demonstrate that the proposed approach obtains better results than one of the most evolved current computational attention model proposed in the literature.

8291-58, Poster Session

A neurobiologically-based, two-stage model for human color vision

C. Q. Wu, Stanford Univ. (United States)

Currently, there are two dominant theories for human color vision: Young-Helmholtz's trichromatic theory and Hering's opponent-color theory. It is now widely accepted that the trichromatic theory holds true for retinal color processing whereas the opponent-color theory works for cortical color processing--this conception has become the "Standard Model" for human color vision. In the present paper I attempt to achieve three goals: First, to demonstrate that the opponent-color theory is fundamentally untenable, based on both logical and empirical grounds; second, to resurrect a two-stage trichromatic model, in which both retinal and cortical color processing are trichromatic, proposed by A. Rollett and W. McDougall more than a century ago; and third, to map the cortical color processing stage in this model to layer 4C within the primary visual cortex of the human brain.

8291-59, Poster Session

The oscillatory activities and its synchronization in auditory-visual integration as revealed by event-related potentials to bimodal stimuli

J. Guo, Beijing Normal Univ. (China); P. Xu, General Hospital Armed Police Forces (China); X. Zhao, Beijing Normal Univ. (China)

Neural mechanism of auditory-visual speech integration is always a hot study of multi-modal perception. The articulation conveys speech information that helps detect and disambiguate the auditory speech. As important characteristic of EEG, oscillations and its synchronization have been applied to cognition research more and more. This study analyzed the EEG data acquired by unimodal and bimodal stimuli using time frequency and phase synchrony approach, investigated the oscillatory activities and its synchrony modes behind evoked potential during auditory-visual integration, in order to reveal the inherent neural integration mechanism under these modes. It was found that beta activity and its synchronization had relationship with gesture N1-P2, which happened in the earlier stage of speech coding to pronouncing action. Alpha oscillation and its synchronization related with auditory N1-P2 might be mainly responsible for auditory speech process caused by anticipation from gesture to sound feature. The visual gesture changing enhanced the interaction of auditory brain regions. These results provided explanations to the power and connectivity change of event-evoked oscillatory activities which matched ERPs during auditory-visual speech integration.

8291-63, Poster Session

Quality assessment of images illuminated by dim LCD backlight

Conference 8291

T. Huang, H. H. Chen, National Taiwan Univ. (Taiwan)

We consider the kind of dimmed images that are seen on a liquid crystal display (LCD) when the backlight of the LCD is set to a low power level. A major characteristic of this energy saving mode is that the perceived image quality decreases with the intensity of the backlight. In particular, some image regions may appear so dark that they become non-perceptible to human eye when the image is illuminated with very dim backlight. Ignoring this very fact and directly applying an image quality assessment metric to the entire image may produce results inconsistent with human evaluation of the image. In this paper, we propose a method to fix the problem. The proposed method works as a precursor of image quality assessment. Specifically, given an image and the backlight intensity level of the LCD on which the image is to be displayed, the method automatically classifies the pixels of an image to perceptible and non-perceptible pixels according to the backlight intensity level and excludes the non-perceptible pixels from quality assessment.

8291-64, Poster Session

Parallax scanning methods for stereoscopic three-dimensional imaging

C. A. Mayhew, C. M. Mayhew, Vision III Imaging, Inc. (United States)

Under certain circumstances, conventional stereoscopic imagery is subject to being misinterpreted. Stereo perception created from two static horizontally separated views can create a "cut out" 2D appearance for objects at various planes of depth. The subject volume looks three-dimensional, but the objects themselves appear flat. This is especially true if the images are captured using small disparities.

One potential explanation for this effect is that although three-dimensional perception comes primarily from binocular vision, a human's gaze (the direction and orientation of a person's eyes with respect to their environment) and head motion contribute additional information. The absence of this information may be the reason that certain stereoscopic imagery appears "odd" and unrealistic

Another contributing factor may be the absence of vertical disparity information in traditional stereoscopy display. Current psychophysical and physiological evidence suggests that vertical disparities influence the perception of three-dimensional depth and object shape, but little is known about the perceptual mechanisms that support these processes. Whatever the specific mechanisms are, it is clear that the motion and gaze direction of the eyes contribute significantly to the process of three-dimensional perception. Conventional stereoscopic image capture goes to great lengths to avoid and/or eliminate any vertical parallax differences in the images.

A third factor may be the stereoscopic production trend of capturing images with disparities that are 50% or less than the average human interocular of 65 mm. This trend is fueled, in part, by a desire to keep the images in a comfortable range for the general viewing public. However, less disparity produces less 3D effect and also leads to a flattening of background scene elements.

Recently, Parallax Scanning technologies have been introduced, which provide (1) a scanning methodology, (2) incorporate vertical disparity, and (3) produce stereo images with substantially smaller disparities than the human interocular distances. To test whether these three features would improve the realism and reduce the posterization of stereo images, we have applied Parallax Scanning (PS) technologies to commercial stereoscopic digital cinema productions, and have tested the results with a panel of stereo experts.

The addition of parallax scan information into the left and right images improves the overall three-dimensional perception. Viewer stereopsis is shown to benefit from the additional PS sub-process information by generating a more unified three-dimensional perception.

PS is accomplished optically through the use of a digital parallax scanning (DPS) iris mechanism. A PS lens is designed to move its iris center from 0 to 4 mm off the lens center along a circular path at frequencies from 0 to 15 Hz. Two parallax scanning lenses can be configured for stereoscopic image capture.

To test the visual effect, we conducted the following experiment. We supplied DPS equipped lenses to a professional stereographer who used them to shoot a 3D test scene at varying scanning parameters including a no scan control (0 mm offset). The results were presented to a group of fourteen stereographic professionals and academics. Eleven in the group reported significantly improved shape definition in the PS imagery when compared to the control. Two reported moderate image improvement in the PS imagery and one reported no difference between the control and the PS imagery. Subsequent screenings of the test imagery to other groups produced similar viewer results.

These informal experiments show that the addition of PS information into the left and right image capture improves the overall perception of three-dimensionality in most viewers. Parallax scanning significantly increases the set of tools available for 3D storytelling while at the same time presenting imagery that is easy and pleasant to view.

8291-65, Poster Session

Reduced reference image quality assessment via sub-image similarity-based redundancy measurement

X. Mou, W. Xue, Xi'an Jiaotong Univ. (China); L. Zhang, The Hong Kong Polytechnic Univ. (Hong Kong, China)

The reduced reference (RR) image quality assessment (IQA) has been attracting much attention from researchers for its loyalty to human perception and flexibility in practice. A promising RR metric should be able to predict the perceptual quality of an image accurately while using as few features as possible. In this paper, a novel RR metric is presented, whose novelty lies in two aspects. Firstly, it measures the image redundancy by calculating the socalled Sub-image Similarity (SIS), and the image quality is measured by comparing the SIS between the reference image and the test image. Secondly, the SIS is computed by the ratio of NSE (Non-shift Edge) between the pairs of sub-images. Experiments on two IQA databases (i.e. LIVE and CSIQ databases) show that by using only 6 features, the proposed metric can work very well with high correlations between the subjective and objective scores. In particular, it works consistently well across different distortion types.

8291-66, Poster Session

Color impact in visual attention deployment considering emotional images

C. Chamaret, Technicolor S.A. (France)

In this study, we propose to study the color impact as well as the emotion aspect of pictures on the visual attention deployment. As expected the eye fixations of both color and black and white images were highly correlated leading to the question of the integration of such cues in the design of visual attention model. This last point was not a pending issue in the literature since it is agreed that color plays a role in visual attention. The compromise is probably the integration of color component within the visual attention models, but weighted by the pre-analysis of image which may determine a priori the potential impact of color within the presented pictures.

8291-23, Session 6

On the development of expertise in interpreting medical images

E. Krupinsky, The Univ. of Arizona (United States)

No abstract available

8291-24, Session 6

Modeling observer performance for optimizing image acquisition and processing

Conference 8291

C. K. Abbey, M. Eckstein, Univ. of California, Santa Barbara (United States)

No abstract available

8291-25, Session 6

Evaluation of HVS models in the application of medical image quality assessment

L. Zhang, C. Cavaro-Ménard, Univ. of Angers (France); P. Le Callet, Polytech' Nantes (France)

In this study, four of the most widely used Human Visual System (HVS) models are applied on Magnetic Resonance (MR) images for signal detection task. Their performances are evaluated against gold standard derived from radiologists' decisions. The task-based image quality assessment requires taking into account the human perception specificities, for which various HVS models have been proposed. Few works were conducted however to evaluate and compare the suitability of these models with respect to the assessment of medical image qualities. Here we propose to score the performance of each HVS model using the AUC and its variance estimates as the figure of merit. The contribution of this work is twofold: firstly the application of MRMC (multiple-reader, multiple-case) estimates independently of the HVS model's output range, secondly the use of radiologists' consensus as gold standard so that the estimated AUC measures the distance between the HVS model and the radiologist perception.

8291-26, Session 6

Perceptual challenges to computer-aided diagnosis

Y. Jang, The Univ. of Chicago (United States)

No abstract available

8291-27, Session 6

Satisfaction of search experiments in advanced imaging

K. S. Berbaum, The Univ. of Iowa Hospitals and Clinics (United States)

No abstract available

8291-28, Session 6

Integrating human- and computer-based approaches to feature extraction and analysis

B. E. Rogowitz, Visual Perspectives (United States); A. Goodman, Harvard-Smithsonian Ctr. for Astrophysics (United States)

Integrating Human- and Computer-based Approaches to Feature Extraction and Analysis

A major goal of imaging systems is to help doctors, scientists, engineers, and analysts identify patterns and features in complex data. Typically, this is accomplished by computer-driven algorithms that extract features automatically, or by visualization systems that allow the analyst to manipulate visual representations to discover features interactively. Although automatic feature-extraction algorithms are often directed by human observation, and human pattern recognition is often supported by algorithmic tools, very little work has been done to explore how to capitalize on the interaction between human and machine pattern recognition. This paper introduces a preliminary roadmap for guiding research in this space. One key concept is the explicit consideration of the user's task, since the effectiveness of different visual representations and interaction methods will depend on the goal. The second is the explicit inclusion of a "human-in-the-loop," who interacts with the data, the algorithms, and representations, to identify meaningful features, using visual, auditory and haptic tools. And, thirdly, this framework explicitly closes the loop, in that features identified or "carved out" by the human are characterized mathematically, for use in comparison, database query or analysis.

8291-20, Session 7A

How do we watch images?: a case of change detection and quality estimation

J. Radun, T. Leisti, T. Virtanen, G. Nyman, Univ. of Helsinki (Finland)

The most common tasks in subjective image estimation are change detection (a detection task) and image quality estimation (a preference task). We examined how the task influences the gaze behavior when comparing detection and preference tasks. The eye movements of 16 naïve observers were recorded with 8 observers in both tasks. The setting was a flicker paradigm, where the observers see a non-manipulated image, a manipulated version of the image and again the non-manipulated image and estimate the difference they perceived in them. The material was photographic material with different image distortions and contents. To examine the spatial distribution of fixations, we defined the regions of interest using a memory task and calculated information entropy to estimate the spatial concentration of fixations. The quality task was faster and needed fewer fixations and the first eight fixations were more concentrated on certain image areas than in the change detection task. The bottom-up influences of the image also caused more variation to the gaze behavior in the quality estimation task than in the change detection task. The results show that the strategies for making the tasks are different and in subjective image estimation studies it is important to think about the task.

8291-21, Session 7A

Measuring saliency in images: which experimental parameters for the assessment of image quality?

C. Fredembach, G. Woolfe, J. Wang, Canon Information Systems Research Australia Pty. Ltd. (Australia)

Predicting which areas of an image are perceptually salient or attended to has become an essential pre-requisite of many computer vision applications. Because observers are notoriously unreliable in remembering where they look a posteriori, and because asking where they look while observing the image necessarily influences the results, ground truth about saliency and visual attention has to be obtained by gaze tracking methods.

From the early work of Buswell and Yarbus to the most recent forays in computer vision there has been, perhaps unfortunately, little agreement on standardisation of eye tracking protocols for measuring visual attention. As the number of parameters involved in the experimental methodology can be large, their individual influence on the final results is not well understood. Consequently, the performance of saliency algorithms, when assessed by correlation techniques varies greatly across the literature.

In this paper, we concern ourselves with the problem of image quality. Specifically: where people look when judging images. We show that in this case, the performance gap between existing saliency prediction algorithms and experimental results is significantly larger than otherwise reported. To understand this discrepancy, we first set out to devise an experimental protocol that is adapted to the task of measuring image quality. In a second step, we compare our experimental parameters with the ones of existing methods and show that a lot of the variability can directly be ascribed to these differences in experimental methodology and choice of variables.

In particular, the choice of a task: judging image quality vs. free viewing has a great impact on measured saliency maps, suggesting that even for a mildly cognitive task, ground truth obtained by free viewing does not adapt well. Careful analysis of the prior art also reveals that

systematic bias can occur depending on instrumental calibration and the choice of test images.

We conclude this work by proposing a set of parameters, tasks and images that can be used to compare the various saliency prediction methods in a manner that is meaningful for image quality assessment.

8291-29, Session 7A

Examining the effect of task on viewing behavior in videos using saliency maps

H. Alers, J. A. Redi, Technische Univ. Delft (Netherlands); I. Heynderickx, Philips Research (Netherlands)

This work is aimed at understanding how the viewing task and level of video quality affect the viewing behavior of the observer. To test that, a database of 50 videos of 20 seconds each was created. These videos were degraded to different levels of quality. Consequently, observers were asked to watch these videos where half of them were given the task of scoring the quality level while the rest were told to simply watch the videos freely. By using an eye tracking device, it was possible to record the viewing behavior data of the observers. This data was used to create saliency maps for each second of video, which made it possible to compare the viewing behavior between the two tasks across different levels of quality. The results suggest that there is a difference in the viewing behavior depending on the task given to the viewer which is not effected with the quality level of the video. It is also shown that there are specific attributes in the videos, which affect the level of similarity in viewing behavior between the two tasks.

8291-23, Session 7B

Investigations of the tone reproduction curves on the perceived image quality for fine art reproductions

J. Jiang, F. Frey, S. Farnand, Rochester Institute of Technology (United States)

Based on the result from a recent web-based experiment conducted to evaluate the perceived image quality without the reference sponsored by Mellon Foundation, a project is being conducted to understand the effect of the contrast adjustment made in the process of fine art reproduction on the perceived image quality. Contrasts are usually adjusted using the tone reproduction curve (TRC) in order to achieve the highest possible dynamic range and to better represent the original artwork. Copies of fine art images are simulated using tone reproduction curves collected from different museums. During the experiment, observers are instructed to rate each test image based on their preference in a web-based environment. Observers will be asked to click on the area that most influences their decisions. More attention may be needed to the areas on which most clicks are made when the tone reproduction curve is adjusted during visual editing. To understand the impact of the original on participants' preference judgments, the experiment will be conducted both with and without the reference. The result may be indicative of whether a separate TRC is needed for fine art reproductions depending on the availability of the original when the reproduced image is viewed.

8291-30, Session 7B

Characterizing eye movements during temporal- and global-quality assessment of h.264 compressed video sequences

C. Mantel, N. Guyader, P. Ladret, G. Ionescu, Gipsa-lab (France); T. Kunlin, STMicroelectronics (France)

Studies have shown that the deployment of visual attention is closely link to the assessment of image or video quality, though this link is not yet fully understood. The influence of rating temporal quality of compressed videos over the way an observer deploys his attention is

investigated in this paper.

We set-up a subjective experiment in which the eye movements of observers are recorded during three different tasks: a free-viewing task (FT), a global quality assessment task and a temporal quality assessment task. The FT acts as a reference to which we compare the eye movements during the two other tasks.

As previously shown, observers assessing global quality gaze at locations dissimilar to those fixated during the FT. For temporal quality assessment, it seems that the fixated locations are closer to FT than the global quality assessment fixated locations.

Our results suggest that the locations observers look at do not depend on the displayed video quality level. Quality however influences the way participants look at videos: the lower the quality, the longer they gaze at a precise location. The area fixated seems to be much smaller during the quality assessment tasks than during the FT for either perfect or poor quality level.

The evolution over time of all indicators suggests that, during the first 1 or 2 seconds, the signal properties of the videos are the main attractors for the participants' eye movements. Instructions only seem to play a role afterwards on the deployment of the participants' visual attention.

8291-31, Session 7B

A compressed sensing model of crowding in peripheral vision

J. Hocke, Univ. zu Lübeck (Germany); M. Dorr, Schepens Eye Research Institute (United States); E. Barth, Univ. zu Lübeck (Germany)

We here model peripheral vision in a compressed sensing framework and find that typical letter-crowding effects arise naturally. Images are re-sampled with samples that are distributed according to a Poisson-disk distribution. These samples are then considered to be inputs to a network of neurons such that the output neurons are randomly connected to the input neurons. Also the weights of the connections are assigned randomly. Convergence is modeled by allowing for fewer output neurons than input neurons. By reconstructing the original image from the activity of the output neurons we estimate the amount of information delivered at the output. To reconstruct the original image, the signal at the output neurons is modeled as a linear superposition in a sparse basis and a sparse-approximation algorithm is applied. The reconstructed images exhibit typical crowding artifacts in the sense that single letters can be recognized but not letters which are flanked by additional letters. When the sampling rate is increased, as would be the case in central vision, both the single and the flanked letters can be recognized.

8291-32, Session 7B

Foveated self-similarity in nonlocal image filtering

A. Foi, Tampere Univ. of Technology (Finland); G. Boracchi, Politecnico di Milano (Italy)

We investigate the role of foveation in nonlocal image filtering and, in particular, in the NLmeans denoising algorithm. The nonlocal self-similarity between regions at different image locations is used in NLmeans to adaptively determine the weights for pointwise estimation. This similarity between different regions is computed through the windowed photometric differences, where the window function decays to zero as we move away from center of the corresponding region.

We replace such windowed difference with a foveated difference: instead of multiplying the photometric differences against a window function, we blur the differences with point-spread functions having increasing standard-deviation (i.e. increasing spread) as the distance from the region's center grows. This foveated difference mimics the actual ability of the human visual system to perceive details at the periphery of the point of fixation.

Our study show that a foveated self-similarity leads to an improvement in the resulting image estimate, according both to objective criteria and

visual appearance, particularly due to better contrast and sharpness.

We discuss this form of self-similarity in the context of natural image statistics as well as its potential implications in explaining the role of eye movements in improving visual perception.

8291-33, Session 7B

A statistical study of the correlation between interest points and gaze points

M. Nauge, M. Larabi, Univ. de Poitiers (France)

In this work, we propose to study the performance of these detectors with a totally different point of view, by studying similarity/correlation between the subjective gaze points and the objective interest points. The aim is to determine whether the interest points can be used to predict salient information on an image like the HVS does. This can help for several applications like quality assessment, simplified saliency maps construction, . . . Even though the interest points have not been originally designed to be close to the gaze points, they may have a particular setting that maximizes the similarity between them. For this study a battery of statistical tools is used to test a large range of settings/configurations for the three interest point detectors mentioned previously.

We used several statistical tools such as Bartlett, ANOVA, . . . to understand the effect and the influence of each parameters for each detector. These studies illustrate that particular parameters can minimize the cost of transformation and predict interest points in accordance with the subjective gaze points. We also proposed a solution to give a scale to facilitate the interpretation of the EMD values by analyzing the mean human behavior. By comparing the best setting for each detector we can also indicate which detector is the most reliable to estimate the subjective gaze points. This study is also a good way to prove than interest points detector share some properties with the HVS.

8291-34, Session 7B

Interest point analysis as a model for the Poggendorff illusion

F. W. M. Stentiford, Univ. College London (United Kingdom)

This paper describes a recognition mechanism based on the relationships between interest points and their properties that is applied to the problem of modelling the Poggendorff illusion. The recognition mechanism is shown to perform in the same manner as human vision on the standard illusion and reduced effects are modelled on a variant without parallels. The results show that the recognition mechanism produces high similarity scores between continuous transversal lines and Poggendorff figures where the line segments are offset from the collinear position. The peaks occur in the same direction of shift as reported in psychophysical investigations of human vision. The most prominent peaks occur with transversals inclined at the smallest angles to the vertical and decrease towards the perpendicular position (90°). The analysis of figures with no parallels also yielded a peak although this was not as prominent as with the conventional Poggendorff figures. This was also in agreement with earlier behavioural experiments. The analysis is also shown to be consistent with psychophysical findings where obtuse angled variants of the Poggendorff figures yield stronger effects than the acute angled variants.

8291-35, Session 8

The perception of art and the science of perception

R. Pepperell, Cardiff School of Art & Design (United Kingdom)

For many centuries artists have studied the nature of visual perception and how to convincingly render what we see using a variety of media. The results of this prolonged investigation, during which knowledge and insights were handed down from generation to generation, can be found in all the countless artworks deposited in museums and galleries around the world. Works of art represent a rich source of ideas and understanding about how the world appears to us, and only relatively recently have those interested in the science of vision started to appreciate the many discoveries made by artists in this field. In this talk I will discuss some of the key insights into vision and perception that artists have revealed through their inquiries, and show how they can help current thinking in science and technology about how best to understand the process of seeing. In particular, I will suggest that some important artistic ideas continue to present fundamental challenges to conventional ideas about how reality is represented.

8291-36, Session 8

Paintings, photographs, and computer graphics are calculated appearances

J. J. McCann, McCann Imaging (United States)

The challenge of a reproduction is to capture the information contained in the physics of light (original 3-D color space) and to render it in a different size and shape of media (reproduction color space).

The reproduction problem is very similar to moving everything in your house to a new one. The reproduction house has different dimensions for the length (amount of red), width (amount of green) and height (amount of blue). On a clear day, shadows cast by the sun are 32 times darker than direct sunlight. The 32:1 range of reflectances in a 32:1 range of illumination creates a 1024:1 HDR range of light. Real-life scene reproduction is analogous to moving a castle into a cottage. Reproductions move everything in the old house into the new house, keeping all contents in corresponding rooms, even though the dimensions of the entire house, and each room, are different. Good reproductions are never exact physical copies of the original, because that is not possible. Good reproductions capture the appearance and relationships of objects in the scene.

Painters reproduce the appearances they see, or visualize. The entire human visual system is the first part of that process, providing extensive spatial processing. Painters have used spatial techniques since the Renaissance to render HDR scenes. Silver halide photography responds to the light falling on single film pixels. Film can only mimic the retinal response of the cones at the start of the visual process. Film cannot mimic the spatial processing in humans. Digital image processing can. So with spatial algorithms, we can calculate appearances and write them to print and display.

8291-37, Session 8

Image integrity and aesthetics: towards a more encompassing definition of visual quality

J. A. Redi, Technische Univ. Delft (Netherlands); I. Heynderickx, Philips Research (Netherlands) and Delft Univ. of Technology (Netherlands)

Visual quality has been so far investigated different, sometimes contrasting perspectives. The electronic imaging community has for long limited visual quality to the perceived integrity of the media, when affected by visual degradations due to signal errors or technological limitations. From another perspective, the media management community has been investigating on a definition of visual quality more related to content pleasantness, and in particular on aesthetic quality assessment. In this work, we aim at moving the first steps towards a more encompassing definition of visual quality, which takes into account both the image integrity and the pleasantness of its content. To do so, we start by investigating the interactions between aesthetic and artifactual attributes in visual quality assessment. We designed a psychometric experiment to investigate both the role of artifactual attributes in aesthetic quality assessment, and the differences in viewing behaviour, between free image observation, image integrity evaluation and aesthetic quality assessment. An eye-tracker supports

the data collection for this purpose. Through the analysis of the resulting data, we aim at collecting useful information to design a more complete set of attributes on which basing future visual quality assessment models.

8291-38, Session 8

Depicting 3D shape using lines

D. DeCarlo, Rutgers, The State Univ. of New Jersey (United States)

No abstract available

8291-39, Session 8

Box spaces in pictorial space: linear perspective versus templates

H. de Ridder, S. C. Pont, Technische Univ. Delft (Netherlands)

No abstract available

8291-40, Session 8

Warping realist art to ensure consistent perspective: a new software tool for art investigations

Y. Chang, Wolfram Research (United States); D. G. Stork, Ricoh Innovations, Inc. (United States)

Many realist paintings, especially ones from the early Renaissance, do not conform to the rules of geometrical perspective: orthogonals (lines in three dimensions perpendicular to the plane of the painting) may not meet at a vanishing point, separate vanishing points may not lie on a unique horizon line, and so on. We have developed a Mathematica-based tool that warps images in response to user-defined control points. Modifications of this software can allow art scholars to click on points on perspective lines-orthogonals and transversals (parallel to the plane of the painting)-and then adjust a single parameter controlling the amount of perspective "correction" between 0 (no warping) and 1 (full warping to ensure consistent perspective). Of course, we are not claiming that a painting that conforms to the rules of geometric perspective is in anyway "superior" to one that does not, or that an image that conforms to perspective rules is "correct." Our novel visualization tool allows art scholars to expose and visualize more clearly the perspective schemes by artists. Quantitative measures of the perspective coherence provided by our software can be used in diachronic studies of the development of works by some artists, such as Cézanne's still lifes. We will demonstrate our software on works by Masolino, Jan van Eyck, and others, and speculate on its uses elsewhere in art studies.

8291-41, Session 8

Sound meets image: freedom of expression in texture description

R. J. Jansen, R. van Egmond, H. de Ridder, Technische Univ. Delft (Netherlands); T. N. Pappas, Northwestern Univ. (United States)

Auditory, visual, and tactile textures provide information about events and objects in the environment. It is interesting to comprehend how sensory modalities can influence each other. We are able to describe various psychophysical characteristics of a texture on a sensorial level. Yet, little is known about their relation to what people hear, see, and feel on a perceptual level. The aim of this study is to find relevant perceptual dimensions to describe auditory and visual textures, and to explore how these dimensions are related to underlying psychophysical measures of both sensory systems. For Experiment 1 two sets of 17 visual textures were prepared: one set taken from the CUReT database, and one set synthesized to replicate the former set. Participants were

instructed to match a sound texture with a visual texture displayed onscreen. They were provided with an interactive physical interface, coupled to a frequency modulation synthesizer. Rather than selecting from a pre-defined set of sound samples, continuous exploration of the auditory space allowed for an increased freedom of expression. While doing so, participants were asked to describe what auditory and visual qualities they were paying attention to. These quality descriptions were used in Experiment 2 to rate the visual textures and sounds generated in Experiment 1. Finally, in Experiment 3 participants were instructed to choose a visual texture for each sound generated in Experiment 1.

8291-42, Session 9

On the dynamics of aesthetic appreciation

C. Carbon, Otto-Friedrich-Univ. Bamberg (Germany)

Personal taste develops over time and is highly susceptible for Zeitgeist-dependent effects. Using an adaptation paradigm often used in the domain of face research (e.g., Carbon & Ditye, 2011), we could show in a series of experiments that not only the representation of designs (e.g., car designs, Carbon, 2010) or artworks (Carbon & Leder, 2006), but also taste quickly adapts towards adaptors (Carbon, Ditye, & Leder, 2006). The present paper will not only show the adaptation towards specific design characteristics in product design, but even so for specific art-specific properties. For instance, one experiment made use of the idiosyncratic style of Amedeo Modigliani who employed manneristic face elongation. In the first part of the experiment, the pre-adaptation phase (T1), participants were asked to rate their liking of a number of portraits by various artists of a variety of different historical epochs on a seven-point Likert-scale. Each portrait was presented in 11 different versions which differed in the amount of compression or extension in heights (cf. Carbon et al., 2007). This block was followed by an adaptation-phase in which Modigliani's paintings were shown. In a succeeding test-phase, the post-adaptation phase (T2), liking of the same paintings of T1 was rated again by the participants. Results revealed systematic changes in aesthetic appreciation for those exemplars that were structurally rather similar in the style to Modigliani (the extended versions). The talk qualifies and discusses such style-based adaptation effects to reveal their cognitive basis.

8291-43, Session 9

An experimental and theoretical analysis of the merits of 'virtual cleaning' of paintings

J. Delaney, National Gallery of Art (United States); A. Casini, Istituto di Fisica Applicata Nello Carrara (Italy); L. Simonot, Univ. de Poitiers (France); M. Thoury, National Gallery of Art (United States); M. Picollo, L. Stefani, Istituto di Fisica Applicata Nello Carrara (Italy); M. Poggesi, Consiglio Nazionale delle Ricerche (Italy); D. Conover, K. Fleisher, E. R. de la Rie, National Gallery of Art (United States)

Image processing to improve color accuracy of images of paintings has expanded with the advent of multispectral and, more recently, high spatial hyperspectral imaging cameras. The increased color accuracy results from a more accurate measurement of the spectral reflectance. Various groups have attempted to restore the color appearance of Old Master paintings by compensating for faded pigments and discolored varnishes using such reflectance information in their models. The latter has been called "virtual cleaning" of paintings. In these models the varnish is often treated as a transmission filter and color is restored to the painting by removing the discolored varnish's absorbance. The absorbance of the aged varnish is determined by either a heuristic process, or altering the reflectance spectra by using reflectance or transmittance spectra of artificially aged varnish.

In this paper the results of experiments designed to test the accuracy of such "virtual cleaning", using several paintings covered with aged varnishes, is presented. The experimental method consists of collecting reflectance spectra before varnish removal and then after application of new varnish. Also the absorbance properties of removed vanish are measured. Two types of experimental studies are performed; the first using a fiber optic reflectance spectrometer (350

to 2500 nm) at selected sites and the second a color accurate imaging hyperspectral mechanical scanning camera system (400 to 900 nm) to collect reflectance spectra over an entire painting. The first set allows determining the degree to which a transmission model can be used to predict the final reflectance spectra of the cleaned and varnished area. The second set allows for a visual comparison of results of virtual cleaning versus actual on a painting. The results will be used to demonstrate whether "virtual cleaning" does or does not account for the scattering inherent in naturally aged varnishes or variation in varnish thickness and what improvements in the models are required.

8291-44, Session 9

In search of Leonardo: computer-based facial image analysis of Renaissance artworks for identifying Leonardo as subject

C. W. Tyler, The Smith-Kettlewell Eye Research Institute (United States); W. Smith, The Univ. of York (United Kingdom); D. G. Stork, Ricoh Innovations, Inc. (United States)

One of the enduring mysteries in the history of the Renaissance is the the adult appearance of the archetypical 'Renaissance Man', Leonardo da Vinci. His only acknowledged self-portrait is from an advanced age, and various candidate images of younger men are difficult to assess in the absence of documentary evidence.

Reconstructing 3D face shape from a single view is a well-studied problem. In general, single image shape estimation is an ill-posed problem even when strong assumptions are made about scene geometry, reflectance and camera properties. In the case of paintings, this process is further confounded by potential errors introduced by the artist in their rendition of perspective, shading and illumination.

For these reasons, we neglect photometric cues and reconstruct 3D face shape using the projected position of fiducial points only. This approach has been shown to perform well on synthetic data. We employ a linear statistical model of 3D face shape (morphable model), learned from a representative sample of human faces. The model is used to constrain the face shape recovery process by transforming it to one of parameter estimation. We form a bilinear system in which the unknowns are the camera projection matrix (approximated by an affine camera) and the face shape parameters. This optimization is convex and can be efficiently solved using alternating least squares.

For robustness, we learn an empirical model of generalization error over the surface of the face. This describes the variance between a true face shape and its best model fit. This allows for feature-sensitive weighting of data-closeness errors and allows us to regularize the optimization problem without manual parameter tuning.

8291-45, Session 9

Non-destructive analytical imaging of metallic surfaces using spectral measurements and ultrahigh-resolution scanning for cultural heritage investigation

J. Kaneko, Y. Murayama, J. A. Toque, A. Ide-Ektessabi, Kyoto Univ. (Japan)

This paper presents an approach for analyzing metallic surfaces using spectroscopic and analytical imaging methods. Spectral luminance, CIE LAB, XYZ and device RGB value of more than 30 metallic surfaces on Japanese papers were measured with a spectrometer and scanned in order to estimate the spectral reflectance. This was done to facilitate the statistical discrimination between individual metallic foils. A reference database with the spectral reflectance of the foils was measured from 400 to 700 nm, with a step width of 10nm. The images were captured by polarizing the light source and the camera to separate the diffused and specular components of the reflection. The angle between the filter in front of the light source and the camera was changed at 10 degrees increments. Results have shown that the spectral reflectance and its peak at a particular wavelength region

differed from each metallic foils, and that the images yielded different characteristics. This could facilitate in distinguishing between foils. This provides some promise that unknown metallic foils may be identified through the measurement of their spectroscopic features. This method could become a useful tool for the analysis of cultural heritage.

8291-46, Session 9

Museum as an integrated imaging device: visualization of ancient Kyoto cityscape from folding screen artifact

K. Miyata, U. Oyabu, M. Kojima, National Museum of Japanese History (Japan)

Museums hold cultural resources such as artworks, historical artifacts, and folklore materials. The National Museum of Japanese History (NMJH) holds over 200,000 of the cultural resources. A role in museums is to exhibit the cultural resources, therefore museums could be referred to as a visualization device for the information-based society. In this research, a method to visualize a historical image from cultural resources with interactive user interface will be mentioned. The material focused on is the oldest extant version of a genre of folding screen paintings that depict the thriving city of Kyoto in the four seasons, named Rekihaku's "Scenes In and Around Kyoto" designated as a nationally important cultural property. Over 1,400 people and a lot of residences, temples, and houses are drawn, and it is very important information resource telling us about city scenes and people's life in Kyoto at that time.

Historical researches were done by using a high resolution digital image obtained by a large scaled scanner that can change the geometry of the sensor, then the research results will be put on the database application for a special exhibition held in NMJH to support for visualizing a cityscape of ancient Kyoto in the visitors' mind.

8291-47, Session 9

Mapping colors from paintings to tapestries: rejuvenating the faded colors in tapestries based on colors in reference paintings

E. Johansson, M. Strom, Chalmers Univ. of Technology (Sweden); D. G. Stork, Ricoh Innovations, Inc. (United States)

We addressed the problem of automatically recovering the original colors in faded tapestries by digitally mapping colors from paintings or cartoons which served as the reference source. There are several deep computational challenges in such a process. First, the images in the source cartoon may differ in shape somewhat from the source cartoon, so any mapping algorithm must be robust and tolerant to minor shape differences. Second, there may be missing or shifted sections (e.g., an added, deleted, or shifted figure). Third, different colors fade by different amounts.

We addressed these challenges with a clustering scheme with a global cost function that incorporates color differences, location differences, and region (area) differences, and manipulated (by hand) relative weightings of these factors in order to map cartoon colors to the tapestry, thereby rejuvenating its colors. We have demonstrated our algorithm on artificial images as well as details of cartoon-tapestry pairs. We will present our refined algorithm on Raphael's cartoons for the tapestries in Sistine Chapel and Francois Boucher's "The forge of Vulcan."

Conference 8292: Color Imaging XVII: Displaying, Processing, Hardcopy, and Applications

Tuesday-Thursday 24-26 January 2012

Part of Proceedings of SPIE Vol. 8292 Color Imaging XVII: Displaying, Processing, Hardcopy, and Applications

8292-01, Session 1

Spatiochromatic detection, appearance, and motivations for wide-gamut color

S. J. Daly, Dolby Labs., Inc. (United States)

No abstract available

8292-02, Session 1

Color assimilation and contrast near absolute threshold

J. J. McCann, McCann Imaging (United States)

Human light response spans more than 10 log units of dynamic range. That is the ratio of radiances from snow on a mountain top to human dark-adapted threshold. This range is possible because two types of retinal cells: rods (high sensitivity) and cones (daylight vision). There are many familiar spatial color experiments in which equal local stimuli make unequal appearances in daylight (Color Assimilation and Color Contrast). Since these experiments are the result of spatial processes, it is possible that the different anatomy and physiology of rods and cones could limit the range of these effects.

Remarkably, spatial effects at the lowest end of our visual HDR range are very similar to those at the top in sunlight. We found that color vision has consistent spatial behavior near absolute threshold and in bright sunlight. Our experiments show that Color Assimilation and Color Contrast behave the same using rod/Lcone color vision. Other achromatic Assimilation/Contrast experiments appeared similar over the entire HDR range of illumination. There were changes in sharpness and color hue, nevertheless, the spatial comparison process was remarkably similar, despite the well known differences in the anatomy and physiology of rods and cones.

8292-03, Session 1

Multi-illuminant color constancy for HDR images through exposure segmentation

H. Zhang, H. Liu, Oregon State Univ. (United States); S. Quan, Broadcom Corp. (United States)

In high dynamic range imaging, a standard practice is to compose an image through exposure bracketing which takes a series of exposures of the same scene and then combines them together, followed by dynamic range compression and some color processing steps. This process can effectively extend the equivalent dynamic range of the image sensor, making mixed lighting condition a more worth-considering issue. Traditional color constancy algorithms first estimate a global white point of the scene and then apply white balance and color correction based on this estimate. This method could exaggerate the difference between the illuminants within the same scene, making part of the image better and part of the image worse, or having to compromise the colors of the whole image.

In this paper, we propose a color constancy algorithm for high dynamic range images targeted for multi-illuminant scenes with a focus on the combination of the exposures with different illuminants. By utilizing the inherent difference between the illuminants, we first create a segmentation mask to differentiate areas lit by different illuminants before they are individually estimated. The final image is formed by dissolving the exposures and is color corrected with parameters interpolated between the illuminants with morphological processing techniques.

8292-04, Session 2

Tone mapping for HDR images with dimidiate luminance and spatial distributions of bright and dark regions

M. Kitaura, F. Okura, M. Kanbara, N. Yokoya, Nara Institute of Science and Technology (Japan)

In order to display high dynamic range (HDR) image on low dynamic range (LDR) display devices, it is necessary to compress a dynamic range of the HDR image approximately by using tone mapping. In related work of tone mapping, there is a limitation to represent HDR images because the magnitude relationship between luminance values of pixels in images is kept. In addition, tone mapping methods based on human's perception, which have been recently proposed to represent the local information of HDR images, cannot utilize a dynamic range effectively because the magnitude relationship between luminance values of adjacent pixels is kept. In this paper, we propose a tone mapping method for HDR images with two separated luminance distributions, where luminance distributions and image regions of light and dark parts are divided clearly. Under the assumption that human's perception does not feel a sense of discomfort if the magnitude relationship between luminance values of pixels in light and dark regions is reversed, our method applies tone mapping to each region independently. In experiments, we confirmed that the result image generated by our tone mapping method was better than results of other related work for an HDR image with two separated luminance distributions.

8292-05, Session 2

Color universal design: analysis of color category dependency on color vision type

N. Kojima, Y. G. Ichihara, Kogakuin Univ. (Japan); K. Ito, The Univ. of Tokyo (Japan); M. G. Kamachi, Kogakuin Univ. (Japan)

This report is a follow-up to SPIE-IS & T / Vol. 7528051-8 and SPIE-IS & T / Vol. 7866 78660J-1-8.

In a previous survey, we investigated color categories of common and uncommon points of each color vision type, trichromat (C-type color vision), protan (P-type color vision) and deuteran (D-type color vision). In the present study, we first validated results from the previous survey. Next, we investigated specific color categories of P-type color vision and D-type color vision that do not depend on the color names common to C-type color vision. Finally, we investigated the difference threshold on the blue-yellow axis of the CIE chromaticity diagram for C-type color vision and P-type color vision. As a result, we propose a common chromaticity of colors that the three color vision types are able to categorize by means of color names common to C-type color vision. We also propose specific color categories of C-type color vision, P-type color vision and D-type color vision and highlight the perception gap of color between C-type color vision and P-type color vision on the blue-yellow axis of the CIE chromaticity diagram.

8292-06, Session 2

Colour perception with changes in levels of illumination

K. F. Baah, Dept. of Health (United Kingdom) and Univ. of the Arts London (United Kingdom)

The display of uniform colour stimuli that relies on appearance stability as an important cue or identifier has dependency on current standards. Typically such stimulus would be judged using a known standard but in the target environment the illuminance may be entirely different. For critical and practical appraisal of colour ISO 3664:2009 is commonly used and requires illuminance levels of 2000 lux (+/-250) and 500 lux (+/-125) respectively under a D50 illuminant. Using ISO 3664:2009 a series of psychophysical experiments were conducted to determine whether changes in illumination levels between 25-3000 lux significantly alter colour perception when considering acceptability and perceptibility thresholds of uniform colour stimulus.

It was found that the average change inter-observer perceived colour difference altered up to a maximum of 1DE00. However, for the perceptual correlate of hue this could be of significance if the accepted error of colour difference was at the threshold thereby yielding the possibility of rejection with illumination level changes. Lightness and chroma on the other hand exhibited greater tolerance and were less likely to be rejected with illuminance changes.

8292-07, Session 3

Reducing the number of calibration patterns for the two-by-two dot centering model

V. Babaei, R. Rossier, R. D. Hersch, Ecole Polytechnique Fédérale de Lausanne (Switzerland)

The two-by-two dot centering model has been designed for predicting the spectral reflectance of color halftones. It requires measuring the reflectances of a large number of two-by-two calibration tile patterns. Spectral measurement of hundreds or thousands of tile patterns is cumbersome and time consuming. In order to limit the number of measurements, we estimate the reflectances of most two-by-two calibration tile patterns from a small subset comprising less than 10% of all tile patterns. By linear regression in the absorptance space we derive a transformation matrix converting tile pattern colorant surface coverages to absorptances. This transformation matrix enables calculating the absorptance of all remaining two-by-two tile patterns. For a cyan, magenta and yellow print, with 72 two-by-two measured calibration tile patterns, we are able to create a two-by-two dot centering model having an accuracy slightly below the accuracy of the model with the fully measured set of 1072 two-by-two tile patterns.

8292-08, Session 3

Spatial gamut mapping for preserving the details of an image

I. Song, H. Ha, W. Kyung, Y. Ha, Kyungpook National Univ. (Korea, Republic of)

A recent trend in gamut mapping algorithm (GMA) research is the use of spatial information to compute the color transformation of pixels from the input to the output gamut. To achieve this, low-/high-pass filtering or pyramidal decomposition is often used. However, these techniques share the problem of halo generation in the case of sharp boundaries, which is induced by the spatial computation itself. Another problem with spatial GMAs is the hue shift induced from compressing a group of pixels towards a gray using the same vector, thus pixels that do not need to be changed are erroneously mapped.

Spatial gamut mapping using a regularization process is proposed with the aim of maintaining the hue and details of the original image as much as possible within the gamut limits of the reproduction medium. Thus, for the details, the difference between the original image and its low-pass filtered image is added, while the hue is preserved through post-processing. The entire process is implemented using CIELAB color space, and the image is filtered in all three orthogonal channels.

8292-09, Session 3

Evaluating color calibration kits with virtual display

W. Cheng, H. Caceres, A. Badano, U.S. Food and Drug Administration (United States)

A new method of evaluating the performance of color calibration kits for LCD monitors is proposed. Routine color calibration is imperative for medical applications that rely on color fidelity such as digital pathology, endoscopy, and colposcopy. However, the commercially available products vary greatly in price and performance with no available evaluation standard. We propose the concept of Virtual Display, a universal display platform that emulates the colorimetric response of real displays. A field programmable gate array (FPGA) board was used to process the video signals based on a pre-programmed look-up table, which contains the characterization data of the real display. A wide-color-gamut display was driven by the FPGA to emulate the colorimetric response of the real display. The virtual display's hardware-based approach is transparent, so the color calibration kits can calibrate it as a regular display without knowing the difference. By changing the look-up table in the FPGA, the virtual display can emulate various real displays for test various color calibration kits, which is different from previous work where only one display was used. Our experimental results show that the virtual display can emulate real displays within 3.64 delta-E-ab discrepancies. Such accuracy was achieved by our method of spectral characterization and modeling.

8292-10, Session 3

Optimizing color fidelity in wide-gamut-display devices when processing images compressed by block-based DCT transforms

F. Lebowsky, STMicroelectronics (France)

High-end monitors based on LCD technology increasingly address wide color gamut implementations featuring precise color calibration within a variety of different color spaces such as extended sRGB or AdobeRGB. However, images are often reconstructed from digitally compressed images files such as JPEG or MPEG where color quality could be questionable. In particular, when such image files are scaled up or zoomed in, different types of image artifacts become visually noticeable. Among these artifacts we find pixelation, blockiness, ringing, and color blotching. While pixellation and ringing appear due to insufficient adaptation to image context using a static or context adaptive filter kernel in temporal domain, blockiness and ringing occur due to image compression in frequency domain, when image compression factors are high. In addition, chrominance channels often undergo an even higher compression ratio that amplifies visibility of artifacts such as color blotches. Consequently, we are interested in improving the quality of images to be displayed depending on the scale factor requested by a user application. We propose to discriminate most relevant visual artifacts using power spectrum analysis in DCT domain as well as after kernel based rescaling combined with statistical analysis based on non-linear regression taking into account characteristic non-stationary behavior of image content and identifiable visual artifacts. A comparative analysis based on some competitive solutions highlights the effectiveness of our approach and identifies its current limitations with regard to wide color gamut representation due to mathematical uncertainty of the studied artifacts.

Conference 8292

8292-11, Session 3

Optimal gamut volume design for three primary and multiprimary display systems

C. E. Rodriguez-Pardo, G. Sharma, Univ. of Rochester (United States); X. Feng, J. Speigle, I. Sezan, Sharp Labs. of America, Inc. (United States)

Primary selection plays a fundamental role in display design. Primaries affect not only the gamut of colors the systems is able to reproduce, but also, they have an impact on the power consumption and other cost related variables. Using more than the traditional three primaries has been shown to be a versatile way of extending the color gamut, widening the angle view of LCD screens and improving power consumption of displays systems. Adequate selection of primaries requires a trade-off between the multiple benefits the system offers, the costs and the complexity it implies, among other design parameters.

The purpose of this work is to present a methodology for optimal design for three primary and multiprimary display systems. We consider the gamut in perceptual spaces, which offer the advantage of an evaluation that correlates with human perception, and determine a design that maximize the gamut volume, constrained to a certain power budget, and analyze the benefits of increasing number of primaries, and their effect on other variables of performance like gamut coverage.

8292-12, Session 4

The dark side of CIELAB

G. Sharma, C. E. Rodriguez-Pardo, Univ. of Rochester (United States)

Standardized in 1976 as a uniform color space, CIELAB is extensively utilized in color science and engineering applications. CIELAB provides both a color difference formula and correlates for common perceptual descriptors of color. Deficiencies in both areas are well-known, and based on these known limitations, numerous fixes have been developed yielding alternative color difference formulae that are derived as modifications of the color difference in CIELAB. In addition, several new color appearance spaces have also been proposed as modifications of the basic CIELAB framework. In this paper, we point out other, lesser-known and poorly-appreciated, limitations of CIELAB that occur particularly in the dark regions of color space. We demonstrate via examples, how these limitations not only cause performance compromises but lead to fundamental breakdowns in system optimization and design problems, making CIELAB completely unusable in these problems. We consider the reasons why these fundamental limitations were overlooked in the original development of CIELAB and analyze the mathematical representations contributing to the undesired behavior. We argue that fundamental new research is required to overcome this dark side of CIELAB; the development of uniform color spaces and new color appearance spaces must be revisited afresh using new experimental data and keeping in mind newer devices and applications.

8292-13, Session 4

Complexitites of complex contrast

E. Peli, Schepens Eye Research Institute (United States)

No abstract available

8292-14, Session 4

It's not the pixel count, you fool

M. A. Kriss, MAK Consultants (United States)

The first thing a "marketing guy" asks the digital camera engineer is "how many pixels does it have, for we need as many mega pixels as possible since the other guys are killing us with their "umpteen" mega pixel pocket sized digital cameras. And so it goes until the pixels get smaller and smaller in order to inflate the pixel count in the never-ending pixel-wars. These small pixels just are not very good. The truth of the matter is that the most important feature of digital cameras in the last five years is the automatic motion control to stabilize the image on the sensor. All the rest has been hype and some "cool" design. What is the future for digital imaging and what will drive growth of camera sales (not counting the cell phone cameras which totally dominate the market in terms of camera sales) and more importantly after sales profits. Well sit in on the Dark Side of Color and find out what is being done to increase the after sales profits and don't be surprised if has been done long ago in some basement lab of a photographic company and of course, before its time.

8292-15, Session 4

Color imaging and aesthetics: is there the cheshire cat?

E. A. Fedorovskaya, Eastman Kodak Co. (United States)

No abstract available

8292-16, Session 4

Dark texture in artworks

C. E. Parraman, Univ. of the West of England (United Kingdom)

No abstract available

8292-17, Session 4

Harmonious colors: from alchemy to science

G. B. Beretta, N. M. Moroney, Hewlett-Packard Labs. (United States)

There is a very long tradition in designing color palettes for various applications. Although color palettes have been influenced by the available colorants, starting with the advent of aniline dyes in the late 1850s there have been few physical limits on the choice of individual colors. This abundance of choices exacerbates the problem of limiting the number of colors in a palette.

The traditional solution is that of "color forecasting." Color consultants assess the sentiment or affective state of a target customer class and compare it with colorants offered by the industry. The color forecasting business is very labor intensive and difficult, thus for years computer engineers have tried to come up with algorithms to design harmonious color palettes, alas with little commercial success.

We argue that the intellectual flaw resides in the belief that a masterful individual can devise a "perfect methodology" that the engineer can then reduce to practice in a computer program. We suggest that the correct approach is to consider color forecasting as an act of distillation, where a palette is digested from the sentiment of a very large number of people. We describe how this approach can be reduced to an algorithm.

8292-44, Poster Session

Characterization of color scanners based on SVR

B. Li, Y. Zhang, Jiangnan Univ. (China)

By researching the principle of colorimetric characterization method and Support Vector Regression (SVR), we analyze the feasibility of nonlinear transformation from scanner RGB color space to CIELAB color space based on SVR and built a new characterization model for scanner. Then we use the MATLABr2009a software to make a data simulation experiment to verify the accuracy of this model and figure

Conference 8292

IS&T / SPIE
Electronic Imaging
SCIENCE AND TECHNOLOGY

out the color differences by CIEDE2000 color difference formula. The experimental results show that the similarity is more than 99% between predicted values and true values of L, a, b. Based on CIEDE2000 color difference formula, the average color differences of training set and text set are 1.2376 and 1.9318. From our results, we can make a conclusion that SVR can realize the nonlinear transformation from scanner RGB color space to CIELAB color space and the model satisfies the accuracy of scanner characterization. Therefore, SVR can be used into the color scanner characterization management.

8292-46, Poster Session

Deducing ink thickness variations of fluorescent print by a spectral prediction model

J. Q. Wang, Y. Zhang, Jiangnan Univ. (China)

By the Clapper-Yule spectral prediction model, we develop a methodology to deduce ink thickness variations of printing on the fluorescent substrate. In the investigatione, we extend the Clapper-Yule spectral prediction model of printing on the fluorescent substrate by considering multiple internal reflections under the fluorescent substrate and the relationship between the transmittance and the ink thickness. At the same time, we deduce the ink thickness variations by introducing the relative ink thickness factor parameters and the ink thickness variation coefficient. By using the least-squares parameter estimation method, the deduction model of the ink thickness variation is developed. The correctness of the conclusions of is verified by experiments. In the experiment, a series of color offset printing is used to verify ink thickness variations of printing on the fluorescent substrate by a spectral prediction model. And the pages comprising the calibration patches and the test patches are printed on uncoated newsprint paper, with classical mutually rotated clustered-dot screens, at a screen frequency of 100 lines per inch.

8292-48, Poster Session

Bio-inspired color sketch for eco-friendly printing

I. V. Safonov, E. Tolstaya, M. N. Rychagov, Samsung Electronics Co., Ltd. (Russian Federation); H. K. Lee, S. Kim, D. Choi, Samsung Electronics Co., Ltd. (Korea, Republic of)

Saving of toner/ink consumption is an important task in modern printing devices. It has a positive ecological and social impact. We propose technique for converting print-job pictures to a recognizable and pleasant color sketches. Drawing a "pencil sketch" from a photo relates to a special area in image processing and computer graphics - non-photorealistic rendering. We describe a new approach for automatic sketch generation which allows to create well-recognizable sketches and to preserve partly colors of the initial picture. Our sketches contain significantly less color dots then initial images and this helps to save toner/ink. Our bio-inspired approach is based on sophisticated edge detection technique for a mask creation and multiplication of source image with increased contrast by this mask. To construct the mask we use DoG edge detection, which is a result of blending of initial image with its blurred copy through the alpha-channel, which is created from Saliency Map according to Pre-attentive Human Vision model. Measurement of percentage of saved toner and user study proves effectiveness of proposed technique for toner saving in eco-friendly printing mode.

8292-49, Poster Session

Reflectance model for recto-verso color halftone images

J. Wang, Y. Zhang, T. D. Wen, Jiangnan Univ. (China)

In the color reproduction process, accurately predicting the color of recto-verso images and establishing a spectral reflectance model

for halftones images are the great concern project of imaging quality control field. The scattering of light within paper and the ink penetration in the substrate are the key factors, which affect the color reproduction. A reflectance model for recto-verso color halftone prints is introduced in this paper which considers these factors. The paper based on the assumption that the colorant is non-scattering and the assumption that the paper is strong scattering substrate. By the multiple internal reflection between the paper substrate and the print-air interface of light, and the light along oblique path of the Williams-Clapper model, we proposed the color spectral reflectance precise prediction model of recto-verso halftone images. In the study, we propose this model for taking into account ink spreading, a phenomenon that occurs when printing an ink halftone in superposition with one or several solid inks. The ink-spreading model includes nominal-to-effective dot area coverage functions for each of the different ink overprint conditions by the least square curve fitting method, so the functions for physical dot gain of various overprint halftones are given. This model provided a theoretical foundation for color prediction analysis of recto-verso halftone images and the development of image quality detection system.

8292-50, Poster Session

The study on physical dot gain of second order FM halftone based on ink spreading in all ink superposition conditions

S. Xi, Y. Zhang, Jiangnan Univ. (China)

In this paper, the study is focused on the second order FM halftone prints. According to the feature of the second order FM halftone dots, a new spectral prediction model for the second order FM halftone prints based on the spectral Neugebauer random mixing model and dot-on-dot mixing model is developed, using a weighting factor to signify the correlation proportion of two models. Due to printing process, interaction among ink, paper and air, there will be physical and optical dot gain. The new spectral prediction model also relies on Yule-Nielsen modified Spectral Neugebauer model by taking into account ink spreading in all ink superposition conditions. When ink dots are printed on different superposition, ink spreading is different, so we create an ink spreading function for each superposition condition, which maps nominal to effective dot surface coverage. When ink dots are printed, their known nominal surface coverage values are converted into effective coverage values. Using ink jet printer to imitate the experiment, the mean and max ΔE difference between measured and predicted values are 2.52 and 6.26, smaller than the difference predicted by the model that is not taken into account ink spreading, which are 5.02 and 12.81. The prediction accuracy is considerably improved by the new model.

8292-51, Poster Session

Tensor decomposition for color printer model lookup table

V. Monga, The Pennsylvania State Univ. (United States); M. Maltz, Z. Fan, Xerox Corp. (United States)

No abstract available.

8292-52, Poster Session

Genetic algorithm for segmentation and classification of colony images automatically

W. Li, Henan Polytechnic Univ. (China)

Images egmentation and classification of colony images plays a key role in automatic visual systems. This paper describes a new algorithm using for segmentation and classification of colony images. It is based on a genetic approach that allow us to consider the segmentation

problem as a global optimization, and the new classifier introduced here is based on fuzzy-integration schemes controlled by a genetic optimization procedure. Two different types of integration are proposed here, and are validated by experiments on real data sets for Machine. Results show the good performance and robustness of the integrated classifier strategies.

8292-53, Poster Session

Color edge detection using edge density

W. Wang, Henan Polytechnic Univ. (China)

This paper presents a novel edge detection algorithm using the local, nonparametric estimation of the color image density. The method analyses the shape information provided by the local probability distribution of the image in the horizontal and vertical direction respectively, then obtains the modulus for edge detection. With the increasing of window size, the other types of distributions can be simplified to the three types of distributions presented in this paper. Experiments show the method is efficient.

8292-54, Poster Session

Spectral prediction model for variable dot-size ink jet presswork

W. Xing, Y. Zhang, Yangtze Univ. (China)

The variable dot-size ink jet printers can base the density of different are in the image, choose ink drop's size achieve variable dot-size. Using variable dot-size FM screening, recent ink jet printers can improve print quality. But the existing FM screening spectral reflectance model can't predict the reflectivity of such printed matter. Now, we found a prediction model, which is capable of predicting variable pixels dot-size spectral reflectance. When ink jet printers' drop size change, based on Clapper-Yule model, study reflectance prediction model for variable dot-size ink jet presswoke. At first, this article introduce variable drop size and the FM screening form of Clapper-Yule model; then, using the change of ink film thickness and FM screening dot, which are caused by drop change, extended Clapper-Yule model; at last, established the Clapper-Yule model which is extended by drop size, in addition, validated the advantage of the model by simulation. Spectral prediction model for variable dot-size ink jet presswoke administer to how can we get higher print quality presswoke use variable dot-size ink jet printers.

8292-18, Session 5

Detection and compensation of backlight images using chrominance

H. J. Park, S. W. Han, Samsung Electronics Co., Ltd. (Korea, Republic of)

An image taken under the backlight condition shows that a main foreground or object appears very dark, but a background appears relatively bright since the exposure time of the main object or foreground is relatively shorter than the one of the background due to high luminance from the background. The determination of a backlight image is generally done by luminance histogram analysis since it is believed that the distinct characteristic of the backlight image is a large luminance difference between foreground and background. However, this conventional detection method may not be adequate for video images since it generally targets on still images. Furthermore, the detection of backlight images would not be performed well if there are abrupt changes in light, motion, or scenes. Inaccurate detection leads to unnecessary compensation that makes image over-highlighted or flickered, especially when consecutive frames of video have different illumination modes. Since an image taken under normal light conditions may also have the similar luminance characteristics of backlight images, using only luminance information would not be sufficient to discriminate between normal and backlight images. Therefore, the analysis of chrominance of images is introduced to detect backlight images more accurately.

8292-20, Session 5

A new method for skin color enhancement

H. Z. Zeng, Hewlett-Packard Co. (United States); R. Luo, Univ. of Leeds (United Kingdom)

it pleasingly is an important factor in photographic color reproduction. Moving skin colors toward their preferred skin color center improves the color preference for skin color reproduction. Two key factors to successfully adjust skin colors are: a method to detect original skin colors effectively even if they are shifted far away from the regular skin color region, and a method to morph skin colors toward a preferred skin color region properly without introducing artifacts. In the same conference last year, we presented a method for skin color enhancement. A statistical skin color model was used for skin color detection, which may miss to detect skin colors that are far away from regular skin tons. In this paper, a new method using the combination of face detection and statistical skin color modeling is proposed to effectively detect skin pixels and to adjust skin colors effectively.

8292-21, Session 6

CIE chromaticity, Planckian locus, and correlated color temperature estimation from raw-image data using colorchecker training images

T. Tajbakhsh, Technische Univ. Hamburg-Harburg (Germany)

The major problem discussed in the paper is how good raw sensor chromaticities, i.e. $r = R/(R+G+B)$ and $b = B(R+G+B)$, can be mapped into the CIE 1931 xy chromaticity space given the constraint that solely training images are being used which are captured from a colorchecker illuminated with a certain emission spectrum. With these mappings we next investigate how good the Planckian and daylight locus can be estimated in the raw sensor chromaticity space. We evaluate a two-dimensional polynomial function to map the locus chromaticities from CIE xy to raw. Our proposed weighting strategy of the training chromaticities improves the locus mapping significantly. However, we show that two one-dimensional parables, similar to those defined by the CIE to describe the daylight locus chromaticities, outperform latter two-dimensional polynomials. The paper furthermore deals with the question how good the correlated color temperature of a raw chromaticity can be determined. We propose linear regression of a 2nd-order reciprocal polynomials to map rb coordinates to CCT values and compare the performance with two other methods. We present a comprehensive set of simulation results where we evaluate three figures of merit for different setups and mapping methods. Finally, we demonstrate the proposed methods on a 5 megapixel sensor and verify the results by measurement.

8292-22, Session 6

Comparative performance analysis of mobile displays

R. Safaee-Rad, Qualcomm Inc. (Canada); M. Aleksic, Qualcomm Inc. (United States)

Cell-phone display performance (in terms of color quality and fidelity) has become a critical factor in creating a positive user experience. As a result, there is a significant amount of effort by cell-phone OEMs to provide a more competitive display solution. This effort is focused on using different display technologies (with significantly different color characteristics) and more sophisticated display processors.

In this paper, the results of a display comparative performance analysis are presented. A few cell-phones from major OEMs are selected and their display performances are measured and quantified.

8292-23, Session 6

Termites: a Retinex implementation based on a colony of agents

G. Simone, Gjøvik Univ. College (Norway); G. Audino, Univ. degli Studi di Milano (Italy); I. Farup, Gjøvik Univ. College (Norway); A. Rizzi, Univ. degli Studi di Milano (Italy)

This paper describes a novel implementation of the Retinex algorithm with the exploration of the image made by an ant swarm. In this case the purpose of the ant colony is not the optimization of some constraint but the exploration as diffused as possible of the image content, with the possibility of tuning the exploration parameters on the image content. For this reason, this approach is called "termites", instead of ants, to underline the idea of the eager exploration of the image. The paper presents the spatial characteristics of locality and discusses differences with other Retinex implementation.

8292-24, Session 6

A color quantization algorithm based on minimization of L_p norm error in a modified CIELAB space

H. Xue, J. P. Allebach, C. A. Bouman, Purdue Univ. (United States)

Color quantization algorithms are used to select a small number of colors which can accurately represent the content of a particular image. Color quantization continues to be important because it can be used to segment images into regions of uniform color, improve image compression, and remove noise. However, in order to be most effective, color quantization algorithms must accurately represent the image with the minimum number of colors, and must be computationally efficient.

In this research, we introduce a novel color quantization algorithm which is based on the minimization of an L_p norm rather than the more traditional L_2 norm associated with MSE. We demonstrate that the L_p optimization approach has two advantages. First, it distributes the colors more uniformly over the regions of the image; and second, the norm's value can be used as an effective criterion for selecting the minimum number of colors necessary to achieve an accurate representation of the image. One potential disadvantage of the L_p norm criterion is that it could increase the computation of the associated clustering methods. However, we solve this problem by introducing a two state clustering procedure in which the first stage (pre-clustering) agglomerates the full set of pixels into a relatively large number of discrete-colors; and the second stage post-clustering) performs L_p norm minimization in a modified CIELAB space using the reduced number of discrete-colors resulting from the pre-clustering step. This two-stage clustering process dramatically reduces computation by merging together colors before the computationally expensive L_p norm minimization is applied.

8292-25, Session 7

Human skin imaging using three-phase spectral matching imager

A. Kimachi, Osaka Electro-Communication Univ. (Japan); S. Ando, The Univ. of Tokyo (Japan); M. Doi, S. Nishi, Osaka Electro-Communication Univ. (Japan)

This paper proposes an imaging method for human skin using the three-phase spectral matching imager (3PSMI), which consists of the time-domain correlation image sensor (CIS) and wavelength-swept monochromatic illumination (WSMI).

Each pixel of the CIS performs temporal correlation between the intensity signal and external three-phase reference signals generated from an orthonormal pair of reference spectra in each frame, producing images of correlation between the object spectrum and the reference spectra.

The similarity of the object spectrum to the reference spectra is evaluated as the azimuth angle of the projection of the object spectrum onto the plane spanned by the reference spectra.

Of the two methods of generating reference spectra, one method uses an average spectral reflectance of human skin and a spectral reflectance representative of non-human skin.

The other method uses only the differential components of spectral reflectances of human skin without an average spectrum, which represent characteristic features of human skin such as congestion and suntan.

Experimental results demonstrate in real time that the use of human/non-human reference spectra distinguishes human skin from non-human objects, while the use of congestion/suntan reference spectra classfies the parts of human body according to the level of congestion and suntan.

8292-27, Session 7

Appearance analysis of human skin with cosmetic foundation

R. Ohtsuki, Kanebo Cosmetics Inc. (Japan) and Chiba Univ. (Japan); S. Tominaga, Chiba Univ. (Japan); R. Hikima, Kanebo Cosmetics Inc. (Japan)

The present paper describes an analysis method of the surface appearance of the make-up face. We propose a method for detecting oily appearance of facial region deteriorated by the sebum. First, a multi-band camera with six spectral channels was used for capturing facial image. Next, we performed subjective assessment of oily appearance for the make-up face. Third, we compared the oily regions with the original regions among the face just after the make-up. It is found that the oily regions are whiter than the original regions, and so the average luminance level is higher than the original. Moreover, the standard deviation of the luminance is larger that the original. Based on these findings, we develop a method for detecting the regions with oily appearance from make-up facial images. First, we extract the candidate regions with oily appearance on the basis of the luminance value Y. The PCA is applied to the luminance and chromaticity image (Y, x, y) of each candidate region. Then a linear cluster of the pixel distribution is formed on the 2D plane. Finally, we can determine the oily appearance by comparison two linear clusters between the original facial image and the image four hour later.

8292-28, Session 7

Color analysis and image rendering of woodblock prints with oil-based ink

T. Horiuchi, T. Tanimoto, S. Tominaga, Chiba Univ. (Japan)

This paper proposes a method for analyzing color characteristics of woodblock prints with oil-based ink and rendering realistic images based on the camera data. The analysis results of woodblock paints show some characteristic features in comparison with oil paintings: (1) Woodblock prints can be divided into some cluster areas with similar surface spectral reflectance, (2) The specular reflection by the influence of overlapping paints is caused only on specific cluster areas. Based on those properties, we develop an algorithm for effective rendering by modifying our previous algorithm for oil paintings. A set of surface spectral reflectances of a woodblock print is represented by using only a small number of the averaged surface spectral reflectances and the registered ratios, while the previous algorithm for oil paintings required high dimensional surface spectral reflectances at all pixels. In the rendering process, we use two sets of parameters of the Cook-Torrance model for cluster areas with/without specular reflection. An

experiment using woodblock paint with oil-based ink is executed for demonstrating the feasibility of the proposed method. It is shown that realistic images of the original woodblock print can be reproduced using a small data set of surface spectral reflectances.

8292-30, Session 8

Pre-RIP color management for soft proofing

I. Tastl, K. Koh, Hewlett-Packard Labs. (United States)

In an ideal situation, a color-managed soft proof is generated by the same raster image processor (RIP) that produces the final raster data for printing. In real-world scenarios, however, this soft proofing capability is often absent. To overcome this limitation, we developed a workflow using the Adobe PDF library for producing color-managed soft proofs before the incoming jobs are sent to the RIP. This paper discusses the color management challenges that presented themselves and how they were addressed. It also highlights a few specific cases where the division of tasks into separate color management and transparency blending steps caused unexpected results and how they can be dealt with. Such a system can be implemented in the form of a cloud-based solution to provide end-customers with a quick, inexpensive, and reasonably accurate preview of the printed artifact before the job is routed to the RIP.

8292-31, Session 8

Parametrically controlled, stochastically seeded clustered halftones

E. A. Bernal, S. Wang, R. P. Loce, Xerox Corp. (United States)

No abstract available.

8292-32, Session 8

Assessing color reproduction tolerances in commercial print workflow

G. B. Beretta, E. Hoarau, S. Kothari, I. Lin, J. Zeng, Hewlett-Packard Labs. (United States)

Color transformations from colorimetric specifications to device coordinates are mostly obtained by measuring a set of samples, inverting the table, and looking up values in the table, and mapping the gamut from input to output device.

The accuracy of a transformation is determined by reproducing a second set of samples and measuring the transformation errors. Accuracy as the average perceptual error is then used as a metric for quality. Accuracy and precision are important metrics in commercial print because a PSP can charge a higher price for more accurate color, or can widen his tolerances when customers want cheap prints.

The disadvantage of determining tolerances through averaging perceptual errors is that the colors in the sample sets are independent and this is not necessarily a good correlate of print quality as determined through psychophysics studies. Indeed, images consist of color palettes and the main quality factor is not color fidelity but color integrity. For example, if the divergence of the field of error vectors is zero, color constancy takes over and humans will perceive the color reproduction as being of good quality, even if the average error is relatively large. However, if the errors are small but in random directions, the perceived image quality is poor.

Our proposal for a standard practice to determine tolerance is to use color discrimination samples for the second set and to evaluate the color transpositions instead of the color differences. There are industry standards for the tolerances of color judges, and the same tolerances and classification can be use for print workflows or its components

(e.g., presses, proofers). We generalize this practice to arbitrary perceptually uniform scales tailored to specific applications and present an implementation.

8292-33, Session 8

Investigation of the paper dependency of laser-printed colors for uncoated papers

S. Gorji Kandi, Institute for Color Science & Technology (Iran, Islamic Republic of)

There are several parameters, which influence the final printed colors in digital printing. One of the most important one is the substrate (paper). The chemical and optical characteristics of the paper may affect the image quality and colorimetric properties of final prints.

In the present study, the effect of uncoated papers on the laser- printed colors is investigated. For this purpose, four kinds of paper with different marks and the same grammage value of 80 was applied. The papers' diversity was almost between their whiteness and texture. The effect of paper was investigated for a Canon Laser Printer LBP-5050.

The effect of paper on the printer's reproducibility, and light-fastness of the printed colors was evaluated for a color chart contains 459 samples, which was generated by Eye-One GretagMacbeth professional maker. The colorimetric characteristics of the final printed samples were studied utilizing a chart of 2125 patches. It was found out that, the reproducibility of the printer is significantly dependent on the paper's type. The average of color change via exposing light is between 1.3 and 1.8 for different papers.

The color difference via the paper change clearly illustrates the effect of paper on the final printed colors. One of the applied papers, which had visually yellow tint and showed the lowest whiteness index (100 CIE units) caused a color difference of 6 CIEDE2000 (1:1:1) units with the others by average. The other papers with a whiteness index between almost 130 and 155 CIE units resulted in a color change about 2.5 to 3 CIEDE2000 (1:1:1) units. The texture properties of the papers were quantified with Edge Frequency technique as an appropriate texture analysis tool. The correlation between texture changes and color differences was also concluded. It appears that the effect of whiteness on the final colors is more than the effect of texture. However, it seems that when the whiteness values reaches to an appropriate range about 145-155 CIE units the effect of whiteness is decreased.

8292-34, Session 9

Color-dependent banding characterization and simulation on natural images

S. Hu, J. Zhang, Purdue Univ. (United States); H. Nachlieli, D. Shaked, Hewlett-Packard Labs. Israel Ltd. (Israel); S. Shiffman, Indigo Ltd. (Israel); J. P. Allebach, Purdue Univ. (United States)

Banding and streaks are important print defects for high-end digital presses. Various mechanisms that cause banding and streaks jointly produce a unique defect appearance from page to page. These defects consist of spatial variations in luminance and chrominance channels, the characteristics of which also depend on the local point within the color space about which the defect variation occurs. This paper focuses on color-dependent characterization and simulation of banding. This framework of banding characterization is based on printing uniform test patches that sample colorant combinations throughout the input color space of the press, scanning these patches, and extracting relevant defect signals. The cloud of the defect signal sample points for each measured colorant combination is analyzed in CIE Lab using 3-dimensional Principal Component Analysis (PCA). Series of features for each measured colorant combination is calculated. This completes the defect characterization framework that can help system designers understand the color-space-dependence of the defects. It also provides the front-end for the framework to simulate the appearance of the defects in the presence of document content. To complete the defect simulation framework, a set of 1-D defect spatial prototypes is selected to represent the common defect appearance

produced by a specific machine. According to the dense defect features LUT obtained by interpolation of the measured defect features, the defect spatial prototype is then modulated in a color-space-dependent fashion that varies spatially across the document content.

8292-35, Session 9

Modeling large-area influence in digital halftoning for electrophotographic printers

Y. Ju, X. Zhang, D. Saxena, Purdue Univ. (United States); T. Kashti, D. Kella, Indigo Ltd. (Israel); D. Shaked, M. Fischer, Hewlett-Packard Labs. Israel Ltd. (Israel); R. A. Ulichney, Hewlett-Packard Co. (United States); J. P. Allebach, Purdue Univ. (United States)

With electrophotography, the deposition of toner within the area of a given printer-addressable pixel is strongly influenced by the halftone values of the immediately neighboring pixels. This local influence can be attributed to fact that the spot size of the laser write beam is larger than a single printer addressable pixel; the complex field interactions that are set up by the charge distribution on the photoconductor and in the toner in the gap between the photoconductor and the developer, and how this influences development; the further spreading of toner during the transfer and fusing processes; and optical scattering of incident light within the media. In our previous work, to predict the measured value of a given printer addressable pixel, we populated a 3 x 3 look-up-table with data based on microanalysis of an exhaustive set of possible printed 3 x 3 binary halftone patterns and summarized the influence of the outer 5 x 5 shell of pixels by forming a weighted sum of these halftone pixels.

In the present paper, we examine the potential influence of a much larger neighborhood of the digital halftone image on the measured value of a printed pixel at the center of that neighborhood. This influence is believed to be due to the scattering of light within the media, rather than the other factors mentioned above. We first demonstrate by printing and analyzing a set of test patterns that indeed the digital halftone within a neighborhood of size on the order of 10s of pixels by 10s of pixels will influence the measured value of the center pixel. We then show that it is possible to account for the influence of this larger neighborhood with a function of only the average value of the digital halftone image within the larger neighborhood. We show that the extended model yields a significant improvement in the accuracy of the prediction of the pixel values of the printed and measured halftone image.

8292-36, Session 9

The octagon screen set: a square N-color, high-order, Moiré-free screen set

Y. Chen, Purdue Univ. (United States); M. Fischer, Hewlett-Packard Labs. Israel Ltd. (Israel); T. Kashti, Indigo Ltd. (Israel); D. Shaked, Hewlett-Packard Labs. Israel Ltd. (Israel); J. P. Allebach, Purdue Univ. (United States)

Periodic clustered-dot screens are widely used for electrophotographic printers due to their print stability. However, moiré is a ubiquitous problem that arises in color printing due to the beating together of the clustered-dot, periodic halftone patterns that are used to represent different colorants. This beating or interference phenomenon introduces spurious low frequency (large period) patterns in the printed output that are very objectionable to the viewer. The traditional solution in the graphic arts and printing industry is to rotate identical square screens to angles that are maximally separated from each other. For example, the classic three-color screen set rotates three identical square screens to the angles 15 degrees, 45 degrees, and 75 degrees, respectively. However, the effectiveness of this approach is limited when printing with more than four colorants, i.e. N-color printing, where N > 4. Moreover, accurately achieving the angles that have maximum angular separation requires a very high resolution plate writer, as is used in commercial offset printing.

In this paper, we propose a systematic way to design color screen sets for periodic, clustered dot screens that offers more explicit control of the moiré properties of the resulting screens when used in color printing applications. This approach is particularly well-suited to the design of tile vector sets for printing with more than four colorants, i.e., N-color printing. We adopt Wang and Loce's method of using a frequency domain basis lattice, but define a specific lattice geometry and a specific halftone frequency vector geometry that have potential advantages. While achieving moiré-free color printing, we also take the symmetry of the resulting halftone patterns into consideration. As a result, in this paper we propose a screen set that consists of completely square screens. We refer to this screen set as the octagon screen set since in the frequency domain, the minimum bounding polygon defined by the fundamental frequencies of the colorants is an octagon. We also propose several symmetry measures, and use them to compare the octagon screen set and the screen sets based on the previous moiré-free N-color non-orthogonal approach. The proposed octagon screen set is shown to have better symmetry properties.

8292-37, Session 9

Ink-saving strategy based on document content characterization and halftone textures

M. V. Ortiz Segovia, Purdue Univ. (United States) and Océ Print Logic Technologies (France); N. Bonnier, Océ Print Logic Technologies (France); J. P. Allebach, Purdue Univ. (United States)

Common ink-saving techniques usually restrict the colorant consumption in a document by replacing a percentage of the colorants by black ink. Even thought such methods achieve a considerable reduction in the amount of colorants used in a page, the visual quality of the printout is affected and unpleasing effects in pastels and skin tones are observed. On the other hand, the quality of the printout is not only affected by the ink-saving algorithm, but also by the way the color halftoning algorithm arranges the dots in the printout. Therefore, the relationship between the contents of the document to be printed and the printing process needs to be addressed by the ink-saving strategy. In order to preserve printing quality, smooth and uniform halftone textures are sought. A color direct binary search halftoning method that strives to minimize both, the ink usage and the perceived error between the continuous-tone color image and the color halftone image is proposed. Our goals are to estimate the effects of the ink-saving module of a printing workflow in individual regions of the document, and to determine the dot arrangement and ink combination that consumes the least amount of ink while preserving printing quality.

8292-38, Session 9

Colour print workflow and methods for multi-layering of colour and decorative inks using UV inkjet for fine-art printing

C. E. Parraman, P. Laidler, Univ. of the West of England (United Kingdom)

The current inkjet market is mainly composed of two dimensional printers, paper and vinyl printing, 3D CAD, solvent wide-format and UV printing onto paper, plastics and board. Inkjet is ubiquitously used as a proofing or prototyping tool in preparation for print production, but it is also being used as the means to print final products and artefacts, for example, printing onto wallpaper, hot-air balloon nylon and furniture. This has been made possible through the more permanent and enduring UV curable inks.

In order to increase density of colour and improve ink coverage when printing onto a range of surfaces, this paper will present research into over-layering of colour and the appearance of colour at 'n' levels of ink coverage. It will also present a study into layering of greys and gloss in order to improve the appearance of printed images onto metal. Returning to our original investigation of artist's requirements when making inkjet prints (1999-), these observations are based on empirical approaches that address the need to present physical data that is

more useful and meaningful to the designer. The study has used colour charts to measure colour to provide users with apriori understanding at a soft-preview level and printed colour circles to demonstrate the appearance of printed colour on different substrates.

Test results relating to the appearance of print on different surfaces, and a series of case studies will be presented.

8292-39, Session 9

Halftone blending between smooth- and detail-screens to improve print quality with electrophotographic printers

S. J. Park, Purdue Univ. (United States); M. Q. Shaw, G. Kerby, T. Nelson, D. Tzeng, V. Loewen, K. Bengtson, Hewlett-Packard Co. (United States); J. P. Allebach, Purdue Univ. (United States)

In this paper, we consider a dual-mode halftoning process that switches between periodic, clustered-dot screens of two different frequencies - a low frequency screen for smooth regions and a high frequency screen for detail regions. These regions are described by an object map that is extracted from a high-level representation of the page content to be rendered. This high-level representation is contained in the page description language (PDL) version of the page to be printed. Our screens obey a harmonic relationship. We implement a blending process based on a transition region. We propose an additive blending process, and a nonlinear blending process in which at each pixel, we choose the maximum of the two weighted halftones. The weights vary according to the position in the transition region, relative to the normal to the boundary between the smooth and detail regions. We show that this blending method better retains the integrity of the two constituent halftone patterns, than does the additive process.

We also show that print quality can be improved by varying the width of the transition region, according to the predicted raggedness of the boundary between the smooth and detail regions. This raggedness measure depends on the relationship between the angles of the smooth and detail screens with respect to the angle of the boundary between the smooth and detail regions. It also depends on the relative gray levels of the image on either side of the boundary. The raggedness measure is computed empirically as a function of these parameters via an off-line training process. Generally speaking, the idea is to use the narrowest transition region that will satisfactorily reduce the raggedness at the boundary. Using a transition region that is wider than this can result in a blurred appearance at the boundary. In addition, we describe an on-line tone-mapping process, based on an off-line calibration procedure that effectively assures the desired tone values within the transition region, and which thereby prevents the occurrence of halo artifacts.

8292-40, Session 10

Spectral transmittance model for piles of transparencies printed in halftone

J. Machizaud, M. Hébert, Lab. Hubert Curien (France) and Univ. Jean-Monnet Saint-Etienne (France) and CNRS (France)

The present work investigates the light transmission property of transparencies printed in halftone. We propose a spectral transmittance model describing the multiple reflections of light between the interfaces of the superposed transparencies as well as the attenuations of light within the plastic bulk and the ink layers. In the case of halftone colors, the model is able to determine the effective surface coverages of the colorants yielded by the partial superposition of the ink dots. Calibration of the model needs printing a few color patches on one transparency and measuring their spectral transmittance. Experimental verification carried out with an inkjet printer shows very good agreement between prediction and measurement in the case of two transparencies. Stochastic halftones are used in order to avoid Moiré effects between the superposed halftone screens. By inversion of the model, we are able to determine the combinations of halftones to print

on the transparencies in order to obtain by superposition one targeted color. An original application of this, called "color matching", consists in creating a pair of color images printed on separated transparencies which yield a uniform colored rectangle once superposed. The prediction accuracy of the proposed model guarantees the good visual uniformity of the resulting colored area.

8292-41, Session 10

Optimal estimation of spectral reflectance based on metamerism

T. Chou, W. Lin, National Taiwan Normal Univ. (Taiwan)

In this paper, we proposed an accurate estimation method for spectral reflectance of objects captured in an image. The spectral reflectance is simply modeled by a linear combination of three basic spectrums of R, G, and B colors respectively, named as spectral reflective bases of objects, which are acquired by solving a linear system based on the principle of color metamerism. Some experiments were performed to evaluate the accuracy of the estimated spectral reflectance of objects. The average mean square error of 24 colors in Macbeth checker between we simulated and the measured is 0.0866, and the maximum is 0.310. In addition, the average color difference of the 24 colors is less than 1.5 under the D65 illuminant. There are 13 colors having their color difference values less than 1, and other 8 colors having the values during the range of 1 and 2. Only three colors are relatively larger, with the differences of 2.558, 4.130 and 2.569, from the colors of No. 2, No. 13, and No. 18 in Macbeth checker respectively. Furthermore, the computational cost of this spectral estimation is very low and suitable for many practical applications in real time.

8292-42, Session 10

Hue-shift model for DLP projector with the white peaking function

I. Park, H. Ha, D. Kim, Y. Ha, Kyungpook National Univ. (Korea, Republic of)

this paper models this hue shift phenomenon and proposes a hue correction method. The modeling of the hue-shift phenomenon is determined based on perceived hue matching experiments. To quantify the hue-shift phenomenon for the whole hue angle, 24 color patches with the same lightness are first created and equally spaced inside the hue angle. These patches are then projected one-by-one on two screens. Next, the hue value for each patch is adjusted by observers using the white peaking function so that the perceived hue for the patches on the two screens appears the same visually. After obtaining the hue-shift values from the perceived hue matching experiment, these values are then used for correction. Essentially, the input red green blue (RGB) values of an image are converted to CIELAB values using a forward characterization model, and the LCh (lightness, chroma, and hue) values are then calculated to obtain the hue values for all the pixels. These hue values are shifted according to the amount calculated using the functions of the hue-shift model. Finally, the corrected CIELAB values are calculated from the lightness, chroma, and corrected hue values, then the output RGB values for all the pixels are estimated using an inverse characterization model. For evaluation, an observer preference test was performed using several test images with other hue shift results, and the compared results are presented with a z score.

8292-43, Session 10

Content-dependent noise reduction for mobile displays

G. Kim, Y. Lee, H. Kim, C. Kim, Inha Univ. (Korea, Republic of)

Mobile imaging systems that can display the contents of the T-DMB have been introduced into markets. Typical size of the T-DMB sequences is QVGA(Quarter Video Graphics Array). In other words,

each frame of T-DMB sequences consists of 320x240 pixels. Also, its frame rate is 30 fps (frames per second). The size of mobile imaging systems are rapidly increasing. Recently, LCD displays whose diagonal size greater than 10 inches are employed for the mobile imaging systems capable of displaying T-DMB sequences. The number of pixels on the large sized mobile displays is about 10 times greater than that of the T-DMB sequences. Thus, the enlargement of more than 3x in both width and height directions is required. Noises resulting from the moving picture compression employed for the T-DMB and the image enlargement procedure affect the visual quality of the T-DMB sequences on the large sized mobile displays. This paper presents a content dependent method to reduce undesirable noises on the enlarged sequences of T-DMB.

Conference 8293: Image Quality and System Performance IX

Tuesday-Thursday 24-26 January 2012

Part of Proceedings of SPIE Vol. 8293 Image Quality and System Performance IX

8293-01, Session 1

Development of the I3A CPIQ spatial metrics

H. Eliasson, Sony Ericsson Mobile Communications AB (Sweden); D. Baxter, STMicroelectronics (R&D) Ltd. (United Kingdom); F. Cao, DxO Labs (France); J. Phillips, Eastman Kodak Co. (United States)

The I3A Camera Phone Image Quality (CPIQ) initiative is aiming at providing a consumer-oriented overall image quality metric for mobile phone cameras. In order to achieve this goal, a set of subjectively correlated image quality metrics has been developed. This paper describes the development of a specific group within this set of metrics, the spatial metrics. Contained in this group are the edge acutance, visual noise and texture acutance metrics. A common feature is that they are all dependent on the spatial content of the specific scene being analyzed. Therefore, the measurement results of the metrics are weighted by a contrast sensitivity function (CSF) and, thus, the conditions under which a particular image is viewed must be specified. This leads to the establishment of a common framework consisting of three components shared by all spatial metrics. First, the RGB image is transformed to a color opponent space, separating the luminance channel from two chrominance channels. Second, associated with this color space are three contrast sensitivity functions for each individual opponent channel. Finally, the specific viewing conditions, comprising both digital displays as well as printouts, are supported through two distinct MTFs.

8293-18, Session 1

A functional-design approach to lens shading correction issues on mobile camera system

S. Yoo, SAMSUNG Electronics Co., Ltd. (Korea, Republic of); T. Kim, SAMSUNG Electronics Co., Ltd. (United States)

This paper proposes the lens shading correction system through combination of functional factors based on recent issues related to mobile camera system. The main issue is designing of the shading algorithm. The second issue is correction of color temperature. The third issue is correction by optical zoom lens. The fourth issue is related with mass production of camera module. The final issue in mobile devices is reducing size of camera module. Previous lens shading correction that situate to stable function through rich hardware resource in DSLR and compact camera for recently the several years is becoming issues again being admitted to part of system of mobile camera that resource is limited. As days go by, size reduction of image sensor and module is essential due to request of marketplace along with the development of mobile devices. The proposed lens shading correction system provides minimizing of accuracy errors in shading evaluation, as well as size reduction of memory and module in hardware. Finally, the proposed system for lens shading correction solves raised issues in mobile environment and guides design of algorithm and hardware for lens shading correction in mobile camera module.

8293-19, Session 1

Rethinking user interfaces for cameraphones

S. A. Brewster, C. McAdam, Univ. of Glasgow (United Kingdom)

Cameraphones are very common but have usability issues that affect their use and reduce the quality of images captured. Users often spend little time post-processing photos, often uploading them immediately to social networking sites. Problems may occur because users look through the LCD to frame shots and often miss icons displayed around the edges of the screen that present important information about camera status.

This may lead to shots being missed, blurred or poorly exposed. Most cameraphones do not take full advantage of the features of the underlying phone platform (such as rich sensing and powerful audio) to try to solve these problems. We created an Android cameraphone application that featured novel interactions and made use of the features of the platform to provide a rich variety of information in more usable forms, such as: sonification of the luminance histogram to ensure better exposure before a picture was taken; measuring phone movement to ensure the phone was being held steady; and the detection of motion in the image to support panning. We evaluated these interactions with users in real settings and showed they could solve some of the most common cameraphone problems and improve the picture taking process.

8293-02, Session 2

Calibration and adaptation of ISO visual noise for camera phone image quality assessment

D. J. Baxter, A. Murray, STMicroelectronics (R&D) Ltd. (United Kingdom)

The I3A Camera Phone Image Quality (CPIQ) visual noise metric described is a core image quality attribute of the wider I3A CPIQ consumer orientated, camera image quality score. This paper describes the motivation behind the choice of a visual noise metric, why the characteristics of cell phone cameras necessitate the adaptation of the ISO 15739 visual noise protocol and the calibration of the adapted visual noise metric in terms of Quality Loss JNDs. Visual noise metrics such as ISO 15739 visual noise, S-CIELAB and vSNR are shown to have the important property of being able to discriminate between different NPS shapes. The optical non-uniformities prevalent in cell phone cameras and higher noise levels pose significant challenges to the ISO 15739 visual noise protocol.

The non- uniformities are addressed using a frequency based high pass filter. Secondly, the data clipping at high noise levels is avoided by use of the Johnson and Fairchild frequency based SCIELAB

Luminance CSF. The final result is a visually based noise metric calibrated in terms of Quality Loss JNDs using Aptina Imaging's subjectively calibrated image set.

8293-03, Session 2

An objective method for evaluating the texture-preserving capability of digital camcorders

K. Zhu, S. Li, D. Saupe, Univ. Konstanz (Germany)

This paper presents a method for evaluating performance of camcorders in terms of texture preservation, taking human visual system (HVS) into account. A metric we call PSDM (power spectral density metric) is the outcome of the method as a quantitative indicator. The relationships between the PSDM value and the motion speed and the compression bitrate were studies. The dead leaves chart was used in our experiments to simulate a scene with textures of different scales. The dead leaves charts is known as a good target invariant to scale, shift, rotation and contrast (exposure) and its radial PSD follows a power law. Experimental results on five camcorders from three different vendors have shown that 1) the PSDM value has a monotonic relationship with the motion speed for all tested camcorders; 2) the PSDM value has a monotonic relationship with the compression bitrates for three tested camcorders but not for the other two. We

attribute the mismatch between the PSDM value and the compression bitrate for some camcorders to the intrinsic nonlinearity of many components in the whole video capturing/recording process. Our study has confirmed PSDM as a useful indicator for measuring a camcorder's performance in terms of preserving textures.

8293-04, Session 2

Improving texture loss measurement: spatial frequency response based on a colored target

U. Artmann, D. Wueller, Image Engineering GmbH & Co. KG (Germany)

The pixel race in the digital camera industry and for mobile phone imaging modules have made noise reduction to a significant part in the signal processing. Depending on the used algorithms and the underlying amount of noise that has to be removed, noise reduction leads to a loss of low contrast fine details, also know as texture loss. The description of these effects became an important part of the objective image quality evaluation in the last years, as the established methods for noise and resolution measurement fail to do so. Different methods have been developed and presented in the last years, but could not fully satisfy the requested stability and correlation with subjective tests. In our paper, we present our experience with the current approaches for texture loss measurement. We have found a critical issue within these methods: the used targets are color neutral. We could show that the test-lab result do not match the real live experience with the cameras under test. We present an approach using a colored target and our experience with this method, based on a huge variety of digital cameras.

8293-20, Session 2

On the performances of computer vision algorithms on mobile platforms

S. Battiato, G. M. Farinella, E. Messina, G. Puglisi, D. Ravì, Univ. degli Studi di Catania (Italy); A. Capra, V. Tomaselli, STMicroelectronics (Italy)

In recent years there is a growing interest in new technology to be employed in the context of mobile devices. Despite today's mobile devices (e.g., smartphone, tablet, etc.) are still limited in terms of resources (e.g., processor speed, available RAM, etc.), novel Computational Photography solutions are available to build appealing imaging applications that cannot be performed before. The main idea is to overcome the limitation of traditional imaging devices by using computational methods which can exploit the different inputs offered by a mobile devices (e.g., from low level data, such as Bayern pattern, GPS position, etc.). Since different cameras are usually embedded in devices of new generation, computer vision algorithms will be extremely useful in many applications of next future. For example, visual tracking can be exploited to interact with video games, or the recognition of the visual content could help in building new applications in the context of cultural heritage (e.g., giving back information on a recognized archeological site).

The main contribution of this work is related to the porting and testing of some classic computer vision algorithms on mobile platforms. Specifically, a few algorithms covering the main tasks of Computer Vision have been considered: keypoint extraction, face detection, image segmentation. The porting has been performed considering the following operating systems: Maemo, typically used in Nokia N900, and Android employed in LG Optimus one, Samsung Galaxy SII. These operating systems have been considered because they can be easily extended with customized libraries and/or programs and provide a standardized and fairly widespread API (Application Program Interface).

It is worth noting that the aforementioned algorithms should be optimized to properly work on low resources devices. For instance, the FCAM library available for N900 Nokia smart phone, allows to interact with the low level algorithms (e.g., demosaicing, white balancing,

denoising, etc.) and data (Bayer pattern) involved in the imaging pipeline. In this way a better design of computer vision algorithms for constrained resources devices can be done.

Finally, comparative tests, conducted on standard datasets, quantitatively and qualitatively evaluate the performance of the algorithms on mobile devices.

8293-05, Session 3

The uncertainty of scanner illumination II

L. C. Cui, Lexmark International, Inc. (United States)

In a preliminary report, we showed the impact of the integrating cavity effect for a typical document scanner with optical ray tracing. The effect was demonstrated by examining the illumination profile after accounting for multiple reflections from the document surface, the contact platen glass surfaces and all reflectors used in the illumination assembly. We identified that the platen glass can contribute just as much as the illumination assembly to the effect. In the second phase, we built an actual scanner model to verify the ray tracing results and the effect. The verification was accomplished by examining the edge profile differences of the scan images of unique patterns before and after a certain reflection component was removed with two different scan configurations. The experimental results are consistent with the simulation results in general.

8293-06, Session 3

Measuring the modulation transfer function of image capture devices: what do the numbers really mean?

X. Zhang, Y. Ju, Purdue Univ. (United States); T. Kashti, D. Kella, T. Frank, Indigo Ltd. (Israel); D. Shaked, Hewlett-Packard Labs. Israel Ltd. (Israel); R. A. Ulichney, Hewlett-Packard Co. (United States); M. Fischer, Hewlett-Packard Labs. Israel Ltd. (Israel); J. P. Allebach, Purdue Univ. (United States)

The modulation transfer function (MTF) is a fundamental tool for assessing the performance of imaging systems. It has been applied to a range of capture and output devices, including printers and even the media itself. In this paper, we consider the problem of measuring the MTF of image capture devices. We analyze the factors that limit the MTF of a capture device. Then, we examine three different approaches to this task based, respectively, on a slant-edge target, a sinewave target, and a grill pattern. We review the mathematical relationship between the three different methods with and without noise in the system, and discuss their comparative advantages and disadvantages. We also consider the impact of linearizing the sensor response on the MTF measurement. Finally, we present experimental results for MTF measurement with a number of different commercially available image capture devices that are specifically designed for capture of 2D reflection or transmission copy. These include camera-based systems, flat-bed scanners, and a drum scanner.

8293-07, Session 3

A fast, automatic camera image stabilization benchmarking scheme

J. Yu, S. A. Craver, Binghamton Univ. (United States)

While image stabilization(IS) has become a default functionality for most digital cameras, there is a lack of automatic IS evaluation scheme, i.e. most publicly known camera IS reviews still require human visual assessment, which is slow and inconsistent. For the first time, we proposed a histogram based automatic IS evaluation scheme, which employs a white noise pattern as shooting target. It is able to produce accurate and consistent IS benchmarks in a very fast manner.

8293-08, Session 3

A comparison of signal-to-noise ratio (SNR) of linear CCD sensors for optical payload

M. Waqas, Pakistan Space and Upper Atmospheric Research Commission (SUPARCO) (Pakistan)

With the growing need and importance of remote sensing data, more and more countries are initiating space missions to fulfill their needs and to benefit from this evolving technology. In this regard, different countries worldwide in general and Asian countries in particular, have initiated high resolution remote sensing satellite projects. The main payload for these satellites is based on high resolution CCD or TDI sensor. Since manufacturing a sensor altogether is not technically and commercially feasible, therefore, we have surveyed various commercially available sensors in order to select a sensor that fulfills the requirements of a high resolution remote sensing mission. In the selection of commercially available sensor, the sensor specifications become important. There are a variety of different sensors commercially available that cover a broad spectrum of remote sensing applications. In order to choose a sensor perfect to user requirements, this paper explores some of the important sensor specifications that drive the selection of a sensor for a high resolution imaging payload; these include: Spectral Range, Quantum Efficiency, Integration Time, Pixel Size, Dark Current, Dynamic Range, SNR etc. This paper presents a comparison of SNR values obtained from different COTS (Commercial off the shelf) based CCD Sensors for an earth imaging payload.

8293-09, Session 4

A sharpness measure on automatically selected edge segments

F. Gasparini, F. Marini, S. Corchs, S. Raimondo, Univ. degli Studi di Milano-Bicocca (Italy)

We address the problem of image quality assessment for natural images, focusing on No Reference (NR) metrics for sharpness.

Among the available methods found in the literature, after detecting the edge pixels, the sharpness measure is defined for each edge pixel. The final metric value is obtained averaging all these values. However, we have observed that in some cases this global measure is not representative of the real sharpness of the images. This fact is mainly due to the averaging process that is applied over all the edge pixels. If the image contains different levels of depth field, this average operation can overestimate the edge spread.

Another important issue in measuring sharpness is to obtain values stable with respect to noise.

In this work, we present an automatic method that selects edge segments, and permits to evaluate image sharpness on more reliable data. Moreover we propose a novel sharpness metric for natural images, inspired by the slanted edge measure adopted by the Imatest in case of synthetic images. This metric makes it possible to cope with noise influence providing more reliable estimations. In the full paper version we will provide examples and experimental results that will confirm the feasibility of our approach.

8293-10, Session 4

Selecting the proper window for SSIM

S. McFadden, Univ. of Waterloo (Canada) and Christie Digital Systems Canada, Inc (Canada); P. A. S. Ward, Univ. of Waterloo (Canada)

The structural similarity index (SSIM) is a popular full-reference metric in the field of image quality assessment (IQA). In spite of its common usage, no analysis has been performed on the effects of modifying the window used in measuring and comparing the luminance, contrast, and structural characteristics of the two images under test.

New experiments involving the modifications of SSIM window characteristics demonstrate a significant impact on metric results. When properly selected, a customized SSIM window can produce superior correlation with subjective results and yield significant gains in computational efficiency relative to the default SSIM window.

Window characteristics relevant to obtaining optimal SSIM results are identified and the effects of their modification demonstrated. SSIM results are obtained using six publicly available subject-rated image quality databases, comprising a total of over 3000 images with various types and levels of distortion. Five commonly used evaluation metrics are computed to demonstrate the improvement of the SSIM algorithm. Analysis of these evaluation metrics is presented to provide insight into the statistical significance of the window modification results, with additional discussion of the computational efficiency improvements. In addition to the full-database outcome presented, results are sub-divided and analyzed according to their specific distortion types.

8293-11, Session 4

Measurement of texture loss for JPEG 2000 compression

P. D. Burns, Carestream Health, Inc. (United States); D. Williams, Image Science Associates (United States)

The capture and retention of image detail is an important characteristic for system design and subsystem selection. An established imaging performance metric that is well suited to certain sources of detail loss, such as optical focus and motion blur, is the Modulation Transfer Function (MTF). As performance standards have developed for digital imaging systems, the MTF concept has been adapted and applied as the spatial frequency response (SFR). Measurement of the SFR is generally done using particular test target features such as edges, repeating patters of square or sign waves.

The use of special image features to derive quality measures is challenged when the effective system characteristics vary with local image (scene) content. This has lead to the development of image quality methods that rely on computed test image content that is some ways resembled natural scenes. An example of this is the measure of texture (image detail) loss using sets of overlapping small objects; circles or rectangles.

In this paper we investigate the application of the above methods to another common adaptive image processing operation - image compression. The derivation of the texture-loss function based on the image noise-power spectrum is described. We then apply several forms of both JPEG and JPEG2000 compression to sets of digital images that include scene content that is amenable to the above standard image quality measures. General conclusions are drawn for application of the proposed image quality measures to digital image compression, as are recommendations for stable estimation of the parameters.

8293-12, Session 4

A no-reference image quality metric for blur and ringing distortions based on weighting process

A. Chetouani, A. Beghdadi, Univ. Paris-Nord (France)

A No Reference Image Quality Metrics (NR-IQM) proposed in the literature are generally developed for a specific degradation, which limits highly its application. To overcome this limitation, we propose in this study a NR-IQM for ringing and blur distortions based on a weighting process. For a given image, we first estimate the quantity of blur and ringing degradations contained in the image using an Artificial Neural Networks (ANN) model through a learning step. Then, the final index quality is given by combining a blur and a ringing metrics according to the obtained weights. The performance of the proposed scheme is evaluated first by testing the weighting process through different specific images. Then the proposed method is evaluated in terms of good correlation with the subjective judgments using the LIVE image database.

Conference 8293

8293-36, Session 4

A new method to identify and quantify image distortion based on Gabor filter bank and multiple regression analysis

B. Ortiz Jaramillo, Univ. Nacional de Colombia (Colombia) and Univ. Gent (Belgium); J. C. Garcia Alvarez, Univ. Nacional de Colombia (Colombia); H. Führ, RWTH Aachen (Germany); S. A. Orjuela Vargas, Univ. Gent (Belgium); G. Castellanos Dominguez, Univ. Nacional de Colombia (Colombia); W. Philips, Univ. Gent (Belgium)

Recently, we have been exploring the possibility to do image quality assessment by using multi-resolution analysis. Thereby, we have proposed an automatic system for full reference image quality assessment based on the Human Visual System model by using multi-resolution analysis. In this sense, previous proposed methodologies shows good performance in objective image quality assessment with goodness of fit between subjective and objectives scores about 92 percent.

Nevertheless, current methodologies are designed assuming that the nature of the distortion is known. But, this is a limiting assumption for practical applications, since in a majority of cases the distortions in the image are unknown.

In this sense, we believe that current studies involving image quality assessment should be addressed to identify and quantify the distortion of images at the same time. Because of perform both studies at the same time could be improve processes like enhancement, restoration, compression, transmission, among others.

We present an approach based on the power of the experimental design to explore effects between components and the joint localization of Gabor filters to split frequency components to identify and quantify distortion in images. With this combination we provide accurate scores and an appropriate differentiability between distortion natures.

8293-37, Poster Session

Optimal patch code design via device characterization

W. Wu, E. N. Dalal, Xerox Corp. (United States)

In many color measurement applications, such as those for color calibration and profiling, "patch code" [1] has been used successfully for job identification and automation to reduce operator errors. A patch code is similar to a barcode, but is intended primarily for use in measurement devices that cannot read barcodes due to limited spatial resolution, such as spectrophotometers. There is an inherent tradeoff between decoding robustness and the number of code levels available for encoding. Previous methods [2, 3] have attempted to address this tradeoff, but those solutions have been sub-optimal. In this paper, we propose a method to design optimal patch codes via device characterization. The tradeoff between decoding robustness and the number of available code levels is optimized in terms of printing and measurement efforts, and decoding robustness against noises from the printing and measurement devices. Effort is drastically reduced relative to previous methods because print-and-measure is minimized through modeling and the use of existing printer profiles. Decoding robustness is improved by distributing the code levels in CIE Lab space rather than in cmyk space.

8293-38, Poster Session

Influence of viewing device and soundtrack in HDTV on subjective video quality

A. Redl, C. Keimel, K. Diepold, Technische Univ. München (Germany)

In many homes today big TV screens and hi-fi systems are common. But is the perception of subjective video quality under professional test conditions the same as in home use?

For this two things are examined: How large is the influence of the presenting device but also the influence of the soundtrack, both in HDTV (1080p50).

Previous works showed that this difference is noticeable, but there were no studies with consumer devices. It was also shown, that there is an influence of the soundtrack, but only in SD or lower resolutions.

Therefore we conducted subjective video tests: One test with different presenting devices, a 23-inch-reference monitor, a high quality 56-inch-LCD-TV and an HD-projector, and one test in which we presented additional to the HD-projector a soundtrack on a 7.1-channel hi-fi system.

The results show two things: The test subjects had a greater quality of experience with the consumer devices than with the reference monitor, although the video quality of the reference monitor itself was rated better in an additional questionnaire and the mean opinion score (MOS). The second result was that there is no significant difference in the MOS between showing the videos on the beamer with or without sound.

8293-39, Poster Session

Influence of viewing experience and stabilization phase in subjective video testing

C. Keimel, A. Redl, K. Diepold, Technische Univ. München (Germany)

In this contribution, we will examine two important aspects of subjective video quality assessment and their overall influence on the test results in detail: the participants' viewing experience and the quality range in the stabilization phase. Firstly, we examined if the previous viewing experience of participants in subjective tests influence the results. We performed a number of single- and double-stimulus tests assessing the visual quality of video material compressed with both H.264/AVC and MPEG2 not only at different quality levels and content, but also in different video formats from 576i up to 1080p. During these tests, we collected additional statistical data on the test participants. Overall, we were able to collect data from over 100 different subjects and analyse the influence of the subjects' viewing experience on the results of the tests. Secondly, we examined if the visual quality range presented in the stabilization phase of a subjective test has significant influence on the test results. Due to time constraints, it is sometimes necessary to split a test into multiple sessions representing subsets of the overall quality range. Consequently, we examine the influence of the quality range presented in the stabilization phase on the overall results, depending on the quality subsets included in the stabilization phase.

8293-40, Poster Session

Perceptual visual image sharpness metric for image-based imager stabilization

F. F. G. Gavant, L. Alacoque, A. Dupret, T. Ho-Phuoc, D. David, CEA-LETI (France)

Image sensors stabilization is usually based on accelerometers. To reduce the number of external components of digital image sensors, an integrated image based image stabilization system is envisaged. Such a system requires modeling the blur due to hand tremor and a general sharpness metric to quantify the gain of such a stabilization system. This paper presents an accurate model of the hand tremor, then its impact as a Point Spread Function. In order to define the specification of the image based image stabilization we have derived perceptual visual quality sharpness metric for camera shake blur. This sharpness metric is based on visual blur test. It fitting well both different approaches such as mean opinion score data base and quality ruler measure of blur.

Conference 8293

8293-41, Poster Session

A unified method for comparison of algorithms of saliency extraction

T. Ho-Phuoc, L. Alacoque, A. Dupret, CEA-LETI (France); A. Guérin-Dugué, Gipsa-lab (France); A. Verdant, CEA-LETI (France)

Extracting salient regions of a still image, which are pertinent areas likely to attract subjects' fixations, can be useful to adapt compression loss according to human attention. In the literature, various algorithms have been proposed for saliency extraction, ranging from region-of-interest (ROI) or point-of-interest (POI) algorithms to saliency models, which also extract ROIs. Implementing such an algorithm within image sensors implies to evaluate its complexity and performance of fixation prediction. However, there have been no pertinent criteria to compare these algorithms in predicting human fixations due to the different nature between ROIs and POIs. In this paper, we propose a novel criterion which is able to compare the prediction performance of ROI and POI algorithms. Aiming at the electronic implementation of such an algorithm, the proposed criterion is based on blocks, which is consistent with processing within image sensors. It also takes into account salient surface, an important factor in electronic implementation, to reflect more accurately the prediction performance of algorithms. The criterion is then used for comparison in a benchmark of several saliency models and ROI/POI algorithms. The results show that a saliency model, which has higher computational complexity, gives better performance than other ROI/POI algorithms.

8293-13, Session 5

Comparative performance analysis of two picture adjustment methods: HSV versus YCbCr

R. Safaee-Rad, M. Aleksic, Qualcomm Inc. (Canada)

Picture adjustments is referred to those adjustemnts that affect the four main subjective perceptual image attributes: Hue, Saturation, Brightness (sometimes called Intensity) and Contrast. The common method used for this type of adjustments in a display processing pipe is based on YCbCr color space. This is expected since usually the incoming content is already in YCbCr color space (TV color space) and thus HSIC adjustments requires less amount of computation.

Picture adjustments using this method, however, leads to multiple problems.

As an alternative, HSV color space (as defined in SIGRAPH 1978) for HSIC picture adjustments is used which leads to multiple advantages.

This paper presents a comparative (picture adjustments) performance analysis based on these two methods.

8293-14, Session 5

Evaluation of preferred lightness rescaling methods for color reproduction

Y. Chang, P. J. Green, London College of Communication (United Kingdom)

Based on the paired comparison experiment result, the Adobe BPC method works well in general with the images. Having very little image dependence on the paired comparison results, there was very little correlation between psychopysical results and the statistical image characteristics analyzed. This suggests that image quality metrics may need to be based on a different set of image characteristics than the ones used in this study.

When the median lightness values are lower than 2.2 L* and higher than 67 L*, linear lightness rescaling method is preferred, while the median lightness values are lower than 5.2 L* and higher than 62.6 L*, spline lightness rescaling method is preferred over the Adobe BPC. As the sample image set was primarily landscape and high contrast, further works need to be performed with larger image set. Also, rather than with image properties, studies on the correlation between image contents and paired comparison results need to be studied in future.

8293-15, Session 5

Investigations of the display white point on the perceived image quality

J. Jiang, F. Abed, J. Voelkel, Rochester Institute of Technology (United States)

Based on the result from a recent web-based experiment conducted to evaluate the perceived image quality without the reference sponsored by Mellon Foundation, a project to understand the impact of the white point setting of the display on the preference judgments of perceived image quality was designed and run, in which two copies of each image were generated to appear either bluish or yellowish. As part of the project, observers were asked to rate the two copies along with the original version of the test image on a display at different white point settings. To include both typical and extreme white point of the display, D50 and D65 were selected besides the two ends of the white point adjustable on a Mac computer. The goal was to determine whether the change in the color of the display would cause any effect on the perceived image quality preferred by observers. In the evaluation of the Analysis of Variance (ANOVA) results, the white point factor was found to be insignificant for the test images in the experiment. Therefore, the perceived image quality based on preference was unlikely to be biased by the color of the display at least for the test images. The result was in agreement with what was found in the Mellon project.

8293-16, Session 6

The mobile image quality survey game

D. R. Rasmussen, Qi Analytics LLC (United States)

In this paper we discuss human assessment of the quality of photographic still images, that are degraded in various manners relative to an original, for example due to compression or noise. In particular, we examine and present results from a technique where observers view images on a mobile device, perform pairwise comparisons, identify defects in the images, and interact with the display to indicate the location of the defects. The technique measures the response time and accuracy of the responses. By posing the survey in a form similar to a game, providing performance feedback to the observer, the technique attempts to increase the engagement of the observers, and to avoid exhausting observers, a factor that is often a problem for subjective surveys. The results are compared with the known physical magnitudes of the defects and with results from similar web-based surveys. The strengths and weaknesses of the technique are discussed. Possible extensions of the technique to video quality assessment are also discussed.

8293-17, Session 6

Evaluation of perceived image sharpness with changes in the displayed image size

J. Park, S. Triantaphillidou, R. E. Jacobson, Univ. of Westminster (United Kingdom)

In this paper an evaluation of the degree of change in the perceived image sharpness with changes in displayed image size was carried out. This was achieved by collecting data from three psychophysical investigations that used techniques to match the perceived sharpness of displayed images of three different sizes. The paper first describes a method employed to create a series of frequency domain Gaussian filters for sharpening and blurring. The filters were designed to achieve one just-noticeable-difference (JND) in sharpness between images viewed from a certain distance and having a certain displayed image size (angle of subtense). During psychophysical experiments, the filtered images were used as a test series for sharpness matching. For test-image capture, a digital SLR camera with a quality zoom lens was

used for recording natural scenes with varying scene content under various illumination conditions. For the psychophysical investigation, a total of sixty-four original test-images were selected and resized, using bi-cubic interpolation, to three different sizes. Results showed that degree of change in sharpness between images of different sizes varied considerably with scene content. Thus, a method, based on the quantification of busyness, to categorize the scene content of the original test images was implemented.

8293-18, Session 6

Towards a perceptual metric for computer-generated images

P. Boulenguez, B. Airieau, M. Larabi, D. Meneveaux, Univ. de Poitiers (France)

Although a significant research effort has been dedicated to the physical validation of Computer-Generated Images (CGIs), few studies addressed the problem from a psycho-visual perspective. This paper engages in the construction of such a perceptual metric for CGIs. To that purpose, an original psycho-visual experiment has been designed and submitted to a representative panel of observers. Each participant was asked to score the overall perceived quality and five perceptual criteria on CGIs rendered with seven different algorithms. An ad hoc analytical quality model, fit to the statistical data, then gives insight into the influence of each criterion on the overall perceived quality. Results tend to show that accuracy in the simulation of shadows, good contrast and absence of noise have a major impact on the overall perceived quality, rather than precise anti-aliasing or faithfull color bleeding.

8293-19, Session 6

Assessing product image quality for online shopping

A. Goswami, S. H. Chung, N. Chittar, A. Islam, eBay Inc. (United States)

Assessing product-image quality is important in the context of online shopping. A high quality image that conveys more information about a product can boost the buyer's confidence and can get more attention. However, the notion of image quality for product-images is not the same as that in other domains. The perception of quality of product-images depends not only on various photographic quality features but also on various high level features such as clarity of the foreground or goodness of the background etc. In this paper, we define a notion of product-image quality based on various such features. We conduct a crowed sourced experiment to collect user judgments on thousands of eBay's images. We formulate a multi-class classification problem for modeling image quality by classifying images into good, fair and poor quality based on the guided perceptual notions from the judges. We then compute a pseudo-regression score with expected average of predicted classes. We design many experiments with various sampling and voting schemes with crowd sourced data and construct various experimental image quality models. Most of our models have reasonable accuracies (greater or equal to 70%) on test data set. We observe that our computed image quality score has a high (0.66) rank correlation with average votes from the crowd sourced human judgments.

8293-20, Session 7A

How do we watch images?: a case of change detection and quality estimation

J. Radun, T. Leisti, T. Virtanen, G. Nyman, Univ. of Helsinki (Finland)

The most common tasks in subjective image estimation are change detection (a detection task) and image quality estimation (a preference task). We examined how the task influences the gaze behavior when comparing detection and preference tasks. The eye movements of

16 naïve observers were recorded with 8 observers in both tasks. The setting was a flicker paradigm, where the observers see a non-manipulated image, a manipulated version of the image and again the non-manipulated image and estimate the difference they perceived in them. The material was photographic material with different image distortions and contents. To examine the spatial distribution of fixations, we defined the regions of interest using a memory task and calculated information entropy to estimate the spatial concentration of fixations. The quality task was faster and needed fewer fixations and the first eight fixations were more concentrated on certain image areas than in the change detection task. The bottom-up influences of the image also caused more variation to the gaze behavior in the quality estimation task than in the change detection task. The results show that the strategies for making the tasks are different and in subjective image estimation studies it is important to think about the task.

8293-21, Session 7A

Measuring saliency in images: which experimental parameters for the assessment of image quality?

C. Fredembach, G. Woolfe, J. Wang, Canon Information Systems Research Australia Pty. Ltd. (Australia)

Predicting which areas of an image are perceptually salient or attended to has become an essential pre-requisite of many computer vision applications. Because observers are notoriously unreliable in remembering where they look a posteriori, and because asking where they look while observing the image necessarily influences the results, ground truth about saliency and visual attention has to be obtained by gaze tracking methods.

From the early work of Buswell and Yarbus to the most recent forays in computer vision there has been, perhaps unfortunately, little agreement on standardisation of eye tracking protocols for measuring visual attention. As the number of parameters involved in the experimental methodology can be large, their individual influence on the final results is not well understood. Consequently, the performance of saliency algorithms, when assessed by correlation techniques varies greatly across the literature.

In this paper, we concern ourselves with the problem of image quality. Specifically: where people look when judging images. We show that in this case, the performance gap between existing saliency prediction algorithms and experimental results is significantly larger than otherwise reported. To understand this discrepancy, we first set out to devise an experimental protocol that is adapted to the task of measuring image quality. In a second step, we compare our experimental parameters with the ones of existing methods and show that a lot of the variability can directly be ascribed to these differences in experimental methodology and choice of variables.

In particular, the choice of a task: judging image quality vs. free viewing has a great impact on measured saliency maps, suggesting that even for a mildly cognitive task, ground truth obtained by free viewing does not adapt well. Careful analysis of the prior art also reveals that systematic bias can occur depending on instrumental calibration and the choice of test images.

We conclude this work by proposing a set of parameters, tasks and images that can be used to compare the various saliency prediction methods in a manner that is meaningful for image quality assessment.

8293-29, Session 7A

Examining the effect of task on viewing behavior in videos using saliency maps

H. Alers, J. A. Redi, Technische Univ. Delft (Netherlands); I. Heynderickx, Philips Research (Netherlands)

This work is aimed at understanding how the viewing task and level of video quality affect the viewing behavior of the observer. To test that, a database of 50 videos of 20 seconds each was created. These videos were degraded to different levels of quality. Consequently, observers

were asked to watch these videos where half of them were given the task of scoring the quality level while the rest were told to simply watch the videos freely. By using an eye tracking device, it was possible to record the viewing behavior data of the observers. This data was used to create saliency maps for each second of video, which made it possible to compare the viewing behavior between the two tasks across different levels of quality. The results suggest that there is a difference in the viewing behavior depending on the task given to the viewer which is not effected with the quality level of the video. It is also shown that there are specific attributes in the videos, which affect the level of similarity in viewing behavior between the two tasks.

8293-23, Session 7B

Investigations of the tone reproduction curves on the perceived image quality for fine art reproductions

J. Jiang, F. Frey, S. Farnand, Rochester Institute of Technology (United States)

Based on the result from a recent web-based experiment conducted to evaluate the perceived image quality without the reference sponsored by Mellon Foundation, a project is being conducted to understand the effect of the contrast adjustment made in the process of fine art reproduction on the perceived image quality. Contrasts are usually adjusted using the tone reproduction curve (TRC) in order to achieve the highest possible dynamic range and to better represent the original artwork. Copies of fine art images are simulated using tone reproduction curves collected from different museums. During the experiment, observers are instructed to rate each test image based on their preference in a web-based environment. Observers will be asked to click on the area that most influences their decisions. More attention may be needed to the areas on which most clicks are made when the tone reproduction curve is adjusted during visual editing. To understand the impact of the original on participants' preference judgments, the experiment will be conducted both with and without the reference. The result may be indicative of whether a separate TRC is needed for fine art reproductions depending on the availability of the original when the reproduced image is viewed.

8293-30, Session 7B

Characterizing eye movements during temporal- and global-quality assessment of h.264 compressed video sequences

C. Mantel, N. Guyader, P. Ladret, G. Ionescu, Gipsa-lab (France); T. Kunlin, STMicroelectronics (France)

Studies have shown that the deployment of visual attention is closely link to the assessment of image or video quality, though this link is not yet fully understood. The influence of rating temporal quality of compressed videos over the way an observer deploys his attention is investigated in this paper.

We set-up a subjective experiment in which the eye movements of observers are recorded during three different tasks: a free-viewing task (FT), a global quality assessment task and a temporal quality assessment task. The FT acts as a reference to which we compare the eye movements during the two other tasks.

As previously shown, observers assessing global quality gaze at locations dissimilar to those fixated during the FT. For temporal quality assessment, it seems that the fixated locations are closer to FT than the global quality assessment fixated locations.

Our results suggest that the locations observers look at do not depend on the displayed video quality level. Quality however influences the way participants look at videos: the lower the quality, the longer they gaze at a precise location. The area fixated seems to be much smaller during the quality assessment tasks than during the FT for either perfect or poor quality level.

The evolution over time of all indicators suggests that, during the first 1 or 2 seconds, the signal properties of the videos are the main attractors for the participants' eye movements. Instructions only seem to play a role afterwards on the deployment of the participants' visual attention.

8293-31, Session 7B

A compressed sensing model of crowding in peripheral vision

J. Hocke, Univ. zu Lübeck (Germany); M. Dorr, Schepens Eye Research Institute (United States); E. Barth, Univ. zu Lübeck (Germany)

We here model peripheral vision in a compressed sensing framework and find that typical letter-crowding effects arise naturally. Images are re-sampled with samples that are distributed according to a Poisson-disk distribution. These samples are then considered to be inputs to a network of neurons such that the output neurons are randomly connected to the input neurons. Also the weights of the connections are assigned randomly. Convergence is modeled by allowing for fewer output neurons than input neurons. By reconstructing the original image from the activity of the output neurons we estimate the amount of information delivered at the output. To reconstruct the original image, the signal at the output neurons is modeled as a linear superposition in a sparse basis and a sparse-approximation algorithm is applied. The reconstructed images exhibit typical crowding artifacts in the sense that single letters can be recognized but not letters which are flanked by additional letters. When the sampling rate is increased, as would be the case in central vision, both the single and the flanked letters can be recognized.

8293-32, Session 7B

Foveated self-similarity in nonlocal image filtering

A. Foi, Tampere Univ. of Technology (Finland); G. Boracchi, Politecnico di Milano (Italy)

We investigate the role of foveation in nonlocal image filtering and, in particular, in the NLmeans denoising algorithm. The nonlocal self-similarity between regions at different image locations is used in NLmeans to adaptively determine the weights for pointwise estimation. This similarity between different regions is computed through the windowed photometric differences, where the window function decays to zero as we move away from center of the corresponding region.

We replace such windowed difference with a foveated difference: instead of multiplying the photometric differences against a window function, we blur the differences with point-spread functions having increasing standard-deviation (i.e. increasing spread) as the distance from the region's center grows. This foveated difference mimics the actual ability of the human visual system to perceive details at the periphery of the point of fixation.

Our study show that a foveated self-similarity leads to an improvement in the resulting image estimate, according both to objective criteria and visual appearance, particularly due to better contrast and sharpness.

We discuss this form of self-similarity in the context of natural image statistics as well as its potential implications in explaining the role of eye movements in improving visual perception.

8293-33, Session 7B

A statistical study of the correlation between interest points and gaze points

M. Nauge, M. Larabi, Univ. de Poitiers (France)

In this work, we propose to study the performance of these detectors with a totally different point of view, by studying similarity/correlation between the subjective gaze points and the objective interest points. The aim is to determine whether the interest points can be used to predict salient information on an image like the HVS does. This can help for several applications like quality assessment, simplified saliency

maps construction, . . . Even though the interest points have not been originally designed to be close to the gaze points, they may have a particular setting that maximizes the similarity between them. For this study a battery of statistical tools is used to test a large range of settings/configurations for the three interest point detectors mentioned previously.

We used several statistical tools such as Bartlett, ANOVA, . . . to understand the effect and the influence of each parameters for each detector. These studies illustrate that particular parameters can minimize the cost of transformation and predict interest points in accordance with the subjective gaze points. We also proposed a solution to give a scale to facilitate the interpretation of the EMD values by analyzing the mean human behavior. By comparing the best setting for each detector we can also indicate which detector is the most reliable to estimate the subjective gaze points. This study is also a good way to prove than interest points detector share some properties with the HVS.

8293-34, Session 7B

Interest point analysis as a model for the Poggendorff illusion

F. W. M. Stentiford, Univ. College London (United Kingdom)

This paper describes a recognition mechanism based on the relationships between interest points and their properties that is applied to the problem of modelling the Poggendorff illusion. The recognition mechanism is shown to perform in the same manner as human vision on the standard illusion and reduced effects are modelled on a variant without parallels. The results show that the recognition mechanism produces high similarity scores between continuous transversal lines and Poggendorff figures where the line segments are offset from the collinear position. The peaks occur in the same direction of shift as reported in psychophysical investigations of human vision. The most prominent peaks occur with transversals inclined at the smallest angles to the vertical and decrease towards the perpendicular position (90°). The analysis of figures with no parallels also yielded a peak although this was not as prominent as with the conventional Poggendorff figures. This was also in agreement with earlier behavioural experiments. The analysis is also shown to be consistent with psychophysical findings where obtuse angled variants of the Poggendorff figures yield stronger effects than the acute angled variants.

8293-24, Session 8

A method of detecting changes in image quality via sensing on customer documents

W. Wu, B. Xu, J. C. Handley, Xerox Corp. (United States)

It is of great value to be able to track image quality of a printing system and detect changes before/when it occurs. To do that effectively, image quality data need to be constantly gathered and processed. A common approach is to print and measure test-patterns over-time at a pre-determined schedule and then analyze the measured image quality data to discover/detect changes. But due to the presence of other printer noises, such as page-to-page instability, mottle etc., it is likely that the measured image quality data for a given image quality attribute of interest (e.g. streak) at a given time is governed by a statistical model rather than a deterministic one. This imposes difficulty for methods intended to detect image quality changes reliably unless sufficient data of test samples are collected. However, these test samples are nonvalue- add to the customers and should be minimized. An alternative is to directly measure and assess the image quality attributes of interest from customer pages and post-processing them for detecting changes. In addition to the difficulty caused by sources of other printer noise, variable image contents from customer pages further impose challenges in the change detection. This paper addresses these issues and presents a feasible solution in which change points are detected by statistical model-ranking.

8293-25, Session 8

Print quality analysis for ink-saving algorithms

M. V. Ortiz Segovia, Purdue Univ. (United States) and Océ Print Logic Technologies (France); N. Bonnier, Océ Print Logic Technologies (France); J. P. Allebach, Purdue Univ. (United States)

Ink-saving strategies for CMYK printers have evolved from their earlier stages where the 'draft' print mode was the main option available to control ink usage. The savings were achieved by printing alternate dots in an image at the expense of reducing print quality considerably. Nowadays, customers are not only unwilling to compromise quality but have higher expectations regarding both visual print quality and ink reduction solutions. Therefore, the need for more intricate ink saving solutions with lower impact on print quality is evident. Printing-related factors such as the way the printer places the dots on the paper and the ink-substrate interaction play important and complex roles in the characterization and modeling of the printing process that make the ink reduction topic a challenging problem. In our study, we are interested in finding the connections between different ink reduction levels of a given ink-saving algorithm and print quality attributes such as sharpness, artifacts, and color saturation.

8293-26, Session 8

Masking mediated print defect visibility predictor

X. Jing, Purdue Univ. (United States); H. Nachieli, D. Shaked, S. Shiffman, Hewlett-Packard Labs. Israel Ltd. (Israel); J. P. Allebach, Purdue Univ. (United States)

In commercial printing contexts, it is quite reasonable that customers expect good quality printouts. Therefore, print shops need to design their workflow to pay attention to this important aspect. This makes print quality assessment quite important for developers of printing systems. However, there are not many well-developed integrated measurements of print quality. There also has been a lack of attention in the existing models to the effect of content-masking on the visibility of print defects. Furthermore, such a print quality assessment should truly represent the viewers' observations. Therefore, generating an integrated print quality model which can predict viewers' observations is difficult but greatly needed.

Compared with other image and print quality research, our work is focused on the print quality of typical printouts printed on a digital commercial printing press. In this paper, we propose a Masking-Mediated Defect Visibility Predictor (MMDVP) to predict the visibility of defects in the presence of customer content. The parameters of the algorithm are trained from ground-truth images that have been marked by subjects. The MMDVP could help the press operator decide whether the print quality is acceptable for specific customer requirements. Ultimately, this model can be used to optimize the print-shop workflow.

8293-27, Session 8

Psychophysical evaluation of banding visibility in the presence of print content

J. Zhang, Purdue Univ. (United States); D. Shaked, H. Nachlieli, S. Shiffman, Hewlett-Packard Labs. Israel Ltd. (Israel); J. P. Allebach, Purdue Univ. (United States)

Observing and evaluating print defects represents a major challenge in the area of print quality research. Visual identification and quantification of print defects becomes a key issue for improving print quality. However, the page content may confound the visual evaluation of print defects in actual printouts. Our research is focused on banding in the presence of print content in the context of commercial printing. In this paper, a psychophysical experiment is described to evaluate the perception of bands in the presence of print content. A number of banding defects are added by way of simulation to a selected set of

commercial print content to form our set of stimuli. The participants in the experiment mark these stimuli based on their observations via a graphical user interface (GUI). Based on the collection of the marked stimuli, we were able to see general consistency among different participants. Moreover, the results showed that the likelihood of an observer perceiving the banding defect in a smooth area is much higher than in a high frequency area. Furthermore, our results also indicate that the luminance of the image may locally affect the visibility of the print defects to some degree.

8293-28, Session 9

No-reference video quality assessment of H.264 video streams based on semantic saliency maps

H. Boujut, J. Benois-Pineau, T. Ahmed, Bordeaux Univ. (France); O. Hadar, Ben-Gurion Univ. of the Negev (Israel); P. Bonnet, Audemat Worldcast Systems (France)

The paper contributes to No-Reference video quality assessment of broadcasted HD video over IP networks and DVB. The NR-quality assessment research is of primarily importance for the community because of the wide range of applications and the inherent difficulty of the task. We propose to enhance the visual saliency model of our No-Reference video quality assessment metric Weighted Macro-Block Error Rate (WMBER) based on saliency maps and macro-block error detection. This enhancement of visual saliency is performed by considering the influence of semantics of the visual scene in the visual attention. Recent works have showed that saliency maps are well suited to measure the perceived quality in the context of lossy video broadcasting networks. However visual saliency models are mainly based on a bottom-up approach which does not take into account the semantics of the visual scene. New works on modeling visual attention have shown that semantics increase visual attention especially on faces. It was observed that areas which contain faces grab the attention 16.6 times more than areas without faces and with the same saliency. Furthermore, S. Hemami and A. Reibman also stress that the perceived quality depends on the usefulness of the content. Thus, in this work, we decide to stay focused on face detection and tracking to enhance our bottom-up visual attention model.

8293-29, Session 9

Linking quality assessment of free-viewpoint video objects up with algorithm development

S. Kepplinger, Technische Univ. Ilmenau (Germany)

This paper presents a new approach using an extended model for the linkage of Quality of Experience with the technical realization, supporting algorithmic developments in the field of free viewpoint video. The interlinking model of Quality of Experience with the technical realization is implemented by taking into account subjective evaluation results as well as possible variances of algorithmic processes. This is used to support the scalability and adaptability of the system based on the end users' requirements. The extended model is defined after a detailed literature review showing the lack of an adequate way to link Quality of Experience with algorithmic developments, taking into account existing approaches. The model includes prior evaluation activities on the subjective quality assessment of free viewpoint video objects used within the context of video communication to support eye contact. However, quality estimation in this particular use case has not been covered yet, and adequate approaches are missing. A methodological approach to define quality influencing factors, and its results, will be presented. A detailed description of the interlinking model taking into account these results will be given, and a way of weighting quality influencing factors will therefore be outlined.

8293-30, Session 9

A perceptual optimization of H.264/AVC bit allocation at the frame and macroblock levels

M. Hrarti, H. Saadane, M. Larabi, XLIM-SIC (France)

Rate control is a key technology for an efficient encoder to achieve a high and consistent quality for the whole video sequence under the channel bandwidth and delay/buffer constraints. In general, rate control includes two parts; the bit allocation used to estimate target bits for encoding the current frame or macroblock; and the QP (Quantization Parameter) determination for achieving accurate bits allocation.

In the H.264/AVC rate control algorithm, the bit allocation process and the QP determination are not optimal. In a previous work we have proposed Rate-Quantization (R-Q) models for Intra and Inter frames used to deal with the QP determination shortcoming. In this paper, we look to overcome the limitation of the bit allocation process at the frame and the macroblock layers. At the frame level, we enhance the bit allocation process by exploiting frame complexity measures. At macroblock level, the visual saliency is used in the bit allocation process. The basic idea is to promote salient macroblocks. Hence, a saliency map, based on a Bottom-Up approach, is generated and a macroblock classification is performed. This classification is then used to accurately adjust UBitsH264 which represents the usual bit budget estimated by H.264/AVC bit allocation process. Simulations have been carried out using JM15.0 reference software, several video sequences and different target bit-rates. In comparison with JM15.0 algorithm, the proposed approach improves the coding efficiency in terms of PSNR (up to +2.05dB), and SSIM (up to 4%). Furthermore, the bandwidth constraint is always satisfied because the actual bit-rate is always lower than or equal to the target bit-rate.

8293-31, Session 9

QoE assessment method for mobile video services based on user motivation

F. Kobayashi, M. Masuda, T. Hayashi, NTT Network Innovation Labs. (Japan)

To ensure that video communication services meet the high expectations of end users, user quality of experience (QoE) must be properly considered. Therefore, various QoE assessment methods for video services have been proposed. However, several QoE assessment methods based on user motivation shows video quality is not the only perspective for QoE. To assess user expectation for video communication services, we have to obtain "user preferences" in which user interest in a video must be considered in addition to video quality, motivation, and level of motivation achievement. Additionally, we consider multiple QoEs, such as preference and motivation, which vary with each participant. We propose a QoE assessment method for mobile video services. We give participants motivation before they watch videos on mobile terminals. After watching, participants assess QoE for video quality, motivation achievement level, and user preference. Simultaneously, participants assess their expectations which are integrated QoEs. We conducted an experiment using proposed method. As a result, we concluded that considering user preference is important for QoE assessment method bases on motivation. We also clarified that the required video quality level to meet certain user expectations is different depending on the classification of participants.

8293-32, Session 9

Quality rules for detection, recognition, and identification in video-surveillance applications

M. Larabi, D. Nicholson, Univ. de Poitiers (France)

This work is focusing on the definition of a procedure for the

qualification of coding schemes for video surveillance applications. It consists in developing and benchmarking tools that learn from the expertise of police and security department. This expertise is intended to be modeled thanks to a campaign of subjective measurement allowing to analyze the way they are using in performing the security tasks like face or license plate recognition, event detection and so on. The results of the previous test are used will be used to tune and to construct a hybrid metric based on basic artifacts detection due to compression and transmission.

8293-33, Session 10

A learning-based approach for automated quality assessment of computer-rendered images

X. Zhang, G. Agam, Illinois Institute of Technology (United States)

Computer generated images are common in numerous computer graphics applications such as games, modeling, and simulation. There is normally a tradeoff between the time allocated to the generation of each image frame and and the quality of the image, where better quality images require more processing time. Specifically, in the rendering of 3D objects, the surfaces of objects may be manipulated by subdividing them into smaller triangular patches and/or smoothing them so as to produce better looking renderings. Since unnecessary subdivision results in increased rendering time and unnecessary smoothing results in reduced details, there is a need to automatically determine the amount of necessary processing for producing good quality rendered images. In this paper we propose a novel supervised learning based methodology for automatically predicting the quality of rendered images of 3D objects. To perform the prediction we train on a data set which is labeled by human observers for quality. We are then able to predict the quality of renderings (not used in the training) with an average prediction error of roughly 20%. The proposed approach is compared to known techniques and is shown to produce better results.

8293-34, Session 10

A comparison of techniques for superresolution evaluation

M. A. Trifas, Jacksonville State Univ. (United States)

Super resolution approaches can be classified in to two main categories: those which infer data and those which simply piece together existing data. Both approaches produce an approximation of the actual high resolution image which may suffer from false positive inclusion and false negative exclusion. Super resolution techniques are generally evaluated either visually or on a pixel-by-pixel comparison basis. The visual approach suffers from both experimenter and respondent biases and, while the pixel-comparison approach is easily understandable and computationally simple, it is quite problematic. Specifically, it does not effectively evaluate super resolution performance for many applications and it is susceptible to noise created by small (and in many cases irrelevant) pixel-value differences.

Testing was conducted with an inference engine which enhances using a database of patterns which it is trained with prior to enhancement-presentation. Three techniques for image evaluation were conducted on each super-resolved image. Super-resolved images were compared visually to the high-resolution originals, pixel-by-pixel difference was calculated and threshold-based pixel-by-pixel comparison (where error within a application-acceptable margin is not averaged in) was also computed. The unsuitability of pixel-by-pixel comparison for evaluating super resolution techniques for many applications was demonstrated.

8293-35, Session 10

Detection of image quality metamers based on the metric for unified image quality

K. Miyata, National Museum of Japanese History (Japan); N. Tsumura, Chiba Univ. (Japan)

In this paper, we propose a criterion showing the degree of the image quality metamerism, named as Metric for Unified Image Quality (MUIQ), to measure overall image quality as a function of image quality components, then MUIQ is applied to detect metameric image quality areas as an application for historical materials. Two colored areas in the image are evaluated by MUIQ and an objective graininess factor Gr, then if MUIQ is the same but Gr is different, the area is classified into the metameric image quality area. The image quality metamerism is an expanded concept of the conventional metamerism defined in the color science. The material investigated in this paper is the oldest extant version of folding screen paintings that depict the thriving city of Kyoto designated as a nationally important cultural property in Japan. The local MUIQ and Gr are calculated to detect gold colored areas in the folding screen as the image quality metamers, which are painted by using high granularity pigments compared with other color areas. As a result of this research, the gold colored areas were detected better than the method used in our previous method.

Conference 8294: Visualization and Data Analysis 2012

Monday-Wednesday 23-25 January 2012

Part of Proceedings of SPIE Vol. 8294 Visualization and Data Analysis 2012

8294-01, Session 1

StreamSqueeze: a dynamic stream visualization for monitoring of event data

F. Mansmann, M. Krstajic, F. Fischer, E. Bertini, Univ. Konstanz (Germany)

While in clear-cut situations automated analytical solution for data streams are already in place, only few visual approaches have been proposed in the literature for exploratory analysis tasks on dynamic information. However, due to the competitive or security-related advantages that real-time information gives in domains such as finance, business or networking, we are convinced that there is a need for exploratory visualization tools for data streams. Under the conditions that new events have higher relevance and that smooth transitions enable traceability of items, we propose a novel dynamic stream visualization called StreamSqueeze. In this technique the degree of interest of recent items is expressed through an increase in size and thus recent events can be shown with more details. The technique has two main benefits: First, the layout algorithm arranges items in several lists of various sizes and optimizes the positions within each list so that the transition of an item from one list to the other triggers least visual changes. Second, the animation scheme ensures that for 50 percent of the time an item has a static screen position where reading is most effective and then continuously shrinks and moves to the its next static position in the subsequent list. To demonstrate the capability of our technique, we apply it to large and high-frequency news and syslog streams and show how it maintains optimal stability of the layout under the conditions given above.

8294-02, Session 1

Interactive data-centric viewpoint selection

H. S. Kim, D. Unat, S. B. Baden, J. P. Schulze, Univ. of California, San Diego (United States)

We propose a new algorithm for automatic viewpoint selection. While most previous algorithms depend on information theoretic frameworks, our algorithm solely focuses on the data itself without any rendering steps, and finds a view direction that best describes the data. The algorithm consists of two main steps: feature selection and viewpoint selection. The feature selection step is an extension of the 2D Harris interest point detection algorithm. This step selects corner and/or high-intensity points as features, which captures the overall structures and local details. The second step, viewpoint selection, takes this set and finds a direction that lays out those points in a way that the variance of projected points of them is maximized, which can be formulated as a Principal Component Analysis (PCA) problem. The PCA solution guarantees that the surfaces captured by corner points are less likely to be degenerative, and that the local details have minimum occlusion among them. The entire algorithm takes under a second for various volume data sets, which allows it to be integrated into real-time volume rendering applications where users can modify the volume with transfer functions.

8294-03, Session 1

Interactive analysis of situational awareness metrics

D. R. Overby, J. A. Wall, J. Keyser, Texas A&M Univ. (United States)

Digital systems are employed to maintain situational awareness of people in various contexts including emergency response, disaster relief, and military operations. Because these systems are often operated in wireless environments and are used to support real-time decision making, the accuracy of the data provided is important to measure and evaluate. Our work has been conducted in conjunction with analysts in the evaluation and performance comparison of different systems designed to provide situational awareness in coordinated operations. To this end, we defined temporal and spatial metrics for measuring the accuracy of the SA data provided by each system. In this paper we discuss the proposed temporal and spatial metrics for SA data and show how we provided these metrics in a linked coordinated multiple view environment that enabled the analysts to effectively perform critical analysis tasks. The temporal metric is used to determine when network performance has a significant effect on SA data, and therefore identify specific time periods in which individuals were provided inaccurate data. Temporal context can be used to determine the local or global nature of the inaccuracy, and the spatial metric can then be used to identify geographic effects on network performance of the wireless system.

8294-04, Session 2

Incremental visual text analytics of news story development

M. Krstajic, Univ. Konstanz (Germany)

Online news sources produce thousands of news articles every day, reporting on local and global real-world events. These events are represented by topics, i.e. event episodes that have to be extracted from the unstructured time-stamped data that continuously arrive over time. News information streams are potentially unbounded in size, making understanding topic temporal dynamics and relationships a challenging task. In this paper, we present a visual analytics system for exploration of news topics in dynamic information streams, which combines interactive visualization and text mining techniques to facilitate the analysis of similar topics that split and merge over time. We employ document clustering techniques to extract news stories that evolve over time in an online fashion and an incremental visualization is designed to represent temporal characteristics of the stories in different time frames. By using interaction, stories can be filtered and explored in full detail with details on demand. To demonstrate the usefulness of our system, case studies with real news data are presented and show the capabilities for detailed dynamic text stream exploration.

8294-05, Session 2

Guided text analysis using adaptive visual analytics

C. A. Steed, C. T. Symons, F. A. DeNap, T. E. Potok, Oak Ridge National Lab. (United States)

This paper demonstrates the promise of augmenting interactive visualizations with semi-supervised machine learning techniques to improve the discovery of significant associations and insight for searching textual information. More specifically, we have developed a system-called Gryffin-that hosts a unique collection of techniques that facilitate individualized investigative search pertaining to an ever-changing set of analytical questions over an indexed collection of open-source publications related to national infrastructure. The Gryffin client hosts dynamic displays of the search results via focus+context record listings, temporal timelines, term-frequency views, and multiple coordinated views. Furthermore, as the analyst interacts with the display, the interactions are recorded and used to label the search records. These labeled records are then used to drive semi-supervised machine learning algorithms that re-rank the unlabeled search records

such that potentially relevant records are moved to the top of the record listing. Gryffin is described in the context of the daily tasks encountered at the Department of Homeland Security's Fusion Centers, with whom we are collaborating in its development. The resulting system is capable of addressing the analyst's information overload that can be directly attributed to the deluge of information that must be addressed in search and investigative analysis of textual information.

8294-06, Session 3

Designing a better weather display

C. Ware, M. Plumlee, The Univ. of New Hampshire (United States)

The variables most commonly displayed on weather maps are atmospheric pressure, wind speed and direction, and surface temperature. But they are usually shown separately, not together on a single map. As a design exercise, we set the goal of finding out if it is possible to show all three variables (two 2D scalar fields and a 2D vector field) simultaneously such that values can be accurately read using keys for all variables, a reasonable level of detail is shown, and important meteorological features stand out clearly. Our solution involves employing three perceptual "channels", a color channel, a texture channel, and a motion channel in order to perceptually separate the variables and make them independently readable. We conducted an experiment to evaluate our new design both against a conventional solution, and against a glyph-based solution. The evaluation tested the abilities of novice subjects both to read values using a key, and to see meteorological patterns in the data. Our new scheme was superior especially in the representation of wind patterns using the motion channel, and it also performed well enough in the representation of pressure using the texture channel to suggest it as a viable design alternative.

8294-07, Session 3

Visualization feedback for musical ensemble practice: a case study on phrase articulation and dynamics

T. Knight, N. Boulliot, J. Cooperstock, McGill Univ. (Canada)

We consider the possible advantages of visualization in supporting musical interpretation. Specifically, we investigate the use of visualizations in making a subjective judgement of a student's performance compared to reference "expert" performance for particular aspects of musical performance---articulation and dynamics. Our assessment criteria for the effectiveness of the feedback are based on the consistency of judgements made by the participants using each modality, that is to say, in determining how well the student musician matches the reference musician, the time taken to evaluate each pair of samples, and subjective opinion of perceived utility of the feedback.

For articulation, differences in the mean scores assigned by the participants to the reference versus the student performance were not statistically significant for each modality. This suggests that while the visualization strategy did not offer any advantage over presentation of the samples by audio playback alone, visualization nevertheless provided sufficient information to make similar ratings. For dynamics, four of our six participants categorized the visualizations as helpful. The means of their ratings for the visualization-only and both-together conditions were not statistically different but were statistically different from the audio-only treatment, indicating a dominance of the visualizations when presented together with audio. Moreover, the ratings of dynamics under the visualization-only condition were significantly more consistent than the other conditions.

8294-08, Session 3

Exploring ensemble visualization

C. G. Healey, M. N. Phadke, L. Pinto, North Carolina State Univ. (United States); F. Alabi, J. M. Harter, R. M. Taylor II, The Univ. of North Carolina at Chapel Hill (United States); X. Wu, Renaissance Computing Institute (United States); H. Petersen, S. A. Bass, Duke Univ. (United States)

An ensemble is a collection of related datasets. Each dataset, or member, of an ensemble is normally large, multidimensional, and spatio-temporal. Ensembles are used extensively by scientists and mathematicians, for example, by executing a simulation repeatedly with slightly different input parameters and saving the results in an ensemble to see how parameter choices affect the simulation. To draw inferences from an ensemble, scientists need to compare data both within and between ensemble members. We propose two techniques to support ensemble exploration and comparison: a pairwise sequential animation method that visualizes locally neighboring members simultaneously, and a screen door tinting method that visualizes subsets of members using screen space subdivision. We demonstrate the capabilities of both techniques, first using synthetic data, then with simulation data of heavy ion collisions in high-energy physics. Results show that both techniques are capable of supporting meaningful comparisons of ensemble data.

8294-09, Session 4

Parallel large-data visualization with display walls

L. Scheidegger, Facebook Inc. (United States); H. Vo, Polytechnic Institute of New York Univ. (United States); J. Kruger, Univ. des Saarlandes (Germany); C. T. Silva, Polytechnic Institute of New York Univ. (United States); J. L. D. Comba, Univ. Federal do Rio Grande do Sul (Brazil)

While there exist popular software tools that leverage the power of arrays of tiled high resolution displays, they usually require either the use of a particular API or significant programming effort to be properly configured. We present PVW (Parallel Visualization using display Walls), a framework that uses display walls for scientific visualization, requiring minimum labor in setup, programming and configuration. PVW works as a plug-in to pipeline-based visualization software, and allows users to migrate existing visualizations designed for a single-workstation, single-display setup to a large tiled display running on a distributed machine. Our framework is also extensible, allowing different APIs and algorithms to be made display wall-aware with minimum effort.

8294-10, Session 4

Visual exploratory analysis of a large volume of SQL log data with the SDSS log viewer

J. Zhang, C. Chen, M. Vogeley, D. Pan, Drexel Univ. (United States); A. Thakar, J. Raddick, The Johns Hopkins Univ. (United States)

User-generated SQL (Structured Query Language) query logs are rich information sources for database analysts, information scientists, and the end users of databases. In this study a group of data scientists and information scientist work together to analyze a large volume SQL log data generated by users of the Sloan Digital Sky Survey (SDSS) data archive in order to better understand users' data seeking behaviors. While statistical analysis of these logs is useful at aggregated levels, efficiently exploring details and patterns of queries is often a challenging task due to the typically large data volume, multivariate features, and text content in SQL queries. To enable and facilitate effective and efficient exploration of the SDSS log data, we designed an interactive visualization tool, called the SDSS Log Viewer, which integrates time series visualization, text visualization, and dynamic query techniques. We describe two analysis scenarios of visual exploration of SDSS log data, including understanding unusually high daily query traffic and modeling the types of data seeking behaviors of massive query generators. The two scenarios demonstrate that the SDSS Log Viewer provides a novel and potentially valuable approach to support these targeted tasks.

Conference 8294

8294-11, Session 5

Comparison of open-source visual analytics toolkits

J. R. Harger, Sandia National Labs. (United States) and The Univ. of New Mexico (United States); P. J. Crossno, Sandia National Labs. (United States)

We present the results of the first stage of a two-stage evaluation of open source visual analytics packages. This stage is a broad feature comparison over a range of open source toolkits. Although we had originally intended to restrict ourselves to comparing visual analytics toolkits, we quickly found that very few were available. So, we expanded our study to include information visualization, graph analysis, and statistical packages. We examine three aspects of each toolkit: visualization functions, analysis capabilities, and development environments.

The first section describes the visualization functionality categories we examined, such as graph visualizations, tree visualizations, geo-spatial, general data plots and diagrams. We provide comparison tables of each category, specifying which algorithms a toolkit used when that information was available.

Next, we describe analysis capabilities, including graph analysis and statistical analysis techniques. These include finding shortest paths, minimum spanning trees, centrality measures, and clustering. We provide tables comparing available functionality for these categories as well.

Finally, we compare the aspects of the develop environments available for each toolkit. We give comparison tables showing programming language bindings, target platforms, GUI interfaces, database support as well as documentation available.

8294-12, Session 5

Evaluation of progressive treemaps to convey tree and node properties

R. Rosenbaum, B. Hamann, Univ. of California, Davis (United States)

In this paper we evaluate progressive treemaps. Progressive refinement has a long tradition in image communication, but is a relatively new approach for information presentation. Besides technical benefits it also promises to provide advantages important for the conveyance of data properties. In this first user study in this domain, we focus on the additional value of progressive refinement for traditional treemaps to convey the topology of a given hierarchical data set and properties of its nodes. To achieve this, we compare the results gained for common squarified treemap displays with and without progression for various related tasks and set-ups. The results we obtained indicate that progressive treemaps allow for a better conveyance of topological features and node properties in most set-ups. We also assessed the opinions of our study participants and found that progressive treemaps also lead to a better confidence about the given answers and provide more assistance and user friendliness.

8294-13, Session 5

Evaluation of multivariate visualizations: a case study of refinements and user experience

M. A. Livingston, J. Decker, U.S. Naval Research Lab. (United States)

One difficulty of multi-variate visualization (MVV) is that the number of perceptual channels may be exceeded. We embarked on a series of evaluations of MVVs in an effort to understand the limitations of attributes that are used in MVVs. In a follow-up study to previously published results, we attempted to use our past results to inform refinements to the design of the MVVs and the study itself. Some of these changes resulted in improved performance, whereas other degraded performance. We report results from the follow-up study and a comparison of data collected from subjects who participated in both studies. One gratifying result includes improved performance with a new MVV, Attribute Blocks, relative to Dimensional Stacking. On the other hand, our refinement to Data-driven Spots resulted in greater errors on the task. Users' previous exposure to the MVVs enabled them to complete the task significantly faster (but not more accurately). Previous exposure also yielded lower ratings of subjective workload. We discuss these intuitive and counter-intuitive results and the implications for MVV design.

8294-14, Session 6

Integrating sentiment analysis and term associations with geo-temporal visualizations on customer feedback streams

M. C. Hao, Hewlett-Packard Labs. (United States); C. Rohrdantz, H. Janetzko, D. A. Keim, Univ. Konstanz (Germany); U. Dayal, Hewlett-Packard Labs. (United States); L. Haug, Hewlett-Packard Co. (United States); M. Hsu, Hewlett-Packard Labs. (United States)

Twitter currently receives over 190 million tweets (small text-based Web posts) and manufacturing companies receive over 10 thousand web product surveys a day, in which people share their thoughts regarding a wide range of products and their features. A large number of tweets and customer surveys include opinions about products and services. However, with Twitter being a relatively new phenomenon, these tweets are underutilized as a source for determining customer sentiments. To explore high-volume customer feedback streams, we integrate three time series-based visual analysis techniques: (1) feature-based sentiment analysis that extracts, measures, and maps customer feedback; (2) a novel idea of term associations that identify attributes, verbs, and adjectives frequently occurring together; and (3) new pixel cell-based sentiment calendars, geo-temporal map visualizations and self-organizing maps to identify co-occurring and influential opinions. We have combined these techniques into a well-fitted solution for an effective analysis of large customer feedback streams such as for movie reviews (e.g., Kung-Fu Panda) or web surveys (buyers).

8294-15, Session 6

A self-adaptive technique for visualizing geospatial data in 3D with minimum occlusion

A. Chaudhuri, H. Shen, The Ohio State Univ. (United States)

Geospatial data are often visualized as 2D cartographic maps with interactive display of detail on-demand. Integration of the 2D map, which represents high level information, with the details pertaining to specific locations is a key design issue in geovisualization. Solutions include multiple linked displays around the map which can impose cognitive load on the user as the number of links goes up, and separate windowed displays on top of the map which causes occlusion of the map. In this paper, we present a self-adaptive technique which reveals hidden layers of information in single display and but minimizes occlusion of the 2D map. The proposed technique creates extra screen space by invoking controlled deformation of the 2D map. We extend our method to allow simultaneous display of multiple windows at different map locations. Since our technique is not dependent on the type of data to display, we expect it to be useful to both common users and the scientists. Case studies are provided in the paper to demonstrate the utility of the method in occlusion management and visual exploration.

8294-16, Session 7

Space/error tradeoffs for lossy wavelet reconstruction

J. Frain, R. D. Bergeron, The Univ. of New Hampshire (United States)

No abstract available

8294-17, Session 7

A configurable data prefetching scheme for interactive visualization of large-scale volume data

B. Jeong, Schlumberger (United States); P. Navratil, K. Gaither, G. Abram, G. P. Johnson, The Univ. of Texas at Austin (United States)

This paper presents a novel data prefetching and memory management scheme to support interactive visualization of large-scale volume datasets using GPU-based isosurface extraction. Our dynamic in-core approach uses a span-space lattice data structure to predict and prefetch the portions of a dataset that are required by isosurface queries, to manage an application-level volume data cache, and to ensure load-balancing for parallel execution. We also present a GPU memory management scheme that enhances isosurface extraction and rendering performance. With these techniques, we achieve rendering performance superior to other in-core algorithms while using dramatically fewer resources.

8294-18, Session 7

A general approach for similarity-based linear projections using a genetic algorithm

J. A. Mouradian, B. Hamann, R. Rosenbaum, Univ. of California, Davis (United States)

A widely applicable approach to visualizing properties of high-dimensional data is to view the data as a linear projection into two- or three-dimensional space. However, developing an appropriate linear projection is often difficult. Information can be lost during the projection process, and many linear projection methods only apply to a narrow range of qualities the data may exhibit. We propose a general-purpose genetic algorithm to develop linear projections of high-dimensional data sets which preserve a specified quality of the data set as much as possible. The obtained results show that the algorithm converges quickly and reliably for a variety of different data sets.

8294-19, Session 7

Image space adaptive volume rendering

A. Corcoran, J. Dingliana, Trinity College Dublin (Ireland)

We present a technique for interactive direct volume rendering which provides adaptive sampling at a reduced memory requirement compared to traditional methods. Our technique exploits frame to frame coherence to quickly generate a two-dimensional importance map of the volume which guides sampling rate optimisation and allows us to provide interactive frame rates for user navigation and transfer function changes. In addition our ray casting shader detects any inconsistencies in our two-dimensional map and corrects them on the fly to ensure correct classification of important areas of the volume.

8294-25, Poster Session

X3DBio1: a visual analysis tool for biomolecular structure exploration

H. Yi, Renaissance Computing Institute (United States); A. Singh, Y. G. Yingling, North Carolina State Univ. (United States)

Protein tertiary structure analysis provides valuable information on their biochemical functions. The structure-to-function relationship can be directly addressed through three dimensional (3D) biomolecular structure exploration and comparison. We present X3DBio1, a visual analysis tool for 3D biomolecular structure exploration, which allows for easy visual analysis of 2D intra-molecular contact map and 3D density exploration for protein, DNA, and RNA structures. A case study is also presented in this paper to illustrate the utility of the tool. X3DBio1 is open source and freely downloadable from http://sourceforge.net/

projects/x3dbio1 . We expect this tool can be applied to solve a variety of biological problems.

8294-26, Poster Session

Increasing the perceptual salience of relationships in parallel coordinate plots

J. M. Harter, X. Wu, The Univ. of North Carolina at Chapel Hill (United States); D. Dougherty, Michigan State Univ. (United States); H. Petersen, S. Bass, Duke Univ. (United States); R. M. Taylor II, The Univ. of North Carolina at Chapel Hill (United States); M. N. Phadke, L. Pinto, North Carolina State Univ. (United States); O. S. Alabi, The Univ. of North Carolina at Chapel Hill (United States)

We present three extensions to parallel coordinates that increase the perceptual salience of relationships between axes in multivariate data sets: (1) luminance modulation maintains the ability to preattentively detect patterns in the presence of overplotting, (2) adding a one-vs.-all variable display highlights relationships between one variable and all others, and (3) adding a scatter plot within the parallel-coordinates display preattentively highlights clusters and spatial layouts without strongly interfering with the parallel-coordinates display. These techniques can be combined with one another and with existing extensions to parallel coordinates, and two of them generalize beyond cases with known-important axes. We applied these techniques to two real-world data sets (relativistic heavy-ion collision hydrodynamics and weather observations with statistical principal component analysis) as well as the popular car data set. We present relationships discovered in the data sets using these methods.

8294-27, Poster Session

Comparative visualization of ensembles using ensemble surface slicing

O. S. Alabi, The Univ. of North Carolina at Chapel Hill (United States); X. Wu, Renaissance Computing Institute (United States); H. Petersen, S. Bass, Duke Univ. (United States); S. Zhong, Michigan State Univ. (United States); M. N. Phadke, L. Pinto, C. G. Healey, North Carolina State Univ. (United States); R. M. Taylor II, The Univ. of North Carolina at Chapel Hill (United States)

By definition, an ensemble is a set of surfaces or volumes derived from a series of simulations or experiments. Sometimes the series is run with different initial conditions for one parameter to determine parameter sensitivity. The understanding and identification of visual similarities and differences among the shapes of members of an ensemble is an acute and growing challenge for researchers across the physical sciences. More specifically, the task of gaining spatial understanding and identifying similarities and differences between multiple complex geometric data sets simultaneously has proved challenging. This paper proposes a comparison and visualization technique to support the visual study of parameter sensitivity. We present a novel single-image view and sampling technique which we call Ensemble Surface Slicing (ESS). ESS produces a single image that is useful for determining differences and similarities between surfaces simultaneously from several data sets. We demonstrate the usefulness of ESS on two real-world data sets from our collaborators.

8294-29, Poster Session

A performance assessment on the effectiveness of digital image registration methods

S. T. Kacenjar, Lockheed Martin Corp. (United States); B. Li, Lockheed Martin Systems Integration-Owego (United States); A. Ostrow, Lockheed Martin Maritime Systems & Sensors (United States)

Digital Image Correlation (DIC) of time-sequenced-imagery (TSI) has

become a very popular method in the study of material deformation, geological terrestrial movement, medicine, and environmental impact assessment such as the before and aftermath of a tsunami. By examining the before-and-after images of such problem domains it is possible to estimate critical information about the scene deformation and structural differences between the imagery.

Derivative methods have also been examined to improve such registration based upon image pre-filtering before DIC (PBDIC) and diffusion-based / order paired statistical (DBOPS) methods. Each of these methods offer distinct trades offs with respect to processing time and lock-in accuracy.

Regardless of the approach, the fidelity of such derived information is influenced by several factors that degrade its accuracy including; camera model uncertainty, surface deformation, sensor noise, illumination variation, and the spatial mobility of object classes within the imagery.

This paper assesses and compares the impact of three such degrading factors (camera model uncertainty, surface deformation, and sensor noise) on TSI image registration accuracy using the DIC, PBDIC, and DBOPS registration approaches. Results are presented in terms of registration accuracy, correlation strength, and computation time.

8294-30, Poster Session

An evaluation of rendering and interactive methods for volumetric data exploration in virtual reality environments

N. Wang, A. Paljic, P. Fuchs, Mines ParisTech (France)

In this paper we evaluate one interaction method and four display techniques for exploring volumetric datasets in virtual reality immersive environments. We propose an approach based on the display of a subset of the volumetric data, as isosurfaces, and an interactive manipulation of the isosurfaces to allow the user to look for local properties in the datasets. We also studied the influence of four different rendering techniques for isosurface rendering in a virtual reality system. The study is based on a search and point task in a 3D temperature field. User precision, task completion time and user movement were evaluated during the test. The study allowed to choose the most suitable rendering mode for isosurface representation, and provided guidelines for data exploration tasks in immersive environments.

8294-32, Poster Session

Efficient, dynamic data visualization with persistent data structures

J. A. Cottam, A. Lumsdaine, Indiana Univ. (United States)

Working with data that is changing while it is being worked on, so called "dynamic data", presents unique challenges to a visualization and analysis framework. In particular, making rendering and analysis mutually exclusive can quickly lead to either livelock in the analysis, unresponsive visuals or incorrect results. A framework's data store is a common point of contention that often drives the mutual exclusion. Providing safe, synchronous access to the data store eliminates the livelock scenarios and responsive visuals while maintaining result correctness. Persistent data structures are a technique for providing safe, synchronous access. They support safe, synchronous access by directly supporting multiple versions of the data structure with limited data duplication. With a persistent data structure, rendering acts on one version of the data structure while analysis updates another, effectively double-buffering the central data store. Pre-rendering work based on global state (such as scaling all values relative to the global maximum) is also efficiently treated if independently modified versions can be merged. The Stencil visualization system uses persistent data structures to achieve task-based parallelism between analysis, pre-rendering and rendering work with little synchronization overhead. With efficient persistent data structures, performance gains of several orders of magnitude are achieved. 8294-33, Poster Session Radial visualizations for comparative data analysis G. Draper, M. G. Styles,

Brigham Young Univ.-Hawaii (United States); R. F. Riesenfeld, Brigham Young Univ. (United States) SQiRL is a novel visualization system for querying and visualizing large multivariate data sets. Although initially designed for novice users, recent extensions to SQiRL facilitate more advanced analysis without sacrificing the simplicity that makes this visualization appealing to beginners. The default view provides a simple-to-learn interface for query evaluation. Intermediate users are provided a straightforward method for comparing the results of two queries. More advanced users can make use of a radial crosstab," a new interactive visualization technique that melds the expressive power of traditional crosstabulation with a drag-and-drop canvas. Through application to multiple data sets, we demonstrate the system's generality.

8294-36, Poster Session

Exploiting major trends in subject hierarchies for large-scale collection visualization

C. Julien, P. Tirilly, Univ. of Wisconsin-Milwaukee (United States); J. E. Leide, C. Guastavino, McGill Univ. (Canada)

Unlike unstructured Web collections, many large digital collections are currently organized by subject; however, these useful information organization structures are large and complex making them difficult to browse. Current online tools and visualization prototypes show small localized subsets and do not provide the ability to explore the predominant patterns of the overall subject structure. This research aims to address this issue by first simplifying the subject structure using two techniques based on the highly uneven distribution of real-world collections: level compression and child pruning. The approach is demonstrated using a sample of 130K records organized by the Library of Congress Subject Headings (LCSH). Promising results show we can reduce the subject hierarchy down to 35% of its initial size, while maintaining access to over 80% of major subject content. The visual impact is demonstrated using a traditional outline view allowing searchers to dynamically change the amount of complexity they feel is necessary for the tasks at hand.

8294-37, Poster Session

Visualization of multidimensional time

L. A. Tychonievich, Brigham Young Univ. (United States) and Univ. of Virginia (United States); R. P. Burton, Brigham Young Univ. (United States)

Time generally is assumed to be a scalar: it can be sorted, is unidirectional, and has only a single dimension. In this work we demonstrate that vector-valued multidimensional time can be defined meaningfully, simulated efficiently, and visualized in an interactive manner. We present two particular simulations, providing a first look at what hypertime may be "like" from both a physical and a navigational perspective. Although similar in many ways to our experience, mT phenomena also differ from 1T phenomena on a fundamental level. Our visualization framework motivates observations of some of these differences and helps us identify a variety of open tasks that will further our understanding of the characteristics of time, whatever its dimensionality. Together, these results form a basis from which arbitrary space-time dimensionalities can be understood.

8294-38, Poster Session

Degeneracy-aware interpolation of 3D diffusion tensor fields

C. Bi, S. Takahashi, The Univ. of Tokyo (Japan); I. Fujishiro, Keio Univ. (Japan)

Visual analysis of 3D diffusion tensor fields has become an important topic especially in medical imaging for understanding physical properties and microscopic structures of biological tissues. However,

it is still difficult to continuously track the underlying features from the discrete tensor samples, due to the absence of appropriate interpolation schemes in the sense that we are able to handle possible degeneracy while fully respecting the smooth transition of tensor anisotropic features. This paper presents such an approach to interpolating 3D diffusion tensor fields. The overall process consists of three primary steps. First, we systematically locate the possible rotational inconsistency of tensor anisotropy around degenerate points. This is accomplished by clustering discrete tensor samples with similar anisotropy and orientation using a minimum spanning tree strategy. Then, we rectify such rotational inconsistency by optimizing their rotational transformation through explicitly introducing the eigenstructure-based representation of tensor samples. Finally, the eigenvectors and eigenvalues of the tensor samples will be separately interpolated using the optimal correspondence between every pair of adjacent tensor samples in the second step. Comparisons with existing interpolation schemes will be provided to demonstrate the advantages of our scheme, together with several results of tracking white matter fiber bundles in human brain.

8294-39, Poster Session

Visualization and analysis of 3D gene expression patterns in zebrafish using web services

D. Potikanond, F. J. Verbeek, Leiden Univ. (Netherlands)

Gene expression patterns analysis with microarray provides quantitative information that shows how a gene is expressed in a particular condition. Whole mount in situ hybridization can be used to capture the spatio-temporal relation of the gene expression patterns. Therefore, visual integration of gene expression data from both techniques with digital atlas data of an organism can help identifying spatial, temporal and quantitative aspects of gene expression in different developmental stages. In this paper, we present an approach to provide an online visualization of gene expression data in zebrafish within 3D reconstruction model of zebrafish in different developmental stages. We developed SOAP web services that provide programmatic access to the 3D reconstruction data and spatial-temporal gene expression data in our previously developed information systems; the 3D digital atlas of zebrafish development and the Gene Expression Management System (GEMS). In this approach, we also created a web application that uses the newly developed web services to retrieve data from local repositories. Web application also uses the web services to retrieve relevant microarray analysis gene expression data from external community resource; ArrayExpress Atlas. All the gene expression patterns data and the 3D reconstruction data are subsequently integrated using ontology based mapping. To delivery the visualization to end users, we developed a Java based 3D viewer that can be integrated with web interface allowing users to visualize the integrated information over the Internet. The viewer supports arbitrary re-sectioning of the volume data and provides visualization for related gene expression information in 3D graphical model and 3D annotations.

8294-40, Poster Session

Vortex core detection: back to basics

A. Van Gelder, Univ. of California, Santa Cruz (United States)

A robust definition of vortex-core axis has eluded researchers for a decade. This paper reviews the criteria described in some early papers, as well as recent papers that concentrate on issues of unsteady flows, and attempts to build on their ideas. In particular, researchers have proposed criteria that are desirable for a vortex-core axis that correspond to nonlocal properties, yet current extraction methods are all based on local properties. Analysis is presented to support the thesis that inaccuracies observed in some popular early methods are due to a mixture of frequencies in the flow field in vortical regions. Such mixtures occur in steady flows, as well as unsteady (time-varying) flows. Thus, the fact that the flows are unsteady is not necessarily the primary reason for inaccuracies recently observed in vortex analysis of such flows. It is hypothesized that time-varying (unsteady) flows tend

to be more complex, hence tend to have mixed frequencies more often than steady flows. We further conjecture that an "effective" lack of Galilean invariance may occur in steady or unsteady flows, due to the interaction of low frequencies with high frequencies.

8294-20, Session 10

Visualization of mappings between the gene ontology and cluster trees

I. Jusufi, A. Kerren, V. Aleksakhin, Linnaeus Univ. (Sweden); F. Schreiber, Leibniz Institute of Plant Genetics and Crop Plant Research (Germany) and Martin-Luther Univ. Halle-Wittenberg (Germany)

Ontologies and hierarchical clustering are both important tools in biology and medicine to study high-throughput data such as transcriptomics and metabolomics data. Enrichment of ontology terms in the data is used to identify statistically overrepresented ontology terms, giving insight into relevant biological processes or functional modules. Hierarchical clustering is a standard method to analyze and visualize data to find relatively homogeneous clusters of experimental data points. Both methods support the analysis of the same data set, but are usually considered independently. However, often a combined view is desired: visualizing a large data set in the context of an ontology under consideration of a clustering of the data. This paper proposes a new visualization method for this task.

8294-21, Session 10

Visualizing uncertainty in biological expression data

C. Holzhüter, Univ. Rostock (Germany); A. Lex, D. Schmalstieg, H. Schulz, Technische Univ. Graz (Austria); H. Schumann, Univ. Rostock (Germany); M. Streit, Technische Univ. Graz (Austria)

Expression analysis of ~omics data using microarrays has become a standard procedure in the life sciences. However, microarrays are subject to technical limitations and errors, which renders the data gathered likely to be uncertain.

While a number of approaches exist to target this uncertainty statistically, it is hardly ever even shown when the data is visualized using for example clustered heatmaps. Yet, this is highly useful when trying not to omit data that is 'good enough' for an analysis, which otherwise would be discarded as too unreliable by established conservative thresholds. Our approach addresses this shortcoming by first identifying the margin above the error threshold of uncertain, yet possibly still useful data. It then displays this uncertain data in the context of the valid data by enhancing a clustered heatmap. We employ different visual representations for the different kinds of uncertainty involved. Finally, it lets the user interactively adjust the thresholds, giving visual feedback in the heatmap representation, so that an informed choice on which thresholds to use can be made instead of applying the usual rule-of-thumb cut-offs. We exemplify the usefulness of our concept by giving details for a concrete use case from our partners at the Medical University of Graz, thereby demonstrating our implementation of the general approach.

8294-22, Session 11

Instant visitation maps for interactive visualization of uncertain particle trajectories

K. Bürger, R. Fraedrich, Technische Univ. München (Germany); D. Merhof, Univ. Konstanz (Germany); R. Westermann, Technische Univ. München (Germany)

Visitation maps are an effective means to analyze the frequency of similar occurrences in large sets of uncertain particle trajectories. A visitation map counts for every cell the number of trajectories passing

Conference 8294

through this cell, and it can then be used to visualize pathways of a certain visitation percentage. In this paper, we introduce an interactive method for the construction and visualization of high-resolution 3D visitation maps for large numbers of trajectories. To achieve this we employ functionality on recent GPUs to efficiently voxelize particle trajectories into a 3D texture map. In this map we visualize envelopes enclosing particle pathways that are followed by a certain percentage of particles using direct volume rendering techniques. By combining visitation map construction with GPU-based Monte-Carlo particle tracing we can even demonstrate the instant construction of a visitation map from a given vector field. To facilitate the visualization of safety regions around possible trajectories, we further generate Euclidean distance transform volumes to these trajectories on the fly. We demonstrate the application of our approach for visualizing the variation of stream lines in 3D flows due to different numerical integration schemes or errors introduced through data transformation operations, as well as for visualizing envelopes of probabilistic fiber bundles in DTI tractography.

8294-23, Session 11

Motion visualization in large particle simulations

R. Fraedrich, R. Westermann, Technische Univ. München (Germany)

Interactive visualization of large particle sets is required to analyze the complicated structures and formation processes in astrophysical particle simulations. While some research has been done on the development of visualization techniques for steady particle fields, only very few approaches have been proposed to interactively visualize large time-varying fields and their dynamics. Particle trajectories are known to visualize dynamic processes over time, but due to occlusion and visual cluttering such techniques have only been reported for very small particle sets so far. In this paper we present a novel technique to solve these problems, and we demonstrate its potential for the visual exploration of large astrophysical particle sequences. We present a new hierarchical space-time data structure for particle sets which allows for a scale-space analysis of trajectories in the simulated fields. In combination with visualization techniques that adapt to the respective scales, clusters of particles with homogeneous motion as well as separation and merging regions can be identified effectively. The additional use of dynamic transfer functions to modulate the color and size of trajectories allows emphasizing various particle properties like direction, speed or mass. Furthermore, tracking of interactively selected particle subsets permits the user to focus on structures of interest.

8294-24, Session 11

Animating streamlines with repeated asymmetric patterns for steady flow visualization

C. Yeh, National Cheng Kung Univ. (Taiwan); Z. Liu, Univ. of Pennsylvania (United States); T. Lee, National Cheng Kung Univ. (Taiwan)

Animation provides intuitive cueing for revealing essential spatial-temporal features of data in scientific visualization. This paper explores the design of Repeated Asymmetric Patterns (RAPs) in animating evenly-spaced color-mapped streamlines for dense accurate visualization of complex steady flows. We present a smooth cyclic variable-speed RAP animation model that performs velocity (magnitude) integral luminance transition on streamlines. This model is extended with inter-streamline synchronization in luminance varying along the tangential direction to emulate orthogonal advancing waves from a geometry-based flow representation, and then with evenly-spaced hue differing in the orthogonal direction to construct tangential flow streaks. To weave these two mutually dual sets of patterns, we propose an energy-decreasing strategy that adopts an

iterative yet efficient procedure for determining the luminance phase and hue of each streamline in HSL color space. We also employ adaptive luminance interleaving in the direction perpendicular to the flow to increase the contrast between streamlines.

Conference 8295A: Image Processing: Algorithms and Systems X

Monday-Wednesday 23-25 January 2012

Part of Proceedings of SPIE Vol. 8295A Image Processing: Algorithms and Systems X

8295A-01, Session 1

Analysis of different image-based biofeedback models for improving cycling performances

D. Bibbo, S. Conforto, I. Bernabucci, M. Carli, M. Schmid, T. D'Alessio, Univ. degli Studi di Roma Tre (Italy)

Sport practice can take great advantage by the quantitative assessment of task execution that is strictly connected to the implementation of optimized training procedures. To this aim biofeedback techniques are interesting to explore even if such a training approach implies a complete chain for information extraction containing instrumented devices, processing algorithms and graphical interfaces to extract valuable information (i.e. kinematics, dynamics, and electrophysiology) to be presented in real-time to the athlete. In cycling , performance indexes displayed in a simple and perceptible way can help the cyclist to optimize the pedaling. To this purpose, in this work a system based on an instrumented pedal for bikes is used to measure the force components exerted during pedaling. A mechanical Index of Efficiency (IE) is then real-time calculated and displayed to the user. Different graphical interfaces have been designed and used in this study in order to understand if and how a graphical IE biofeedback can influence the cycling performance. A group of subjects underwent the experimental protocol and pedaled with (i.e. the interfaces have been used in a randomized order) and without graphical biofeedback. Preliminary results show how the effective perception of the biofeedback influences the motor performance.

8295A-02, Session 1

Textured areas detection and segmentation in circular harmonic functions domain

L. Costantini, L. Capodiferro, Fondazione Ugo Bordoni (Italy); M. Carli, A. Neri, Univ. degli Studi di Roma Tre (Italy)

In this paper a novel technique for detecting and segmenting textured areas in natural images is presented. Although it is very difficult to define what is a texture, it is easily recognizable by users. Textured areas are an important feature of natural images, as well as edges. Nowadays, the image enhancement algorithms automatically improve the quality of images, by removing noise, by sharpening the edges, and by smoothing the flat areas. These algorithms handle the textured area as flat areas or as edges, thus resulting in information and quality loss. Therefore an automatic system for detecting textured areas is the preliminary fundamental task for any enhancement technique. Furthermore, the detection of the textured areas could be exploited in many other image processing application, such as image retrieval or image quality assessment. Content-based image retrieval (CBIR) systems make use of the low level features to represent the content of images. The most common low level features used are color, texture and edges. In this context an algorithm that detects the textured area could improve the performances of the both texture characterization methods and the overall CBIR systems.

8295A-03, Session 1

Searching for streamer trajectories on synoptic maps of the sun corona

A. Llebaria, Observatoire Astronomique de Marseille-Provence (France)

Synoptic maps of the solar corona are useful tools in order to study the evolution of coronal activity for long periods of time (years). They show

the coronal activity as function of time. This activity is measured as brightness in a circular profile around the solar disk as function of the latitude. The time series of these profiles form the synoptic maps. There are three conspicuous types of events on solar corona and therefore on synoptic maps: 1) the coronal mass ejections (CMEs), 2) the streamers 3) the coronal holes. Because all these structures are optically thin there add their brightness in the synoptic map. Streamers and coronal holes define the medium term appearance of the solar corona. CMEs are frequent and very strong transients (in general). In this paper we show how to disentangle these additive components of very diverse appearance: CME appear as temporal transients, streamers appear as wandering traces and coronal holes appear as low level smooth regions. Median filtering is efficient for transients, thresholding and morphologic analysis is good for coronal holes but the existence of multiple streamers need a combined approach to disentangle these components using Hough transform techniques and multirésolution analysis.

8295A-04, Session 1

Performance evaluation for 2D and 3D filtering methods of noise removal in color images

V. V. Lukin, A. A. Zelensky, N. N. Ponomarenko, National Aerospace Univ. (Ukraine); K. O. Egiazarian, J. T. Astola, Tampere Univ. of Technology (Finland)

Color images formed by modern digital cameras are often noisy and it is makes desirable to remove this noise by image pre-filtering. Filters to be recommended for the considered application have to effectively suppress noise with introducing negligible distortions into processed images. Moreover, such filters have to be fast enough and tested for a wide variety of natural images and noise environments. Visual quality of processed images has to be paid specific attention. For testing, the recently designed database TID2008 of distorted images provides good opportunities. Our study is performed in terms of the standard PSNR and some visual quality metrics, in particular, PSNR-HVS-M. It is shown that 3D (vector) filtering of color images is able to provide additional improvement compared to the 2D filtering especially for textural images. On the average, the 3D filter produces about 2 dB increase of PSNR-HVS-M compared to the 2D DCT-based denoising. The results obtained for spatially correlated noise case are similar, i.e., the 3D filter performance is better than the component-wise filter performance in terms of PSNR and PSNR-HVS-M by about 2 dB. The reasons are high degree of signal component correlation in color images and the independence of noise in color components.

8295A-05, Session 2

Integrated text detection and recognition in natural images

N. S. Roubtsova, Technische Univ. Eindhoven (Netherlands) and ViNotion B.V. (Netherlands); R. Wijnhoven, ViNotion B.V. (Netherlands) and Technische Univ. Eindhoven (Netherlands); P. H. N. de With, Technische Univ. Eindhoven (Netherlands)

Text detection and recognition in natural images have conventionally been seen in the prior art as autonomous tasks executed in a strictly sequential processing chain with limited information sharing between sub-systems. This approach is flawed because it introduces (1) redundancy in extracting the same text properties multiple times and (2) error by prohibiting verification of hard (often binarized) detection results at later stages. We explore the possibilities for integration of detection and recognition modules by a feedforward multidimensional

information stream. Integration involves suitable characterization of the text string at detection and application of this knowledge to ease recognition by a given OCR system. The choice of characterization properties generally depends on the OCR system, although some of them have proven universally applicable.

We show that the proposed integration measures enable more robust recognition of text in complex, unconstrained natural environments. Specifically, integration by the proposed measures (1) eliminates textual input irregularities that recognition engines cannot handle and (2) adaptively tunes the recognition model for each input image. The former function boosts correct detections, while the latter mainly reduces the number of false positives.

8295A-06, Session 2

Ear recognition based on edge potential function

F. Battisti, M. Carli, Univ. degli Studi di Roma Tre (Italy); F. G. B. De Natale, Univ. degli Studi di Trento (Italy); A. Neri, Univ. degli Studi di Roma Tre (Italy)

In this paper a biometric system for human identification based on ear recognition is presented. The ear is modeled as set of contours extracted from the ear image with an edge potential function. The proposed matching algorithm has been tested in presence of several image modifications. Two human ear databases have been used for the tests and the achieved experimental results show the effectiveness of the proposed scheme.

8295A-07, Session 2

Feature extraction from ladar data using modified GPCA

P. F. Stiller, Texas A&M Univ. (United States)

In this paper we present a method for extracting feature information from ladar data in the form of a point cloud. The method exploits a modified version of Generalized Principal Component Analysis (GPCA) to extract planar or other non-linear surface elements from this sort of data. The essential difficulty is that, depending on the aspect of the object, certain surfaces will be minimally exposed. As a result we cannot say in advance how many surfaces we are looking for and we cannot reliably detect surfaces that are hit by only a few of the points in the cloud. An additional difficulty occurs when recontructing the surface normal at points where two surfaces join. The algorithm handles both issues and captures enough essential surface features to allow accurate alignment to say a CAD model for detailed recognition.

8295A-08, Session 2

Recognition of rotated images using the multi-valued neuron and rotation-invariant 2D Fourier descriptors

E. Aizenberg, I. J. Bigio, E. Rodriguez-Diaz, Boston Univ. (United States)

In this paper, we extend the well-established Fourier descriptors paradigm to images, in order to obtain features which are invariant to image rotation. By transferring an image from Cartesian to polar coordinates, we reduce rotation to a circular shift along a single dimension. While computing the Discrete Fourier Transform (DFT) of the polar image, a phase normalization algorithm is applied along the dimension of circular shift, cancelling its effect on the image spectrum. This algorithm retains the uniqueness of the signal's phase, and therefore vital information is not lost.

The phase of the derived normalized spectrum was used to train a single Multi-Valued Neuron (MVN) to recognize satellite (LIMA) and human face (PICS) images rotated by an arbitrary angle. 100% and 96.43% classification rates were achieved for each image set,

respectively. Additional testing sets were formed by rotating original samples that went through MATLAB's JPEG routine, and addition of Gaussian noise. JPEG compression yielded 96.57% - 100% classification rate on the LIMA set, and 89.76% - 97.14% classification rate on the PICS set. Addition of Gaussian noise (0.1 of signal's variance) resulted in 98.00% and 75.24% classification rates for the LIMA and PICS sets, respectively. Preliminary results show that the suggested scheme for obtaining rotation-invariant DFT features combined with the MVN provides a promising approach for efficient image recognition.

8295A-10, Session 3

Amoeba-based superpixel partitioning of multispectral images into elementary, uniform, connected units

J. Grazzini, L. Prasad, Los Alamos National Lab. (United States)

Segmentation is a difficult task because of the high complexity of images, where complexity refers to the large variety of pictorial representations of objects with the same semantic meaning and also to the extensive amount of available details. It seems therefore natural, and presumably more efficient, to work with perceptually meaningful entities obtained from low-level grouping processes instead of the pixel representation. In that context, superpixels obtained from conservative over-segmentation are a common pre-processing step for recovering image features. In this paper, we introduce a new algorithm that works essentially like a k-means based local clustering of pixels, but enforces connectivity, so that it can efficiently generate compact, connected, and nearly uniform superpixels. Our approach is based on the estimation of amoeba-like neighborhoods around selected cluster centers that exploit the connections between successive image pixels along geodesic paths in the image. The resulting superpixels capture the spatial/spectral redundancy in images and greatly reduce the complexity of subsequent image processing tasks. They provide convenient primitives from which to compute local image features when objects present in the scene have diverse scales or when they are not known in advance. One possible application is provided by a scale/shape adaptive technique for change detection in satellite images.

8295A-11, Session 3

Smooth partition of unity with Hermite interpolation: applications to image processing

L. T. Dechevsky, P. Zanaty, A. Lakså, B. Bang, Narvik Univ. College (Norway)

We shall consider a general approach for interpolation, representation and edition of 2D (planar) images and 3D (spatial) shapes. This approach uses a new mathematical apparatus (see [1]) for smooth convex resolution of unity on a very general class of domain partitions together with Hermite-type interpolation on a scattered-point set, the distribution of which is consistent with the domain partition in consideration.

We shall be using the tensor-product version of the construction in [1] as developed in [2], for 3D geometric representation of images sampled on possibly non-uniformly distributed (scattered) data point sets. In particular, we shall discuss how to handle the conversion from Taylor-Hermite interpolatory form to Bernstein-Bezier control-polygon form. We will further demonstrate how this can be used to edit the interpolated image interactively using standard technique for geometric modeling typical for Computer-Aided Geometric Design (CAGD).

To establish the needed isomorphism between the color values and 3D point coordinates we are using the classical RGB coding [3], as well as some other methods for color coding based on recent developments of ideas from the early work [4].

References:

[1] L.T. Dechevsky, "Smooth Convex Resolution of Unity and/or

Scattered-Point Hermite-Interpolation by Generalized Expo-Rational B-Splines, I: General simply connected covers and partitions", Int. J. Pure Appl. Math. 65(4), 413-460 (2010).

[2] L.T. Dechevsky, "Multivariate Hermite interpolation on scattered-point sets using tensor-product expo-rational B-splines", Communication at the 37th Int. Conf. on Applications on Mathematics in Engineering and Economics, AMEE, Sozopol 2011

[3] J. Gundersen, L.T. Dechevsky, "On the scientific visualization of complex-valued functions of one complex variable", In: Proceedings of the 35th Int. Conf. on Applications on Mathematics and Engineering and Economics, Sozopol 2009, AIP Conf. Proc. 1184, 324-334. Amer. Inst. Of Phys., Melville, NY, 2009

[4] R.S. Hunter, "Photoelectric Color-Difference Meter", Proceedings of the Winter Meeting of the Optical Society of America. JOSA 38(7), 661 (1948).

8295A-12, Session 3

An algorithm for GPGPU-computing of multidimensional DWTs based on bijective mapping of tensor-product wavelet bases of different number of variables

L. T. Dechevsky, J. Bratlie, B. Bang, A. Lakså, Narvik Univ. College (Norway)

An algorithm for computation of multivariate wavelet transforms (DWTs) on graphics processing units (GPUs) was proposed in [1]. This algorithm was based on mapping the indices of orthonormal tensor-product wavelet bases of different number of variables and a tradeoff between the number of variables versus the resolution level, so that the resulting wavelet bases of different number of variables are with different resolution, but the overall dimension of the bases is the same. In the above-said paper, the algorithm was developed only up to mapping of the indices of {em blocks} of wavelet basis functions. This was sufficient to prove the consistency of the algorithm, but not enough for the {em mapping of the individual basis functions} in the bases needed for a programming implementation of the algorithm. In [2] we began the study of an upgrade of this construction by passing from block-matrix index mapping on to the detailed index mapping of the individual basis functions for the case of an n-variate tensor-product wavelet basis being bijectively mapped onto a univariate wavelet basis. Here we continue this study by developing a further upgrade of the basis-matching algorithm which works in the most general case of bijective mapping between n_1-variate and n_2-variate tensor-product wavelet bases. Potential applications of the results obtained include: begin{itemize} item{ In computational imaging: for computing multivariate DWT via GPGPU-programming. } item{ In telecomunications: for transforming large volumes of multidimensional geometric information as univariate signals. } item{ In data encryption: if multiwavelet tensor-product bases are considered possible, the encryption keys may include the number of scaling functions and the values of the filter coefficients generating them. } end{itemize}

[1]L.T. Dechevsky, J. Gundersen, B. Bang,

Computing n-variate orthogonal discrete wavelet transforms on graphics processing units.

In: I. Lirkov, S. Margenov, J. Wasniewski (Eds.)

LSSC'2009, LNCS 5910, Springer-Verlag, Berlin-Heidelberg, 2010,

730--737

[2]L.T. Dechevsky, J. Bratlie, J. Gundersen,

Index mapping between tensor-product wavelet bases of different number of variables, and computing multivariate orthogonal discrete wavelet transforms on graphics processing units,

To appear in:LSSC'2011, LNCS

8295A-13, Session 4

Quantitative evaluation of image mosaicing in multiple scene categories

D. Ghosh, S. Park, N. Kaabouch, W. Semke, R. Fevig, The Univ. of North Dakota (United States)

In this paper, we propose an evaluation methodology for image mosaicing algorithms. Majority of previous evaluation methodologies lack sufficient number of performance metrics; while others suffer from computational complication. The proposed evaluation method involves four metrics, percentage of mismatches, difference of pixel intensities, peak signal-to-noise ratio (PSNR), and mutual information to measure the quality of the mosaicing outputs. These outputs are obtained using a mosaicing algorithm based on Scale Invariant Feature Transform (SIFT), Best Bins First (BBF), and Random Sample Consensus (RANSAC) algorithms. In order to evaluate mosaicing performance objectively, the proposed method compares mosaic image and the ground-truth image that depict the same scene view.The evaluation has been performed using 36 test sequences from 3 different categories: images of 2D surface, images of outdoor 3D scenes, and airborne images from an Unmanned Aerial Vehicle (UAV) during the University of North Dakota flight test. Exhaustive testing shows that the proposed metrics are efficient in assessing the quality of mosaicing outputs.

8295A-14, Session 4

Curvelet transform with adaptive tiling

H. Al-Marzouqi, G. Al-Regib, Georgia Institute of Technology (United States)

The curvelet transform is a recently introduced non-adaptive multi-scale transform that have gained popularity in the image processing field. In this paper, we study the effect of customized tiling of frequency content in the curvelet transform. Specifically, we investigate the effect of the size of the coarsest level and its relationship to denoising performance. Based on the observed behavior, we introduce an algorithm to automatically choose the optimal number of decompositions. Its performance shows a clear advantage, in denoising applications, when compared to default curvelet decomposition. We also examine how denoising is affected by varying the number of divisions per scale.

8295A-15, Session 4

Tetrachromatic colour space

A. Restrepo, Univ. de Los Andes (Colombia)

Based on the well known RGB colour space and on the recently introduced Runge colour space, we present the analogous colour spaces called the "WXYZ hypercube" (that corresponds to the Cartesian product of the allowed normalized output [0, 1] of the 4 sensors of the corresponding vision system) and a space of the type hue-colourfulness-luminance, called "4-Runge" colour space that is a four-dimensional solid ball (that corresponds to a "rounding" of sorts of the hypercube). The spaces find applications for example in the modeling of the vision systems of tetrachromatic animals and in the processing of 4-spectral images. The hue variable in these 4-spaces is bidimensional; in particular, it lives on a topological 2-sphere. The hue space is the (nonconvex) dodecahedron (called the "chromatic dodecahedron") that results from considering the faces in the boundary of the hypercube that do not have the points [0 0 0 0] ("black") nor [1 1 1 1] ("white") as a vertex; the boundary of the hypercube is a polytope of the type {4 3 3}, in Scläfli notation; the chromatic dodecahedron is thus a surface subset of the topological 3-sphere that is the boundary of the hypercube WXYZ. Depending on the ordering of the tetrad WXYZ, 24 basic types of hue result. In 4-Runge space, the hue is given by two angles (as in spherical coordinates) that determine a point in the equatorial 2-sphere of the boundary 3-sphere of the Runge 4-ball the point in the boundary is the intersection of the ray from the center of the ball to the colour point and the 3-sphere; the colourfulness is given by the distance from the colour point to the center of the ball and the

lightness is given as an angular measurement of the ray from the center ("intermediate gray") of the 4-ball to the colour point in the hyper-ball, and the ray that goes from the center to the point "white".

8295A-16, Session 5

Smooth image inpainting by least square oriented edge prediction

E. Pallotti, L. Capodiferro, F. Mangiatordi, Fondazione Ugo Bordoni (Italy); P. Sità, Univ. degli Studi di Roma Tre (Italy)

This paper introduces an new digital edge oriented algorithm for automatic digital inpainting. The basic idea is the Laguerre Gauss analysis of the structure information of the regions around the damaged portions of the image to extrapolate the gradient of the luminance in missing area by a least square prediction from simplified edge map. Completion of unknown parts is automatically obtained by a variational method using the predicted gradient information with smoothing constraints on luminance and color level. Experiments on a number of images show the effectiveness of the proposed algorithm.

8295A-17, Session 5

Image inpainting using cubic spline-based edge reconstruction

V. Voronin, V. I. Marchuk, South-Russian State Univ. of Economics and Service (Russian Federation); K. O. Egiazarian, Tampere Univ. of Technology (Finland); A. Sherstobitov, South-Russian State Univ. of Economics and Service (Russian Federation)

We have presented in this paper an image inpainting method which combines restoration of the edges, non texture regions and texture. First, we are using LPA-ICI method to realize segmentation in damaged image. Second, cubic spline curve is applied to reconstruct structure and connect edge and contour in damaged area. In next step we use new inpainting procedure to reconstruct texture and non texture areas independently. For texture we use exemplar based methods which find a best matching patch from other source region and copy it in damaged image. For non texture we use fast marching method. Proposed technique allows to obtain a smaller reconstruction error than that of the state-of-the-art image inpainting methods. We demonstrate the performance of a new approach via several examples, showing the effectiveness of our algorithm in reconstruction of missing small and large objects on the test images.

8295A-18, Session 5

Global registration and stabilization of jittered and noisy airborne image sequences

N. M. Namazi, The Catholic Univ. of America (United States); W. Scharpf, U.S. Naval Research Lab. (United States); J. Obermark, DCS Corp. (United States); J. Caron, Research Support Instruments, Inc. (United States)

This paper is concerned with the development and implementation of registration and stabilization method in conjunction with airborne imaging applications. We consider the situations for which the camera motion and vibration collectively affect the noisy image sequence. The general routine presented in this work is a combination of two algorithms for global image registration and image stabilization. We use and present experiments with real image sequences to track a moving object in the direction of its motion trajectory.

8295A-19, Session 6

Image and video restoration via Ising-like models

O. Hadar, Ben-Gurion Univ. of the Negev (Israel); E. Cohen, Tel Aviv Univ. (Israel)

During the last decades, statistical models, such as the Ising model, have become very useful in describing solid state systems. These models excel in their simplicity and intuitiveness. Furthermore, their results get quite often accurate experimental proofs. Leading researchers, such as Tanaka, have used them successfully during the last years to restore images [1]. Kandes has used recently a simple model to restore B/W and grayscale images and achieved fair results [2]. In this paper we will present the results we have achieved using a similar, though more complex model in order to get better results. Moreover, we shall describe the XY model which enables us to restore colored images. Additionally, we will describe the general method in which similar algorithms enable us to restore degraded video signals. Finally, we will present some other physical models which can be helpful in general to process images and confront their results with results which are well known from previous works [1,2,3,4].

8295A-20, Session 6

Region adaptive correction method for radial distortion of fish-eye image

K. S. Song, Y. S. Han, M. G. Kang, Yonsei Univ. (Korea, Republic of)

Most of cameras follow pinhole camera model. However, result of this model makes some undesirable effects in wide angle lens. The most serious problem among these effects is radial distortion which appears heavily in fish-eye images. Several geometric models for correcting radial distortion of fish-eye lens are developed. Most of these models require only one parameter. However, correcting with one parameter is limited to correct both central and outer part simultaneously. Aim of this paper is to solve this problem. The proposed method is able to correct radial distortion of both areas using region adaptive distortion parameter. Each parameter is determined by considering amount of distortion in each region respectively. Also, the proposed method modifies the existing division model to correct radial distortion of both regions. Experimental results show that radial distortions in both areas are corrected.

8295A-21, Session 6

Super-resolution image reconstruction with edge adaptive weight in video sequence

J. Y. Kwon, D. S. Yoo, M. G. Kang, Yonsei Univ. (Korea, Republic of)

Digital images and videos are used in many digital devices recently. Also, the resolution of display became larger than that of previous years. Image up-scaling algorithm is important issue since original input source is limited in transferring within data bandwidth. Among various up-scaling algorithms, Super-Resolution (SR) image reconstruction method is able to estimate high-resolution (HR) image using multiple low-resolution (LR) images. Conventional approaches to estimate HR image with Lease Square (LS) method and Weighted Least Square (WLS) method are not able to reconstruct high-frequency region effectively in case its blur kernel is assumed Gaussian kernel in unknown system. Also, these methods produce jagging artifacts from deficiency of LR frames. The proposed SR algorithm uses edge adaptive WLS to reconstruct high-frequency region considering local properties and is applied to video sequence with block process to cope with local motions. Moreover, to apply video sequence with complex motions, we use selectively correct information of reference frame to avoid errors from incorrect information. For accurate additional information from reference frames, the proposed algorithm determines additional information in reference frame by comparing with current frame and reference frame. The experiments demonstrate superior performance of the proposed algorithm.

Conference 8295A

8295A-22, Session 7

Color image enhancement based on genetic algorithm and ensemble empirical mode decomposition

S. Bakhtiari, S. S. Agaian, M. Jamshidi, The Univ. of Texas at San Antonio (United States)

In this paper, we introduce a new method for enhancement of the color images. The proposed approach utilizes Ensemble Empirical Mode Decomposition (EEMD), an extension of EMD method, and Genetic Algorithm (GA). The HSV color space is selected for brightness modification purpose since it has shown a higher effectiveness in the applications under study. The algorithm is tested on the underwater images and the underexposed scenes. For the dark images, a nonlinear transform is first performed on the luminance channel of the original image for a preliminary illumination correction. For underwater images, the mean of each RGB channel is initially corrected based on a GA based technique. In fact, GA is deployed for optimizing the offsets each channel should have with respect to the near optimal values. The enhancement measure, EME, is applied as the fitness function of the GA. After the preprocessing steps, an EEMD based method is performed on the luminance channel of the resulting image to further correct the brightness of the image. The Genetic Algorithm enables the presented method to set the required parameters automatically so that the applications can be feasible in autonomous missions. The experimental results demonstrate the effectiveness of the algorithm in visualizing the details which were unperceivable before enhancement.

8295A-23, Session 7

Image classification and interpolation

A. Khemka, KLA-Tencor Corp. (United States); C. A. Bouman, Purdue Univ. (United States)

We have developed a novel interpolation method for images containing text, graphics and natural scenes. The method allows us to select the best interpolation algorithm for different regions of an image. In particular, we segment the image into graphical and natural regions and use the appropriate algorithm for each region. The natural regions are interpolated using a current state-of-the-art algorithm. However, when applied to graphical images, the current state-of-the-art interpolators tend to produce artifacts at edge discontinuities. Thus, we developed a novel approach which we call Low Entropy Interpolation (LEI) algorithm for the graphical images. The LEI algorithm is highly non-linear and produces very sharp edges with very few defects necessary for good quality interpolation of graphical images.

8295A-24, Session 7

Optimal fractional filter for image segmentation

A. Nakib, Y. Schulze, E. Petit, Univ. Paris 12 - Val de Marne (France)

In this paper, we present a new image thresholding algorithm based on fractional filter (FF). Our experiments showed that a good segmentation result corresponds to an optimal order of the filter. Then, we propose new alternative based on Legendre moments' to find the optimal order. The whole algorithm is called FLM. This algorithm allows to include contextual information such as the global object shape and exploits the properties of the two-dimensional fractional integration. The efficiency of FLM was illustrated by the comparison to other six competing methods recently published and it was tested on real-world problem.

8295A-25, Session 7

Multi-scale image enhancement using a second derivative-like measure of contrast

S. C. Nercessian, K. Panetta, Tufts Univ. (United States); S. S.

Agaian, The Univ. of Texas at San Antonio (United States)

Image enhancement algorithms aim to improve the visual quality of images for human or machine perception. Most direct multi-scale image enhancement methods are based on enhancing either absolute intensity changes or the Weber contrast at each scale, and have the advantage that the visual contrast is enhanced in a controlled manner. However, the human visual system is not adapted to absolute intensity changes, while the Weber contrast is unstable for small values of background luminance and potentially unsuitable for complex image patterns. The Michelson contrast measure is a bounded measure of contrast, but its expression does not allow a straightforward direct image enhancement formulation. Recently, a second derivative-like measure (SDME) of contrast has been used to assess the performance of image enhancement algorithms. The SDME is a Michelson-like contrast measure for which a direct image enhancement algorithm can be formulated. Accordingly, we propose a new direct multi-scale image enhancement algorithm based on the SDME in this paper. Experimental results illustrate the potential benefits of the proposed algorithm.

8295A-26, Session 8

A new denoising method in high-dimensional PCA space

Q. B. Do, A. Beghdadi, M. Luong, Univ. Paris 13 (France)

Kernel-design based method such as Bilateral filter (BIL), non-local means (NLM) filter is known as one of the most attractive approaches for denoising. We propose in this paper a new noise filtering method inspired by BIL, NLM filters and principal component analysis (PCA). The main idea here is to perform the BIL in a multidimensional PCA-space using an anisotropic kernel. The filtered multidimensional signal is then transformed back onto the image spatial domain to yield the desired enhanced image. We will show that the proposed method is a generation of all kernel-design based methods. The obtained results are highly promising.

8295A-27, Session 8

Intelligent detection of impulse noise using multilayer neural network with multi-valued neurons

I. Aizenberg, G. Wallace, Texas A&M Univ.-Texarkana (United States)

Impulse noise filtering is a popular and important problem in image processing. A common disadvantage of even very sophisticated filters is smoothing of edges and smallest details during the filtering process. To take care of this problem, it is possible to detect noisy pixels prior to filtering and to apply a filter only to the pixels detected as noisy. Many impulse detectors are known. Most of them are based on the analysis of local statistical characteristics in a window around each pixel. There are some intelligent impulse noise detectors, which are based on the standard multilayer feedforward neural network (MLF) and fuzzy logic. However, all of them are based on the use of the same local statistical characteristics (which are employed by regular detectors) as the features. This makes process of noise detection multi-stage and complicated.

Since MLMVN significantly outperforms MLF in terms of learning speed, functionality, flexibility and generalization capability, when solving benchmark and different real-world problems, it is very attractive to apply it for impulse noise detection. Moreover, it is especially attractive to use just intensity values in a local window around a pixel of interest as the features (and the MLMVN inputs, respectively).

The main result, which will be presented in the paper, is a proven ability of MLMVN to detect both salt and pepper and random impulse noise with a high accuracy. Moreover, it is shown that to train MLMVN, it is enough to use a single image for the learning purposes. Then other images, which did not participate in the learning, can be processed with a high accuracy. After noisy pixels are detected, they can be efficiently filtered even using a classical median filter.

8295A-28, Session 8

An homomorphic filtering and expectation maximization approach for the point spread function estimation in ultrasound imaging

S. Benameur, Eiffel Medtech, Inc. (Canada); M. Mignotte, Univ. de Montréal (Canada); F. Lavoie, Eiffel Medtech, Inc. (Canada)

In modern ultrasound imaging systems, the spatial resolution is severely limited due to the effects of both the finite aperture and overall bandwidth of ultrasound transducers and the non-negligible width of the transmitted ultrasound beams. This low spatial resolution

remains the major limiting factor in the clinical usefulness of medical ultrasound images. In order to recover clinically important image details, which are often masked due to this resolution limitation, an image restoration procedure should be applied. To this end, an estimation of the Point Spread Function (PSF) of the ultrasound imaging system is required. This paper introduces a novel, original, reliable, and fast Maximum Likelihood (ML) approach for recovering the PSF of an ultrasound imaging system. This new PSF estimation method assumes as a constraint that the PSF is of known parametric form. Under this constraint, the parameter values of its associated Modulation Transfer Function (MTF) are then efficiently estimated using a homomorphic filter, a denoising step, and an expectation-maximization (EM) based clustering algorithm. Given this PSF estimate, a deconvolution can then be efficiently used in order to improve the spatial resolution of an ultrasound image and to obtain an estimate (independent of the properties of the imaging system) of the true tissue reflectivity function. The experiments reported in this paper demonstrate the efficiency and illustrate all the potential of this new estimation and blind deconvolution approach.

8295A-35, Poster Session

Intensity-constrained, flat-kernel filter for local dark feature suppression: application to removal of hair artifacts in dermatoscopic images

A. A. Gutenev, Retiarius Pty Ltd. (Australia)

The Intensity Constrained Flat Kernel Filtering (ICFK) scheme is a dual domain (spatial and intensity) nonlinear framework which has been shown to generate useful filters for image processing. This paper proposes a new filter developed within the ICFK framework. Although local in nature the filter is designed to suppress large scale spatial features within the image. As in every other filter derived within the scheme the suppressed features are defined by two parameters: size of the kernel and intensity range. The filter, a single-step procedure, is applied to removal of hair artifacts in skin lesion epiluminescence microscopy images, the task essential in assisting in automated segmentation of imaged area into lesion and surrounding skin. Results of the experiments on 400 dermatoscopic images of lesions with hair indicate suitability of the method as an aid in lesion segmentation by suppressing hair or vascular features near the lesion borders.

8295A-36, Poster Session

New decision support tool for acute lymphoblastic leukemia classification

M. Madhukar, S. S. Agaian, A. Chronopoulos, The Univ. of Texas at San Antonio (United States)

The analysis of blood cells in microscope image can provide useful information concerning the patient's health. Leukemia is the cancer of the blood which develops in the soft spongy center of long bones called bone marrow.There are four major different forms or types of leukemia (Acute lymphoblastic leukemia (ALL), Acute myelogenous leukemia (AML), Chronic lymphocytic leukemia (CLL), and Chronic myelogenous leukemia (CML)[1]. Acute Lymphoblastic Leukemia is characterized by accumulation of malignant cells called lymphoblasts in the bone marrow. However, the identification of ALL is usually difficult due to the variety of features and the often unclear images might result in wrong diagnosis. In this paper, we develop a new decision support tool for ALL classification. The developed system will effectively extract the cell properties. The blood images are subjected to series of pre-processing steps that effectively bring out the leukocytes from other blood components. Segmentation is done using K-means clustering for evaluating shape and texture properties of the nuclei of the leukocytes. The proposed idea effectively brings out the nuclei features of the leukocytes and thereby presents an effective means of ALL classification.

8295A-37, Poster Session

Sharpness metric for no-reference image visual quality assessment

V. V. Lukin, N. N. Ponomarenko, O. Eremeev, National Aerospace Univ. (Ukraine); K. O. Egiazarian, J. T. Astola, Tampere Univ. of Technology (Finland)

This paper proposes a metric that evaluates the sharpness of a color image. The metric also takes into account the presence of blocking artifacts and the effect of macro photography. A given image is transformed from RGB color space to color space YCbCr. Next, the contribution of blocking effect to energy of high-frequency components of the image is assessed separately for each color component. A high-frequency energy of the image is estimated in wavelet domain and corrected by the previously obtained estimate of blocking effect. And, finally, an estimate of image sharpness is formed for all color components as weighted sum of energies for color components. The effectiveness of the proposed metric has been analyzed using the test image database containing 500 color images. The correspondence between the proposed metric and mean opinion scores for the database is evaluated. Spearman correlation factor between the proposed metrics and the MOS is equal to 0.71. This is considerably higher than for other well-known metrics.

8295A-38, Poster Session

A new system of computer-aided diagnosis of skin lesions

I. A. Sanchez, S. S. Agaian, The Univ. of Texas at San Antonio (United States)

No abstract available

8295A-39, Poster Session

Image denoising using a combined criterion

E. Semenishchev, V. I. Marchuk, South-Russian State Univ. of Economics and Service (Russian Federation); K. O. Egiazarian, Tampere Univ. of Technology (Finland); V. Voronin, South-Russian State Univ. of Economics and Service (Russian Federation)

A new image denoising method is proposed in this paper. We are considering an optimization problem with a linear objective function based on two criteria, namely, L2 norm and the first order square difference; then solving it and proving a convergence and uniqueness of solutions of the developed method. This method is a parametric one, by a choice of the parameters one can tune a given criteria of the objective function. The denoising algorithm consists of the following steps: 1) multiple denoising estimates are found on local areas of the image; 2) image edges are determined; 3) parameters of the method are fixed and denoised estimates of the local area are found; 4) local window is moved to the next position (local windows are overlapping) in order to produce the final estimate.

A proper choice of parameters of the introduced method is discussed.

A comparative analysis of a new denoising method with existed ones is performed on a set of test images.

Conference 8295A

8295A-40, Poster Session

Non-cooperative stationary ground targets detection based on IRST

D. Yang, Y. Qin, X. Zhang, X. Wei, H. Wang, National Univ. of Defense Technology (China)

There are many difficulties in detection for IRST as follows: (1) complex background and lack of target's motion information; (2) lack of prior information, the non-cooperative target is difficult to match; (3) there is big span of the imaging scene distance, and the image pixel gray cannot reflect the radiation of scene and target accurately. To overcome the difficulties, the paper presented an algorithm of non-cooperative stationary ground target detection based on feature set. Firstly, the spatial station model of imaging scene was founded to count the scene distance, and then we pretreat the IR image to get the apparent temperature image which can reflect the radiation of the scene truly. Secondly, the feature set was extracted in the sliding window of the pixel neighbourhood, and the hypothesis test was used to detect the stationary targets. Lastly the sliding window was adjusted adaptively according to the target size. Simulation results demonstrate that the proposed algorithm has better detection performance to the ground non-cooperative stationary targets.

8295A-41, Poster Session

Motion-compensated spatial-temporal filtering for noisy color filter array sequence

M. S. Lee, Yonsei Univ. (Korea, Republic of)

Spatial-temporal filters have been widely used in video denoising module. The filters are commonly designed for monochromatic image. However, most digital video cameras use a color filter array (CFA) to get color sequence. We propose a recursive spatial-temporal filter using motion estimation (ME) and motion compensated prediction (MCP) for CFA sequence. In the proposed ME method, we obtain candidate motion vectors from CFA sequence through hypothetical luminance maps. With the estimated motion vectors, the accurate MCP is obtained from CFA sequence by weighted averaging, which is determined by spatial-temporal LMMSE. Then, the temporal filter combines estimated MCP and current pixel. This process is controlled by the motion detection value. After temporal filtering, the spatial filter is applied to the filtered current frame as a post-processing. Experimental results show that the proposed method achieves good denoising performance without motion blurring and acquires high visual quality.

8295A-42, Poster Session

Application of 1D FIR filter methods to 3D polygonal meshes

W. S. Ward, The Univ. of Texas at San Antonio (United States)

This paper discusses a procedure of filtering a three dimensional polygonal mesh by utilizing the basic methods of finite impulse response (FIR), one dimensional filtering. Most three dimensional filters are limited to either smoothing an object or enhancing its features. By taking a more general approach, a method was developed that can be used to either smooth or enhance objects. This method allows for either multiple iterations of a simple filter, or a single pass of a much more complex filter. This procedure uses the absolute vertex locations, as defined in the .ply file format, as the basis for filtering. Each vertex is filtered independently, using only the original unfiltered object's vertices, with the resultant vertices being stored separately. In order to use the standard methods associated with one dimensional filtering, the vertices must first be linearized in some fashion. To accomplish this, the vertex to be filtered is defined as the "origin" or "layer 1". The

vertices connected to this "layer" by polygons, extending outward, are defined as "layer 2", and so on and so forth, until the requisite number of "layers" are calculated. The number of requisite layers is defined by the number of FIR filter coefficients being used. The points in each "layer" are then averaged and are then treated as a linear set of real numbers, which are then filtered using the FIR coefficients.

8295A-43, Poster Session

An automatic approach for 3D registration of CT scans

Y. Hu, E. Saber, S. Dianat, S. R. Vantaram, Rochester Institute of Technology (United States); V. Abhyankar, DataPhysics Research, Inc. (United States)

This paper presents an automatic 3D registration algorithm which is capable of aligning freshly acquired volumetric CT scans from an arbitrary patient called a 'Study' to a reference original known as an 'Atlas' that serves as ground truth created a priori by experts personnel. The proposed algorithm consists of three steps :(i) locating a region of interest suitable for registration, (ii) 3D interpolation to bring the study and atlas to a common sampling grid, and (iii) aligning the two volumes in a registration step. The first step removes all the irrelevant objects and locates the body region in the CT volume by a series of morphological operations. The second step applies a 3D interpolation scheme to the retrieved dataset and Atlas so that they are identically sampled in x, y and z directions. The last step registers the retrieved datasets to the corresponding slices in the Atlas using a similarity measurement technique that is based on 3D correlation. The proposed technique was tested on several datasets with enhanced performance. It can be used in an automatic and unsupervised 3D segmentation of human body organs from CT scans.

8295A-44, Poster Session

Boundary handling mechanism for lifting-based spatial adaptation of filter banks

D. Jayachandra, A. Makur, Nanyang Technological Univ. (Singapore)

Time/space varying filter banks (FBs) are proved to be useful in building signal adaptive transforms. Lifting factorization of FBs allows to spatially adapt between arbitrary FBs, avoiding the need to design border FBs to complete perfect reconstruction (PR) during the transition. However, lifting based switching between arbitrarily designed FBs induces spurious transients into the resulting subbands during the transition. In this paper we propose a boundary handling mechanism that maintains good frequency response and eliminates the transients during the transition. We successfully show spatial adaptation between JPEG2000 9/7 and 5/3 FBs to reduce the ringing artifacts in images.

8295A-45, Poster Session

A simple and efficient algorithm for connected-component labeling in color images

M. E. Celebi, Louisiana State Univ. Shreveport (United States)

Connected component labeling is a fundamental operation in binary image processing. A plethora of algorithms have been proposed for this low-level operation with the early ones dating back to the 1960s. However, very few of these algorithms were designed to handle color images. In this paper, we present a simple algorithm for labeling connected components in color images using an approximately linear-time seed fill algorithm. Experiments on a large set of photographic and synthetic images demonstrate that the proposed algorithm provides fast and accurate labeling without requiring excessive stack space.

Conference 8295A

8295A-46, Poster Session

An adaptive and deterministic method for initializing the Lloyd-Max algorithm

J. Vicory, M. E. Celebi, Louisiana State Univ. Shreveport (United States)

Gray-level quantization (reduction) is an important operation in image processing and analysis. The Lloyd-Max algorithm (LMA) is a classic scalar quantization algorithm that can be used for gray-level reduction with minimal mean squared distortion. However, the algorithm is known to be very sensitive to the choice of initial centers. In this paper, we introduce an adaptive and deterministic algorithm to initialize the LMA for gray-level quantization. Experiments on a diverse set of publicly available test images demonstrate that the presented method outperforms the commonly used uniform initialization method.

8295A-47, Poster Session

Multi-resolution analysis for region of interest extraction in thermographic, nondestructive evaluation

B. Ortiz Jaramillo, H. A. Fandino Toro, Univ. Nacional de Colombia (Colombia); H. D. Benitez Restrepo, Pontificia Univ. Javeriana, Cali (Colombia); S. A. Orjuela Vargas, Univ. Gent (Belgium); G. Castellanos Dominguez, Univ. Nacional de Colombia (Colombia); W. Philips, Univ. Gent (Belgium)

The ultimate goal of a thermographic inspection is to automatically analyze images providing a pass or fail diagnostic to the operator. Primary difficulties in making useful interpretations of a thermal image is the presence of non-uniform heating caused by the uneven excitation of the surface. Moreover, non-uniform heating is hard to remove because is an inherent time variant aberration. In summary, region growing approach for image segmentation in Infrared Nondestructive Testing (INDT) is affected by ROI low contrast and non-uniform heating.

In this paper, a methodology for ROI extraction in INDT using multi-resolution analysis is proposed, which it is robust to ROI low contrast and non-uniform heating. The former methodology includes local correlation, Gaussian scale analysis and clustering analysis. In this methodology local correlation between image and Gaussian window provide interest points related to ROIs. We use a Gaussian window because thermal behavior is well modeled by Gaussian smooth contours. Besides, Gaussian scale is used to analyze details in the image using multi-resolution analysis avoiding low contrast, non-uniform heating and selection of the Gaussian window size. Finally, clustering analysis is used to provide a good combination between spatial information and gray level values obtained by Gaussian scale decomposition.

8295A-48, Poster Session

Estimation of deformations in ultrasound images using dynamic programming

S. S. Furuie, F. M. Cardoso, Escola Politécnica da Univ. de São Paulo (Brazil)

Dynamic medical images may provide valuable information such as contraction rate, deformation and elasticity. For this purpose, it is fundamental to estimate the displacement of each point of interest. However, in ultrasound this task is hampered by speckle noise. The objective of this work is the estimation of structure deformation and contraction using robust tracking of a set of representative points in a sequence of ultrasound images. The proposed approach is based on discrete optimization of joint displacement estimation, where the criteria involve joint intensity and morphology similarity. We investigated the following modifications: a) initialization of the graph by maximization of Bhattacharayya coefficient, which is invariant to translation and rotation; b) incorporation of cost due to angle differences between consecutive points in the mapping; c)

use of speckle index instead of plain intensity. We have evaluated in realistic numerical phantoms with speckle noise and compared with traditional approaches. Ten points were considered in the phantom and we applied several affine transformations to generate the deformed images. The average displacement error has decreased in almost all cases. In conclusion, we have shown that the proposed approach has improved consistently the estimation of displacements.

8295A-49, Poster Session

Combining skin texture and facial structure for face identification

R. E. Manoni, Photon Research Associates, Inc. (United States); R. L. Canosa, Rochester Institute of Technology (United States)

This research introduces a new technique for facial identification in low resolution images that combines facial structure with skin texture to accommodate changes in lighting and head pose. The model combines skin texture with facial structure. The accuracy of this model is tested on low resolution images containing different illumination and head poses. Skin texture captures the visual details of skin, mapping unique textures to a mathematical space. Facial structure uses multi-scale edges to define gross and fine structures of the face, resulting in a set of features that show the global outline and local structure of the face along with the fine structure of all facial components. These feature images are combined together to create a unique representation for each person, which is then used for identification. The facial structure model used alone had a low overall accuracy, yet showed some promise of being invariant to pose. The skin texture model used alone had a higher average accuracy and showed some invariance to illumination. The combined model had an average accuracy higher than the two models used alone. Most promising, the combined model retained the best features of the individual models with invariance to illumination and most pose angles.

8295A-50, Poster Session

Development of a human vision simulation camera and its application

H. Okumura, M. Fukusaki, S. Takubo, K. Arai, Saga Univ. (Japan)

Human eye has a lot of photoreceptor cells in its retina. Human photoreceptor cells consist of "cone" type and "rod" type.

Cone cells that are responsible for color vision are concentrated at the macula and rod cells are concentrated arround the macula and are used in peripheral vision. More sensitive than cone cells, rod cells are almost entirely responsible for scotopic vision. Purkinje effect for mesopic and scotopic vision and adaptation are caused by uneven distribution and sensitivity of photoreceptor cells. In this study, HuVisCam, a human vision simulation camera, that can simulate not only Purkinje effect for mesopic and scotopic vision but also dark and light adaptation, abnormal miosis and abnormal mydriasis caused by the influence of mydriasis medicine or nerve agent and specific color perception is developed. In this article, the system and its method are described. The applied example to onboard visibility support camera which can enhance the object which visibility decreased by Purkinje effect under the mesopic vision is also demonstrated.

8295A-51, Poster Session

Reconstruction from divergent ray projections

C. S. Sastry, International Institute of Information Technology (India); S. Singh, Siemens Information Systems Ltd. (India)

Despite major advances in x-ray sources, detector arrays, gantry mechanical design and specially computer performances, but computed tomography (CT) enjoys the filtered back projection (FBP) algorithm as their first choice for the CT image reconstruction in the commercial scanners. Over the years, a lot of fundamental work has

been done in the area of finding the sophisticated solutions for the inverse problems using different kinds of optimization techniques. Recent last few years have really been dominated by the compressive sensing techniques and/or sparse reconstruction techniques. Still there is a long way to go for translating these newly developed algorithms in the clinical environment. The reasons are not obvious and seldom discussed. Knowing the fact that filtered back projection is one of the most popular CT image reconstruction algorithms, it allows us to do research work to improve the different error estimates at different steps preformed in filtered back projection.

In this paper, we present a back projection formula for the reconstruction of divergent beam tomography with unique convolution structure. Using such a proposed approximate convolution structure, the approximation error mathematically justifies the low reconstruction error for a suitable choice of parameters.

In order to minimize the exposure time and possible distortions due to the motion of the patient, fan beam method of collection of data is used. Rebinning transformation is used to connect fan beam data into parallel beam data so that the well developed method of image reconstruction for parallel beam geometry can be used. The computational errors involved in the numerical process of rebinning, some degradation of image is inevitable. However, to date a little work has been done for the reconstruction of fan beam tomography. There has been some recent results on wavelet reconstruction of divergent beam tomography. In this paper, we propose a convolution algorithm for the reconstruction of divergent beam tomography, which is simpler than wavelet methods and better approximate error methods. As the formula is approximate in nature, we prove an estimate for the error associated with the formula. Using the estimate, we deduce condition that minimizes approximation error.

8295A-52, Poster Session

Fusing electro-optic and infrared signals for high-resolution night images

X. Huang, Stevens Institute of Technology (United States); R. Netravali, Columbia Univ. (United States); H. Man, V. B. Lawrence, Stevens Institute of Technology (United States)

Electro-optic image exhibits the properties of high resolution and low noise level. Therefore, we propose a novel framework of infrared image enhancement based on the information of electro-optic image for the first time. On one hand, we adopt the Russell C. Hardie et al. developed theoretical point spread function (PSF) of the infrared image system, which is contributed by the modulation transfer function (MTF) of a uniform detector array and the incoherent optical transfer function (OTF) of diffraction-limited optics. The effect of the integration of light intensity over the span of the detectors can be modeled as a linear convolution operation with a PSF determined by the geometry of a single detector. The unknown parameters in this PSF include the detector square detector size and the system cutoff frequency. On the other hand, we regard a registered image pair as the electro-optic and infrared output images. Additionally, the designed inverse filter and Wiener filter are adopted to reconstruct the image without and with noise, respectively. The infrared image enhancement is achieved by establishing an optimization model, which aims at obtaining the optimal values of variable parameters in the theoretical PSF (e.g., system cutoff frequency). This model subjects to minimize the error between the reconstructed infrared image and the transformed electro-optic image. Simulation results compare the difference between the reconstructed infrared image with and without the information of electro-optic image, respectively.

8295A-53, Poster Session

Texture and color descriptors as a tool for context-aware patch-based image inpainting

T. Ruzic, A. Pizurica, W. Philips, Univ. Gent (Belgium)

Image inpainting is an image processing task of lling in missing, damaged or unwanted parts of the image. State-of-the-art results are obtained with patch-based methods that ll in the missing region patch-by-patch by searching for similar patches in the known region and placing them at corresponding locations. In this paper, we introduce a context-aware patch-based inpainting method, where the context is represented by texture and color features of a block surrounding the patch to be lled in. We use this context to recognize other blocks in the image that have similar features and then we constrain the search for similar patches within them. Such an approach guides the search process towards less ambiguous matching candidates, while also speeding up the algorithm. Results demonstrate the qualitative improvement over the related technique with less visual artefacts, but also with decrease in computation time.

8295A-29, Session 9

Hybrid gesture recognition system for short-range use

A. Minagawa, Fujitsu Labs. (Japan); W. Fan, Fujitsu Research and Development Center Co., Ltd. (China); Y. Katsuyama, H. Takebe, N. Ozawa, Y. Hotta, Fujitsu Labs. (Japan); J. Sun, Fujitsu Research and Development Center Co., Ltd. (China)

In recent years, various gesture recognition systems have been studied for use in television and video games. In such systems, motion areas ranging from 1 to 3 meters deep have been evaluated. However, with the burgeoning popularity of small mobile displays, gesture recognition systems capable of operating at much shorter ranges have become necessary. The problems related to such systems are exacerbated by the fact that the camera's field of view is unknown to the user during operation, which imposes several restrictions on his/her actions.

To overcome the restrictions generated from such mobile camera devices, and to create a more flexible gesture recognition interface, we propose a hybrid hand gesture system, in which two types of gesture recognition modules are prepared and with which the most appropriate recognition module is selected by a dedicated switching module. The two recognition modules of this system are shape analysis using a boosting approach (detection-based approach) and motion analysis using image frame differences (motion-based approach).

We evaluated this system using sample users and classified the resulting errors into three categories: errors that depend on the recognition module, errors caused by incorrect module identification, and errors resulting from user actions. In this paper, we show the results of our investigations and explain the problems related to short-range gesture recognition systems.

8295A-30, Session 9

Tracking white road line by particle filter from the video sequence acquired by the camera attached to a walking human body

S. Takahashi, J. Ohya, Waseda Univ. (Japan)

This paper proposes a method for tracking and recognizing the white line marked in the surface of the road from the video sequence acquired by the camera attached to a walking human, towards the actualization of an automatic navigation system for the visually handicapped. Our proposed method consists of two main modules: (1) Particle Filter based module for tracking the white line, and (2) CLAFIC Method based module for classifying whether the tracked object is the white line. In (1), each particle is a rectangle, and is described by its centroid's coordinates and its orientation. The likelihood of a particle is computed based on the number of white pixels in the rectangle. In

(2), in order to obtain the ranges (to be used for the recognition) for the white line's length and width, Principal Component Analysis (PCA) is applied to the covariance matrix obtained from valid sample particles (centroid's coordinates and orientation). At each frame, PCA is applied to the covariance matrix constructed from particles with high likelihood, and if the obtained length and width are within the above-mentioned ranges, it is recognized as the white line. Experimental results using real video sequences show the validity of the proposed method.

8295A-31, Session 9

Driver/passenger discrimination for the interaction with the dual-view touch screen integrated to the automobile centre consol

E. Herrmann, A. Makrushin, J. Dittmann, Otto-von-Guericke-Univ. Magdeburg (Germany); C. Vielhauer, Fachhochschule Brandenburg (Germany)

In an attempt to further develop and evaluate the optical recognition systems for distinguishing between driver and front-seat passenger during their interactions with dual-view touch screen integrated to the automobile centre consol, this work focuses on the enhancement of both image processing algorithms and experimental environment. In addition to the motion based forearm and hand segmentation and the texture based arm direction analysis, the boosting classifiers with haar-like features have been engaged for the learning of driver's and passenger's hand patterns. The user discrimination system was completely reproduced in a laboratory, including passenger compartment with genuine dashboard, touch screen, camera and infrared lamps, so that different illumination conditions could be modeled. The new acquisition system allows automatic and unambiguous registration of all touch screen interactions and their synchronization with the video stream. This results in credible evaluation of the image processing routines. The adjustment of the camera position and the active infrared illumination made it possible to reduce the recognition error rates and to achieve superior discrimination performance comparing to previous works. In the further work the prototypic user discrimination system will be migrated from the standing car simulator to a real car and the experiments will be provided for moving car scenarios as well.

8295A-32, Session 10

A linear filter design technique for equalizing document scanners

A. H. Eid, Lexmark International, Inc. (United States)

In this paper, we propose a new technique to automatically restore the sharpness of blurred documents by equalizing the frequency response of given scanners using linear filters.

To measure the blur characteristics of a scanning device, we measure its both horizontal and vertical Spatial Frequency Response (SFR). Starting from the measured SFR of the scanning device, our goal is to design an equalizing filter so that the combined SFR of the equalizing filter and the scanner resembles a perfect SFR.

The desired 2D frequency response of the filter is computed using bilinear interpolation of the horizontal and vertical responses derived from the corresponding SFRs of the scanner. The filter design technique is two steps. First, a linear system of equations is constructed using the unknown filter coefficients and the desired filter 2D frequency response. The linear least square method is used to solve the linear system of equations. The second step of the filter design uses a non-linear optimization technique to refine the results of the first step.

Our experimental results show that this automated process can be applied to different document scanning devices to equalize their spatial frequency response resulting in consistent output sharpness levels.

8295A-33, Session 10

Application of spatial contrast techniques

on satellite imagery for cloud shape differentiation

J. R. Dim, H. Murakami, Japan Aerospace Exploration Agency (Japan)

Pixels' edges can yield useful information on physical properties of objects featured on satellite images. These properties can be derived through the use of the imagery spatial contrast techniques. To differentiate various cloud types based on their shapes, one of these techniques is applied on thermal images from a polar orbiting satellite, the National Oceanic and Atmospheric Administration/Advanced Very-High-Resolution Radiometer (NOAA-AVHRR). Edge detectors gradients extracted from daily global cloud temperature images of this satellite and the spatial relationship between these gradients permit the distinction of nine major cloud shapes distributed along three cloud pressure levels (high, middle and low). The cloud shape differentiation method utilized is a histogram-based gradient scheme describing the occurrence of different gradients' levels (high, middle and low) in each block of pixels. A detailed analysis of the distribution of the cloud shapes obtained is conducted, and the frequency of each cloud shape is evaluated with another cloud classification method (based on cloud optical properties) for validation purposes. Finally, implications of the results obtained, on the estimation of the impact of cloud shapes variations on the recent climate are discussed.

8295A-34, Session 10

A multi-step system for screening and localization of hard exudates in retinal images

A. S. Bopardikar, V. Bhola, R. B. S., R. Narayanan, Samsung Electronics, India Software Operations Ltd. (India)

The number of people being affected by Diabetes mellitus worldwide is increasing at an alarming rate. Monitoring of the diabetic condition and its effects on the human body are therefore of great importance. Of particular interest is diabetic retinopathy (DR) which is a result of prolonged, unchecked diabetes and affects the visual system. DR is a leading cause of blindness throughout the world. At any point of time 25-44% of people with diabetes are afflicted by DR. Automation of the screening and monitoring process for DR is therefore essential for efficient utilization of healthcare resources and optimizing treatment of the affected individuals. Such automation would use retinal images and detect the presence of specific artifacts such as hard exudates, hemorrhages and soft exudates (that may appear in the image) to gauge the severity of DR. In this paper, we focus on the detection of hard exudates. We propose a two step system that consists of a screening step that classifies retinal images as normal or abnormal based on the presence of hard exudates and a detection stage that localizes these artifacts in an abnormal retinal image. The proposed screening step automatically detects the presence of hard exudates with a high sensitivity and positive predictive value (PPV). The detection/localization step uses a k-means based clustering approach to localize hard exudates in the retinal image. Suitable feature vectors are chosen based on their ability to isolate hard exudates while minimizing false detections. The algorithm was tested on a benchmark dataset (DIARETDB1) and was seen to provide a superior performance compared to existing methods. The two-step process described in this paper can be embedded in a tele-opthamology system to aid with speedy detection and diagnosis of the severity of DR.

Conference 8295B: Parallel Processing for Imaging Applications II

Monday 23 January 2012

Part of Proceedings of SPIE Vol. 8295B Parallel Processing for Imaging Applications II

8295B-52, Session 11

GPGPU-based surface inspection from structured white light

M. Bordallo Lopez, Univ. of Oulu (Finland); K. Niemelä, VTT Technical Research Ctr. of Finland (Finland); O. J. Silvén, Univ. of Oulu (Finland)

Automatic surface inspection has been used in the industry to reliably detect all kinds of surface defects and to measure the overall quality of a produced piece. Structured light systems (SLS) are based on the reconstruction of the 3D information of a selected area by projecting several phase-shifted sinusoidal patterns onto a surface.

Due to the high speed of production lines, surface inspection systems require extremely fast imaging methods and lots of computational power. The cost of such systems can easily become considerable. The use of standard PCs and Graphics Processing Units (GPUs) for data processing tasks facilitates the construction of cost-effective systems.

We present a parallel implementation of the required algorithms written in C with CUDA extensions. In our contribution, we describe the challenges of the design on a GPU, compared with a traditional CPU implementation. We provide a qualitative evaluation of the results and a comparison of the algorithm speed performance on several platforms.

The system is able to compute two megapixels height maps with 100 micrometers spatial resolution smaller than 100 micrometers in less than 200ms on a mid-budget laptop. Our GPU implementation runs about ten times faster than our previous C code implementation.

8295B-54, Session 11

IMPAIR-GPU: massively parallel deconvolution algorithm for GPUs

M. Sherry, A. Shearer, National Univ. of Ireland, Galway (Ireland)

Medical image processing has been fundamentally restricted by CPU power or by the high costs of large parallel systems. In the past we have developed a suite of software which demonstrated the standard deconvolution techniques, such as Richard-Lucy deconvolution when combined with a wavelet based denoising can produce significant

enhancements of large radiographic images. In this paper we review our previous approaches and limitations and then move onto describing the use of GPUs to implement these algorithms for a number of different medical imaging modalities. In particular we concentrate upon our demonstration code developed on an NVIDIA Tesla platform and we

compare its performance with a conventional MPI approach. Furthermore we will also report on the first results of a semantic approach to determining the coefficients of the deconvolution and de-noising processes to enhance specific image features. Our real-time image processing requirement comes from the application area - radio-therapy portal imaging.

8295B-55, Session 11

Parallel processing architectures for H.264 deblocking filter on multicore platforms

D. P. Prasad, S. Sonachalam, M. K. Kunchamwar, N. R. Gunupudi, Parallel Prisms (United States)

Massively parallel computing (multi-core) chips offer outstanding new solutions that satisfy the increasing demand for high resolution and

high quality video compression technologies such as H.264. Such solutions not only provide exceptional quality but also efficiency, low power, and low latency, previously unattainable in software based designs. While custom hardware and Application Specific Integrated Circuit (ASIC) technologies may achieve low-latency, low power, and real-time performance in some consumer devices, many applications require a flexible and scalable software-defined solution.

The deblocking filter in an H.264 encoder/decoder poses difficult implementation challenges because of heavy data dependencies and the conditional nature of the computations. Deblocking filter implementations tend to be fixed and difficult to reconfigure for different needs. The ability to scale up for higher quality requirements such as 10-bit pixel depth or a 4:2:2 chroma format often reduces the throughput of a parallel architecture designed for lower feature set. A scalable architecture for deblocking filtering, created with a massively parallel processor based solution, means that the same encoder or decoder will be deployed in a variety of applications, at different video resolutions, for different power requirements, and at higher bit-depths and better color sub sampling patterns like YUV, 4:2:2, or 4:4:4 formats.

This work describes a scalable parallel architecture for an H.264 compliant deblocking filter. Parallel techniques such as parallel processing of independent macroblocks, sub blocks, and pixel row level are examined in this work. The deblocking architecture consists of a basic cells called deblocking filter unit (DFU) and dependent data buffer manager (DFM). The DFU can be used in several instances, catering to different performance needs; the DFM serves the data required for the different number of DFUs, and also manages all the neighboring data required for future data processing of DFUs. This approach achieves the scalability, flexibility, and performance excellence required in deblocking filters.

8295B-57, Session 12

Interactive plenoptic rendering with GPUs

A. Lumsdaine, G. N. Chunev, Indiana Univ. (United States); T. G. Georgiev, Adobe Systems Inc. (United States)

Processing and rendering of plenoptic camera data requires significant computational power and memory bandwidth. At the same time, interactive rendering performance is highly desirable so that users can explore the infinite variety of images that can be rendered from a single plenoptic image. In this paper we describe a GPU-based approach for lightfield processing and rendering, with which we are able to achieve interactive performance for focused plenoptic rendering tasks such as refocusing and novel-view generation. We present a progression of rendering approaches for focused plenoptic camera data and analyze their performance on popular GPU-based systems. Our analyses are validated with experimental results on commercially available GPU hardware. Even for complicated rendering algorithms, we are able to render 39Mpixel plenoptic data to 2Mpixel images with frame rates in excess of 500 frames per second.

8295B-58, Session 12

Three-level GPU accelerated Gaussian mixture model for background subtraction

Y. Li, G. Wang, X. Lin, Tsinghua Univ. (China)

Gaussian Mixture Model (GMM) for background subtraction (BGS) is widely used for detecting and tracking objects in video sequences. Despite the satisfying results provided by GMM, low processing speed has become the bottleneck for real-time applications. We propose a

novel method to accelerate the GMM algorithm based on graphics processing unit (GPU). As GPU excels at performing massively parallel operations, the novelty lies in how to adopt various optimization strategies to fully exploit GPU's resources. The parallel design consists of three levels. On the basis of first-level implementation, we employ techniques such as memory access coalescing and memory address saving to the second-level optimization and the third-level modification, which reduces the time cost and increases the bandwidth greatly. Experimental results demonstrate that the proposed method can yield performance gains of 145 frames per second (fps) for VGA (640*480) video and 505 fps for QVGA (320*240) video which outperform their CPU counterparts by 24X and 23X speedup respectively. Compared with the previous related work of 10X speedup, the proposed method gains 22X speedup averagely. The resulted surveillance system can process five VGA videos simultaneously with strong robustness and high efficiency.

8295B-59, Session 12

Plane-dependent error diffusion on a GPU

Y. Zhang, Univ. of California, Davis (United States); J. L. Recker, Hewlett-Packard Labs. (United States); R. A. Ulichney, I. Tastl, Hewlett-Packard Co. (United States); J. D. Owens, Univ. of California, Davis (United States)

In this paper, we study a plane-dependent technique that reduces dot-on-dot printing in color images, and apply this technique to a GPU-based error diffusion halftoning algorithm. We design image quality metrics to preserve mean color and minimize colorant overlaps.

We further use randomized intra-plane error filter weights to break periodic structures. Our GPU implementation achieves a processing speed of 200 MegaPixels/second for RGB color images, and a speedup of 30-37x over a multi-threaded implementation on a dual-core CPU. Since the GPU implementation is memory bound, we essentially get the image quality benefits for free by adding arithmetic complexities for inter-plane dependency and error filter weights randomization.

8295B-60, Session 12

An analysis of OpenCL for portable imaging

R. J. Moore, 3M Co. (United States); B. Zimmer, 3M Co. (United States) and Univ. of Wisconsin-Eau Claire (United States)

In this paper, we examine how well OpenCL delivers on this multi target promise for different image processing algorithms. Both GPU (Nvidia and AMD) and CPU (AMD and Intel) platforms are explored to determine how OpenCL performs using the same code on different hardware and driver configurations. We also compare OpenCL with optimized CPU and GPU (CUDA) versions of the same imaging algorithms. We present our findings and share some interesting observations on the process of using OpenCL. The imaging algorithms include basic color transformation, convolution, and visual attention saliency map calculation. The last algorithm includes many different imaging calculations: difference of Gaussian, color features, image statistics, FFT filtering, and assorted other algorithms. The saliency map calculation provides a complex real world imaging application to use for comparison of the test platforms.

8295B-61, Poster Session

Speed-up on GPU in Bayesian image reconstruction

S. Horiuchi, S. Yoshida, Z. Ushiyama, M. Yamamoto, Tokyo Univ. of Science (Japan)

There is a method for image processing that uses Bayes' theorem which is the statistical technique. This technique can be applied to optical systems, and can handle shift-variant system. Therefore, this technique is effective and can be used for wide application. However, it requires high computational cost because order of complexity is in proportion to $O(n^6)$ in two dimensional problems. Then, we propose a calculation method using GPGPU (General-purpose computing on GPU) in the Bayesian image reconstruction for speed-up of calculation.

Conference 8296: Computational Imaging X

Monday-Tuesday 23-24 January 2012

Part of Proceedings of SPIE Vol. 8296 Computational Imaging X

8296-01, Session 1

Imaging with electrons: a review of modern modalities

M. DeGraef, United States (United States)

No abstract available.

8296-02, Session 1

Improving boundary localization in the statistical image segmentation of materials micrographs

M. L. Comer, Purdue Univ. (United States)

No abstract available.

8296-03, Session 1

Scanning transmission electron tomography and applications to materials science

L. F. Drummy, Air Force Research Lab. (United States)

No abstract available.

8296-04, Session 1

Combining global labeling and local relabeling for metallic image segmentation

J. W. Waggoner, Univ. of South Carolina (United States); J. Simmons, Air Force Research Lab. (United States); S. Wang, Univ. of South Carolina (United States)

Analyzing the microstructure of metals plays an important role in materials science and industry. Typically, the microstructure of a metal sample is made up of a large number of grains which can be imaged by a microscope. In this paper, we develop an effective approach to automatically extract these grains by segmenting a sequence of 2D microscopic images that are taken from continuous serial sections of a metal sample. In our approach, we repeatedly propagate a 2D segmentation from one slice to another, and we formulate this process as an optimal labeling problem that can be efficiently solved by the graph-cut algorithm. While the underlying 2D grain topology in neighboring slices is largely consistent, it may show local variations when the serial-section plane moves into a new grain or moves out of an existing grain in 3D space. We address this problem by first running a global labeling to propagate the global grain topology, followed by a series of local relabelings to identify and correct the local topology changes. Finally, we test the performance of the proposed approach on a sequence of serial-section microscopic images of a titanium sample and show it performs better than several existing image segmentation methods.

8296-05, Session 1

Towards automated detection of active colitis in images of H&E-stained tissue samples

M. McCann, R. Bhagavatula, Carnegie Mellon Univ. (United States); M. Fickus, Air Force Institute of Technology (United States); J. A. Ozolek, Children's Hospital of Pittsburgh (United States); J. Kovacevic, Carnegie Mellon Univ. (United States)

No abstract available.

8296-06, Session 1

Computer-aided fiber analysis for crime scene forensics

M. Hildebrandt, A. Makrushin, J. Dittmann, C. Arndt, Otto-von-Guericke-Univ. Magdeburg (Germany)

The forensic analysis of fibers is currently completely manual and, therefore, time consuming process. The automation of analysis steps can significantly support forensic experts and reduce the time required for the investigation. Moreover, a subjective expert belief is extended by objective machine estimation. This work proposes the pattern recognition pipeline containing the digital acquisition of a fiber media, the pre-processing for fibers segmentation, and the extraction of the distinctive characteristics of fibers. Currently, the length, width, color, frequency of occurrence and distribution of fibers are investigated. In order to support the automatic classification of fibers the supervised and unsupervised machine learning algorithms are engaged. The first experimental setup includes a car seat and two different clothes. Staying in the seat, a test person leaves textile fibers behind. The test aims at automatic distinguishing of clothes through the fiber traces gained from the seat with the help of adhesive tape. The digitalization is provided by flatbed scanner and a chromatic white light sensor. The non-destructive contact-less sensors studied in this work, can be potentially used for the direct acquisition of the suspected surface, making adhesive tape unnecessary. Furthermore, an automated acquisition and analysis allow the processing of large areas with numerous different fibers.

8296-07, Session 1

An automated diagnostic aid for otitis media

A. Kuruvilla, Carnegie Mellon Univ. (United States); P. H. Hennings Yeomans, Ontario Institute for Cancer Research (Canada); P. Quelhas, Instituto de Engenharia Biomédica (Portugal); A. Hoberman, Univ. of Pittsburgh (United States); J. Kovacevic, Carnegie Mellon Univ. (United States)

No abstract available.

8296-08, Session 1

3D reconstruction based on single-particle cryo electron microscopy images as a random signal in noise problem

Q. Wang, Cornell Univ. (United States); Y. Zheng, Lawrence Berkeley National Lab. (United States); P. C. Doerschuk, Cornell Univ. (United States)

No abstract available.

Conference 8296

8296-09, Session 1

Highly scalable methods for exploiting a label with unknown location in order to orient a set of single-particle cryo electron microscopy images

C. J. Prust, Milwaukee School of Engineering (United States); P. C. Doerschuk, Cornell Univ. (United States); J. E. Johnson, The Scripps Research Institute (United States)

A highly scalable method for determining the projection orientation of each image in a set of cryo electron microscopy images of a labeled particle is proposed. The method relies on the presence of a label that is a sufficiently strong scatterer such that its 2-D location in each image can be restricted to at most a small number of sites by processing applied to each image individually. It is not necessary to know the 3-D location of the label on the particle. After first determining the possible locations of the label in the 2-D images in parallel, the information from all images is fused to determine the 3-D location of the label on the particle and then the 3-D location is used to determine the projection orientation for each image by processing each image individually. With projection orientations, many algorithms exist for computing the 3-D reconstruction. The performance of the algorithm is studied as a function of the label SNR.

8296-10, Session 2

Plenoptic camera with freely movable microlenses

T. G. Georgiev, Adobe Systems Inc. (United States); S. Goma, Qualcomm Inc. (United States)

The idea of capturing the 4D radiance, or plenoptic function, for the purpose of 3D photography was introduced by Lippmann in 1908. Since then it has gone through several modifications, including digital processing(Adelson 1992, the "plenoptic camera"), digital refocusing (Ng 2005), superresolution, and others).

The main practical problem of plenoptics is the low resolution of the final rendered image relative to that of a conventional camera using the same sensor. This leads to the idea of building a plenoptic camera that captures "the best of both worlds": It should easily convert into a mode equivalent to a 2D camera of high resolution.

We present the results of our work on building such plenoptic camera, based on a microlens array that can be moved freely with micromotors that position it at prcise distances from silicon die of the CCD. As a result the camera changes from capturing 3D at low spatial resolution -- to capturing 2D with very high spatial resolution after computational processing. Switching between modes is done on the fly, within a second.

We also use the two motors to tilt the microlens array, thus implementing shift-tilt plenoptic camera, greatly increasing usable depth of field. Our paper shows experimental results.

8296-11, Session 2

Image reconstruction using projections from a few views by discrete steering combined with DART

J. Kwon, S. M. Song, B. Kauke, D. P. Boyd, TeleSecurity Sciences, Inc. (United States)

In this paper, we propose an algebraic reconstruction technique (ART) based discrete tomography method to reconstruct an image accurately using projections from a few views. We specifically consider the problem of reconstructing an image of bottles filled with various types of liquids from X-ray projections. By exploiting the fact that bottles are usually filled with homogeneous materials, it is possible to obtain accurate reconstruction only with a few projections by ART. In order to deal with various types of liquids in our problem, we first introduce our discrete steering that is a modification of binary steering approach of Censor et al for our multi-valued discrete reconstruction. The main idea of steering approach is to use slowly varying thresholds instead of fixed ones. We further increase reconstruction accuracy by reducing the number of variables in ART by combining our discrete steering with the discrete ART (DART) that fixes the values of interior pixels of segmented regions considered as reliable. By various simulation studies, we show that our discrete steering combined with DART yields superior reconstruction than both discrete steering only and DART only cases and quite accurate results even with a few views such as less than five.

8296-12, Session 2

One-dimensional control grid interpolation-based demosaicing and color image interpolation

C. M. Zwart, D. H. Frakes, Arizona State Univ. (United States)

We recently reported good results with our image interpolation algorithm, One-Dimensional Control Grid Interpolation (1DCGI), in the context of grayscale images. 1DCGI has high quantitative accuracy, flexibility with respect to scaling factor, and low computational cost relative to similarly performing methods. Here we look to extend our method to the demosaicing of Bayer-Patterned images and interpolation of colored images. 1DCGI-based demosaicing performs quantitatively better (higher PSNR) than the gradient-corrected linear interpolation method of Malvar. 1DCGI interpolation of full color images performs quantitatively better than covariance-based approaches, such as iNEDI and NEDI, at substantially lower computational cost.

8296-13, Session 2

Limited view angle iterative CT reconstruction for transportation security application

S. J. Kisner, C. A. Bouman, Purdue Univ. (United States)

No abstract available.

8296-15, Session 2

Variational semi-blind sparse image reconstruction with application to MRFM

S. U. Park, A. O. Hero, Univ. of Michigan (United States); N. Dobigeon, Univ. de Toulouse (France)

This paper addresses the problem of joint image reconstruction and point spread function (PSF) estimation when the PSF of the imaging device is only partially known. To solve this semi-blind deconvolution problem, prior distributions are specified for the PSF and the 3D image. Joint image reconstruction and PSF estimation is then performed within a Bayesian framework, using a variational algorithm to estimate the posterior distribution. The image prior distribution imposes an explicit atomic measure that corresponds to image sparsity. Simulation results demonstrate that the semi-blind deconvolution algorithm compares favorably with previous Markov chain Monte Carlo (MCMC) version of myopic sparse reconstruction. It also outperforms non-myopic algorithms that rely on perfect knowledge of the PSF. The algorithm is illustrated on real data from magnetic resonance force microscopy (MRFM).

Conference 8296

8296-16, Session 3

Moon search algorithms for NASA's Dawn mission to asteroid Vesta

N. Memarsadeghi, L. A. McFadden, D. R. Skillman, NASA Goddard Space Flight Ctr. (United States); B. McLean, M. Mutchler, Space Telescope Science Institute (United States)

A moon or natural satellite is a celestial body that orbits a planet or planet-like body. Scientists seek understanding the origin and evolution of our solar system by studying moons of planets and planet-like bodies, called asteroids and dwarf planets. Additionally, searches for satellites of planetary bodies can be important to protect the safety of a spacecraft as it approaches or orbits a planetary body. If a satellite of a celestial body is found, the mass of that body can be calculated once its orbit is determined. Ensuring the Dawn spacecraft's safety on its mission to the asteroid (4) Vesta primarily motivated the work of Dawn's Satellite Working Group (SWG) in summer of 2011. Dawn mission scientists and engineers utilized various computational tools and techniques for Vesta's satellite search. The objectives of this paper are to 1) introduce the natural satellite search problem, 2) present the computational challenges, approaches, and tools used when addressing the problem, and 3) describe applications of various image processing and computational algorithms for performing satellite searches to the electronic imaging and computer science community. Furthermore, we hope that this communication would enable Dawn mission scientists to improve their satellite search algorithms and tools and be better prepared for performing the same investigation in 2015, when the spacecraft is scheduled to approach and orbit the dwarf planet (1) Ceres.

8296-17, Session 3

CLEAN: a false alarm reduction method for SAR CCD

R. Phillips, MIT Lincoln Lab. (United States)

No abstract available.

8296-18, Session 3

Insertion of synthetic features in SAR CCD imagery

E. Turner, R. Phillips, M. Cha, MIT Lincoln Lab. (United States)

No abstract available.

8296-19, Session 3

Multichannel hierarchical image classification using multivariate copulas

A. Voisin, V. Krylov, INRIA Sophia Antipolis - Méditerranée (France); G. Moser, S. B. Serpico, Univ. degli Studi di Genova (Italy); J. Zerubia, INRIA Sophia Antipolis - Méditerranée (France)

This paper focuses on the classification of multichannel images. The proposed supervised Bayesian classification method, applied to histological (medical) and SAR (remote sensing) imagery, consists of two steps.

The first deals with the statistical modeling of the coregistered input images (for instance RGB optical image, or single-channel SAR image combined with an extracted textural feature). For each class and each channel in this stacked-vector input dataset, the class-conditional marginal probability density functions (PDFs) are estimated by finite mixtures of well-chosen parametric families. For optical imagery, the normal distribution is a generally accepted model. For SAR imagery, we selected generalized Gamma, log-normal, Nakagami and Weibull distributions. Next, the multivariate d-dimensional Clayton copulas

(d being the number of input channels) are applied to estimate multivariate joint class-conditional statistics.

As a second step, we plug the estimated joint PDFs into a contextual model by using two different Markov-based methods. The first model uses a spatial context via a hidden Markov random field (MRF) model. The second model uses a multiscale approach via a hierarchical Markovian model based on a quad-tree structure. Multiscale features are extracted by discrete wavelet transforms. We integrate an exact estimator of the marginal posterior mode (MPM) with the SEM-copula PDF estimates. Such a hierarchical model supports a feasible future extension to input multi-resolution imagery.

8296-32, Session 3

Figure-ground organization is easier than previously thought

Y. Li, T. Kwon, Purdue Univ. (United States); L. J. Latecki, Temple Univ. (United States); Z. Pizlo, Purdue Univ. (United States)

No abstract available.

8296-20, Session 4

Denoising and deblurring of Fourier-trasform infrared spectroscopic imaging

T. H. Nguyen, R. K. Reddy, M. J. Walsh, M. Schulmerich, G. Popescu, M. N. Do, R. Bhargava, Univ. of Illinois at Urbana-Champaign (United States)

Fourier Transform-Infrared (FT-IR) Spectroscopic Imaging is a valuable tool not only to obtain chemical information of tissue samples but also to greatly assist pathology process such as cancer tumor localization. However, current limitations of FT-IR imaging include long acquisition time, vast data storage, low signal-to-noise ratio (SNR), and low spatial resolution due to contaminated noise, pixelation and optical effects such as optical diffraction and Mie scattering. This paper proposes a framework to tackle these problems by exploiting low-rank property of the data, modeling diffraction-limited incoherent point spread function, simultaneously deblurring and estimating unknown information of the image, additive noise and the point spread function with a Bayesian variational approach. By increasing magnification factor in the optical setup to reduce pixelation and combining with a deconvolution algorithm, we will show that more information below conventional diffraction limit can be obtained from not only the short wavelength region but also from the long wavelength region, which was currently limited by optical diffraction.

8296-21, Session 4

Iterative weighted risk estimation for nonlinear image restoration with analysis priors

S. Ramani, J. Rosen, Z. Liu, J. A. Fessler, Univ. of Michigan (United States)

Image restoration algorithms depend on a number of parameters that require proper selection to obtain meaningful results. Mean squared error (MSE) is a widely used metric for assessing image quality and parameter selection, but requires the knowledge of the unknown noise-free image and is therefore not practicable. Stein's unbiased risk estimate (SURE) provides a means for the estimation of MSE purely from the data and some first order derivative information of the algorithm (i.e., the divergence of the linear / nonlinear restoration operator with respect to the data) for the case of additive Gaussian noise. In this work, we use the principle underlying SURE to estimate the predicted mean squared error (PMSE) for parameter selection for image restoration. Earlier works that apply the SURE-principle for parameter selection in restoration have focused either on linear algorithms or nonlinear algorithms for synthesis priors. In this work, our

goal is to estimate PMSE for nonlinear algorithms with analysis priors which include popular criteria such as total variation and smooth edge-preserving regularizers. We propose to derive an analytical expression for the Jacobian matrix (whose trace yields the desired divergence) associated with the algorithm that can be recursively updated as the algorithm evolves. For this, we focus on the iterative reweighed least-squares, which is gradient-descent-based algorithm that can handle a variety of regularization criteria. Through numerical simulations, we show that PSURE (the estimate of PMSE) accurately predicts PMSE and yields regularization parameters close to those that minimize the "unknown" MSE, indicating its potential for parameter selection in image restoration.

8296-22, Session 4

Nonlocal transform-domain denoising of volumetric data with groupwise adaptive variance estimation

M. T. Maggioni, A. Foi, Tampere Univ. of Technology (Finland)

We propose an extension of the BM4D volumetric filter to the denoising of data corrupted by spatially non-uniform noise.

BM4D implements the grouping and collaborative filtering paradigm, where similar cubes of voxels are stacked into a four-dimensional "group". Each group undergoes a sparsifying four-dimensional transform, that exploits the local correlation among voxels in each cube and the nonlocal correlation between corresponding voxels of different cubes. Thus, signal and noise are effectively separated in transform domain. In this work we take advantage of the sparsity induced by the four-dimensional transform to provide a spatially adaptive estimation of the noise variance by applying a robust median estimator of the absolute deviation to the spectrum of each filtered group. The adaptive noise estimates are then used during coefficients shrinkage. Finally, the inverse four-dimensional transform is applied to the filtered group, and each individual cube estimate is adaptively aggregated at its original location.

Experiments on medical data corrupted by spatially varying Gaussian and Rician noise demonstrate the efficacy of the proposed approach in volumetric data denoising. In case of magnetic resonance signals, the adaptive variance estimate can be also used to compensate the estimation bias due to the non-zero-mean errors of the Rician-distributed data.

8296-23, Session 4

Non-uniform contrast correction for coded source neutron imaging

H. J. Santos-Villalobos, P. R. Bingham, Oak Ridge National Lab. (United States)

There is a clear demand for cost-effective scintillator-based neutron imaging systems that achieve resolutions of 1μm or less. Such imaging system would dramatically extend the application of neutron imaging to micro-scale structures such as microchannel heat exchangers, fuel cell components, biological microscopy for pharmacology, drug delivery research, fuel injector sprayers for efficient diesel engine technology, and biofuels research. For such purposes a coded source imaging system is under development. The leading challenge is to reduce artifacts in reconstructed coded source images. Neutrons interact with several optical components before they interact with the imaged object, and finally hit the detector. These interactions introduce patterns to the neutron source, which translate to artifacts in the reconstructed image. As the system magnification increases, the effects of the non-uniformity of the source are more evident and serious. In this paper, we describe a coded source neutron imaging workflow. In particular, the algorithms to correct the non-uniformity of the neutron source are discussed. We assess non-parametrized (e.g., arithmetic correction) and parametrized (e.g., maximum likelihood estimator) correction methods. Finally, all methods are complemented with simulated and experimental examples.

8296-24, Session 4

Image enhancement and quality measures for dietary assessment using mobile devices

C. Xu, F. Zhu, N. Khanna, C. J. Boushey, E. J. Delp III, Purdue Univ. (United States)

Measuring accurate dietary intake is considered to be an open research problem in the nutrition and health fields. Dietary assessment methods perceived as less burdensome and less time- consuming may improve compliance. The use of a mobile telephone's built-in digital camera has been shown to provide unique mechanisms for reducing user burden and improving the accuracy and reliability of dietary assessment. We are developing a system, known as the mobile device food record (mdFR), to automatically identify and quantify foods and beverages consumed based on analyzing meal images captured with a handhold mobile device.

To automatically identify and estimate portions of foods in a meal image, it is crucial to obtain good quality images, which fulfill the requirements of different image analysis steps. This paper presents low complexity methods image quality measurement and image enhancement which will improve the accuracy and robustness of existing dietary assessment methods deployed on mobile devices. These include methods for fiducial marker detection and blur assessment on mobile devices. The methods proposed in this paper have been deployed on the iPhone without adding any perceptible delay in the image capture step and thus enhances the user experience.

8296-25, Session 5

Risk estimates for MRI denoising

P. J. Wolfe, Harvard Univ. (United States)

No abstract available.

8296-26, Session 5

Subjective evaluations of example-based, total variation, and joint regularization for image processing

H. S. Anderson, M. R. Gupta, Univ. of Washington (United States); J. Hardeberg, Gjøvik Univ. College (Norway)

We present subjective evaluations of example-based regualrization, total variation regularization, and a proposed joint example-based and total variation regularization for image estimation problems. We focus on the noisy deblurring problem, which generalizes image superresolution and denoising. Controlled subjective experiments show that the proposed joint regularization can yield signicant improvement over only using total variation or example-based regularization, particularly when the example images contain similar structural elements as the test image. We also investigate whether the regularization parameters can be trained by cross-validation, and the dierence in cross-validation judgments made by humans or by fully automatic image quality metrics. Experiments show that of ve image quality metrics tested, the structural similarity index (SSIM) correlates best with human judgement of image quality, and can be probably used to cross-validate regularization parameters. However, there is a signicant quality gap depending on whether the parameters are cross-validated by humans or with the best image quality metric.

8296-27, Session 5

Removal of haze and noise from a single image

E. Matlin, P. Milanfar, Univ. of California, Santa Cruz (United States)

Images of outdoor scenes often contain atmospheric degradation, generally referred to as haze, caused by particles in the atmospheric

medium absorbing and scattering light. For a variety of reasons (e.g. computer vision or sensing algorithms) one may wish to remove the haze from these images. Furthermore, all images contain some corruption from noise due to measurement (sensor) error, which is amplified by the haze removal process if not addressed. The aim of this paper is to present an effective method for removing both haze and noise from only a single digital image. We propose two different methods: denoise the image with a state-of-the-art denoising algorithm as a pre-processing step to haze removal, and simultaneously denoise and dehaze using an iterative, adaptive, kernel-regression based method. Findings show that denoising prior to dehazing works well when the level of noise is known a priori; however, latent errors from either "under"-denoising or "over"-denoising are amplified in the dehazing process. In this situation, the iterative approach can yield superior results.

8296-14, Session 6

Image reconstruction from nonuniformly spaced samples in Fourier domain optical coherence tomography

J. Ke, E. Y. Lam, R. Zhu, The Univ. of Hong Kong (Hong Kong, China)

In reconstructing an object in Fourier Domain OCT (FD-OCT), conventionally FFT is used with system measurements. However, FFT normally requires data that are uniformly sampled in wavenumber, while here measurements are uniformly sampled in wavelength instead. The nonlinear correspondence between wavelength and wavenumber causes non-uniformly spaced sampling in FD-OCT. To deal with this issue, system measurements are resampled before FFT is applied in the conventional method.

In this work, we use inverse imaging for object reconstruction. We first model FD-OCT system with a linear system of equations, where the source power spectrum and the non-uniformly spaced sample position are represented accurately. Then we reconstruct object signal directly from the non-uniformly spaced wavelength measurements. With the inverse imaging method, we also estimate 2D cross-sectional object image directly instead of a set of independent A-line signals. By using Total Variation (TV) as a constraint in the optimization, we reduce the noise in the 2D object estimation. Besides TV, we also use object sparsity in spatial domain as a regularization to further improve inverse imaging reconstruction performance. Experimental results demonstrate the advantages of our method, as we compare it with conventional and NUDFT method for reconstruction.

8296-28, Session 6

Finding saliency in noisy images

C. Kim, P. Milanfar, Univ. of California, Santa Cruz (United States)

Recently, many computational saliency models have been introduced to transform a given input image into a scalar-valued map that represents visual saliency of the corresponding input image. These approaches, however, assume the input image given is clean and fail to provide good results when the input image is degraded. Accordingly, the question is how to compute saliency in a reliable way when a noise-corrupted image is given. In order to deal with this problem, we investigated the sensitivity of the saliency function measuring the self-resemblance and found it to be sensitive to small changes in its arguments. As such, we modify the saliency function with an appropriate correction term to find a more robust estimate in terms of mean squared error (MSE). More specifically, the optimal parameters of this correction will be determined by minimizing an estimate of MSE between saliency values calculated from clean image and the ones estimated from the noisy image. For the MSE estimate here, we shall use Stein's unbiased risk estimate (SURE) that depends on the noisy data alone. The paper will present an estimation theoretic approach to the computation of saliency for noisy images and show the performance of the proposed estimator by comparing it with existing methods.

8296-29, Session 6

Automatic loop closure detection using multiple cameras for 3D indoor localization

N. Corso, J. Kua, J. Chen, A. Zakhor, Univ. of California, Berkeley (United States)

Automated 3D modeling of building interiors is useful in applications such as virtual reality and environment mapping. We have developed a human operated backpack data acquisition system equipped with a variety of sensors such as cameras, laser scanners, and orientation measurement sensors to generate 3D models of building interiors, including uneven surfaces and stairwells. An important intermediate step in any 3D modeling system, including ours, is accurate 6 degrees of freedom localization over time. In this paper, we propose two approaches to improve localization accuracy over existing methods. First, we develop an adaptive localization algorithm which takes advantage of the environment's floor planarity whenever possible. Secondly, we show that by including all the loop closures resulting from two cameras facing away from each other, it is possible to significantly reduce localization error in scenarios where parts of the acquisition path is retraced. We experimentally characterize the performance gains due to both schemes.

8296-30, Session 6

An information theoretic trackability measure

S. T. Acton, A. Aksel, Univ. of Virginia (United States)

There exists no measure to quantify the difficulty of a video tracking problem. Such difficulty depends upon the quality of the video and upon the ability to distinguish the target from the background and from other potential targets. We define a trackability measure in an information theoretic framework. The Shannon-born tools of information theory, and more specifically, rate distortion theory, allow a measure of trackability that seamlessly combines the video-dependent aspects with the target-dependent aspects of tracking difficulty using measure of rate and information content. Specifically, video quality is encapsulated into a term that measures spatial resolution, temporal resolution, signal-to-noise ratio (assuming an additive white Gaussian noise model), and quantization/coding artifacts by way of a Shannon-Hartley analysis. Then, the ability to correctly match a template to a target is evaluated through mutual information. The target-dependent term also incorporates a distance-weighted penalty for nearby clutter and targets that resemble the template. The overall trackability measure is compared to the performance of a recent tracker based on scale space features computed via connected filters. The results show a high Spearman's rank correlation between the trackability measure and actual performance.

8296-31, Session 6

Text replacement on cylindrical surfaces: a semi-automatic approach

H. Ding, R. Bala, Z. Fan, Xerox Corp. (United States); C. A. Bouman, J. P. Allebach, Purdue Univ. (United States)

Image-based customization that incorporates personalized text strings into photorealistic images in a natural and appealing way has been of great interest lately. We describe a semi-automatic approach for replacing text on cylindrical surfaces in images of natural scenes or objects. The user is requested to select a boundary for the existing text and align a pair of edges for the sides of the cylinder. The algorithm erases the existing text, and instantiates a 3-D cylinder forward projection model to render the new text. The parameters of the forward projection model are estimated by optimizing a carefully designed cost function. Experimental results show that the text-replaced images look natural and appealing.

Conference 8296

8296-41, Session 6

An efficient and iterative two-step depth camera self-calibration technique using depth measurements

R. S. Pahwa, D. Babacan, M. N. Do, Univ. of Illinois at Urbana-Champaign (United States)

In this paper, we propose a novel algorithm for calibrating Time of Flight (ToF) depth cameras using known geometry present in the captured images. The

depth images captured by the cameras are first denoised using a Poisson denoising model and then corrected for barrel distortion. We use these corrected and denoised depth images along with the estimated intensity images captured by the depth camera to develop an effective Two step calibration scheme that performs a local optimization per scene and then a global optimization to estimate the depth camera parameters. We demonstrate the high performance and accuracy of the proposed method by comparing it with the standard calibration scheme on synthetic and real world data.

8296-42, Session 6

Registration and integration of multiple depth images using signed distance function

D. Kubacki, H. Q. Bui, D. Babacan, M. N. Do, Univ. of Illinois at Urbana-Champaign (United States)

Depth camera is a new technology that has potential to radically change the way humans record the world and interact with 3D virtual environments. With depth camera, one can have access to depth information up to 30 frames per second, which is much faster than previous 3D scanners. This speed enables new applications, in that objects are no longer required to be static for 3D sensing. There is, however, a trade-off between the speed and the quality of the results. Depth images acquired with current depth cameras are noisy and have low resolution, which poses a real obstacles to incorporating the new 3D information into computer vision techniques. In this paper, we propose a new method to register and integrate multiple depth frames over time onto a global model represented by an implicit moving least square surface.

8296-33, Poster Session

Analysis of practical coverage of uniform motions for approximating real camera shakes

H. Cho, S. Cho, Pohang Univ. of Science and Technology (Korea, Republic of); Y. S. Moon, J. Cho, S. Lee, Samsung Electronics Co., Ltd. (Korea, Republic of); S. Lee, Pohang Univ. of Science and Technology (Korea, Republic of)

Motion blur is usually modeled as the convolution of a latent image with a motion blur kernel, which assumes a translational camera motion has caused uniform blurs in a blurred image. However, real camera shakes often contain non-translational components, such as rotations, resulting in non-uniform blurs. To better utilize the existing deblurring methods in practice, it is necessary to understand how much uniform motions (i.e., translations) can approximate real camera shakes.

In this paper, we analyze the effects of real camera motions on image pixels and present the coverage of uniform motions for approximating complicated camera shakes. We first perform mathematical analysis on the difference of the motion extents at the optical axis and image boundary under camera shakes. We then derive the practical coverage of the uniform blur model when used for real blurred images. The coverage analysis can provide guidelines for the cases that can be effectively handled by existing uniform deblurring methods, and the cases for which more complicated non-uniform blur models are needed.

8296-34, Poster Session

Real-time computational camera system for high-sensitivity imaging by using combined long/short exposure

S. Sato, Y. Okada, T. Azuma, Panasonic Corp. (Japan)

In this work, we propose a real time computational camera system to process 24fps with 62 FPGAs for 4K2K camera using combined long/short exposure. Our proposed system will contribute to future camera systems based on the following two points:

1. High-sensitivity: Our proposed camera system can capture images with four times higher sensitivity than conventional imager with no motion blur.

2. Small-size and high resolution: The new image reconstruction method achieved the high-sensitivity imaging even for a small imager (the pixel pitch size is 1.43um and the resolution is 3840 x 2160 pixels).

The simulation results showed that the Peak-Signal-to-Noise-Ratio (PSNR) of the image reconstruction algorithm is over 35 dB. Experimental results showed that our proposed camera system is effective even in conditions of low illumination.

8296-35, Poster Session

Color correction with edge preserving and minimal SNR decrease using multi-layer decomposition

B. K. Park, W. Choe, J. Lim, S. Lee, C. Kim, Samsung Electronics Co., Ltd. (Korea, Republic of)

This paper describes the method related to correcting color distortion in color imaging. Acquiring color image from CMOS or CCD digital sensors can suffer from color distortion, which means that the image from sensors is different from the original image in the color space. The main reasons are the crosstalks between adjacent pixels, the color pigment characteristic's mismatch with human perception and infra-red (IR) influx to visible channel (RGB) due to IR cutoff filter imperfection. To correct this distortion, existing methods exploit multiplying gain coefficients in each color channel and this multiplication can cause noise boost and loss of detail information. This paper proposes the novel method which can not only preserve color distortion correction ability, but also suppress noise boost and loss of detail information in the color correction process of IR corrupted pixels. In the case of non-IR corruption pixels, the use of image before color correction instead of IR image makes this kind of method available. Specifically the color and low frequency information in luminance channel is extracted from the color corrected image. And high frequency information is from the IR image or the image before color correction. The method extracting the low and high frequency information use multi-layer decomposition skill with edge preserving filters.

8296-36, Poster Session

Bayesian image superresolution for hyperspectral image reconstruction

Y. Murayama, A. Ide-Ektessabi, Kyoto Univ. (Japan)

This study aims to achieve a more practical spectral imaging method, and proposes a method to estimate high spatial resolution spectral image from low spatial resolution multispectral image. Theories of spectral reflectance estimation, such as Wiener estimation, have reduced the time and trouble in spectral imaging. They increase wavelength resolution of multispectral image and reconstruct hyperspectral image. However there is demand for a more practical spectral imaging. Proposed method can treat not only wavelength region but spatial region of hyperspectral image and makes it possible to increase both wavelength resolution and spatial resolution. It is implemented by combining Wiener estimation and Bayesian superresolution in the framework of Bayesian statics. In proposed

method, Maximization of the marginal likelihood function is carried out to estimate high resolution hyperspectral image and registration parameter of each band of acquired multispectral image. Proposed method allows acquiring images with lower resolution and then increasing it to the desired resolution. This could greatly reduce the image acquisition time making it more practical. This study shows the method and illustrates its effectiveness in an experimental approach applying acquired multispectral image of a Japanese traditional painting. Results show obvious effect of increasing the spatial-resolution in estimating hyperspectral image.

8296-37, Poster Session

ToF depth image deblurring using 3D blur shape models and motion blur saliency map (MBSM)

S. Lee, K. Shim, J. D. K. Kim, C. Kim, Samsung Advanced Institute of Technology (Korea, Republic of)

Time-of-flight depth camera gives direct 3D information enabling faster and easier 3D scene capturing and reconstruction. The depth camera, however, also suffers from motion blur caused by any movement of camera or subjects. Because the 3D depth information is used to reconstruct 3D geometry of the scene, blurred region in a depth image comes out seriously in the final reconstructed 3D model causing color and structural confusions.

Our observation is that the motion blur observed in depth image shows quite different aspect from that of color image. In this manuscript, we analyze the motion blur mechanism of ToF depth camera in mathematical manner and derive a set of motion blur 3D shape models. Base on this theoretical inspection, we propose a unique depth image deblurring algorithm.

8296-38, Poster Session

Computational imaging of defects in commercial substrates for electronic and photonic devices

R. Kashiwagi, M. Fukuzawa, M. Yamada, Kyoto Institute of Technology (Japan)

Computational defect imaging has been performed in commercial substrates for electronic and photonic devices by combining the transmission profile acquired with an imaging type of linear polariscope and the computational algorithm to extract strain-induced birefringence. The imaging results exhibited defect structures specific to substrate materials such as residual strain distribution in GaP substrates, dislocations and subgrain boundaries in optical-grade LiNbO3 substrates, and micropipes in SiC substrates, which were not detected by conventional 'visual inspection' based on simple optical refraction or transmission because of poor sensitivity. The typical imaging time was less than 30 sec for 3-inch diameter substrate with the special resolution of 200 μm, while that by scanning polariscope was 2 hours to get the same special resolution. Since high sensitivity, short imaging time, and wide coverage of substrate materials are practical advantages over the laboratory-scale apparatus such as X-ray topography and electron microscope, our proposed technique is useful for nondestructive inspection of various commercial substrates in production of electronic and photonic devices.

8296-39, Poster Session

Nondestructive three-dimensional measurement of gas temperature distribution by phase tomography

S. Tomioka, S. Nishiyama, Hokkaido Univ. (Japan)

This study presents a nondestructive three-dimensional (3-D) measurement of gas temperature distribution around a heater by

a coupling method using computed tomography (CT) and an interferometer. The gas temperature distribution is related to a distribution of refractive index. Therefore, once a series of two-dimensional (2-D) image of phase modulation that is an integral of refractive index along an optical path is obtained, a 3-D distribution of the gas temperature can be ideally determined in the same way as the widely used CT to determine a distribution of attenuation factor.

The interferometer detects a 2-D fringe pattern, and the fringe pattern is a converted phase modulation map using phase unwrapping algorithm. However, when the fringe pattern is noisy, conventional phase unwrapping algorithm induces many errors. Evaluations of the accuracy by several phase unwrapping algorithms are presented.

In contrast to the conventional CT imaging, the 2-D phase images taken from all the direction of probe beam is not obtained in the system to measure the gas temperature, because of the setting restriction of the interferometer, and also because the object to be measured cannot rotate. To reconstruct the temperature distribution with such incomplete data set, this study presents two approaches: a back-projection method and an algebraic method with Tikhonov's regularization.

8296-40, Poster Session

Closed-form inverses for the mixed pixel/multpath interference problem in AMCW lidar

J. P. Godbaz, M. J. Cree, A. Dorrington, The Univ. of Waikato (New Zealand)

The mixed pixel/multipath interference problem is a significant systematic error in Amplitude Modulated Continuous Wave lidar. Whereas many off-the-shelf commercial ranging systems such as the Swissranger SR4000 advertise centimeter level ranging precision, the accuracy of the ranging measurements is commonly one or more orders of magnitude worse. This poor accuracy is caused by individual pixels integrating light from more than one backscattering source; often due to scattering within the camera optics. In this paper we present two novel approaches to the problem: one using attenuation ratios in the exactly-determined case, modelling each component return as a scaled Dirac delta over range; and one in the over-determined case, modelling volumetric scattering using a Cauchy distribution. Both these new inversion methods have comparatively straightforward closed-form inverses that allow real-time implementation; unlike previous numerical approaches. Whereas many previous approaches have only partially ameliorated the problem, these new approaches determine both the amplitude and phase of the individual backscattering signals. In this paper we evaluate the performance of both methods on real and simulated data and examine the influence of relative phase, the Cauchy range-spread coefficient and SNR on algorithm performance. We also compare the methods to our previously published exactly-determined numerical approach.

Conference 8297: Document Recognition and Retrieval XIX

Wednesday-Thursday 25-26 January 2012

Part of Proceedings of SPIE Vol. 8297 Document Recognition and Retrieval XIX

8297-28, Poster Session

Bleed-through removal in degraded documents

R. F. Rowley-Brooke, A. Kokaram, Trinity College Dublin (Ireland)

A restoration method for bleed-through degraded document images using a Bayesian approach for bleed-through reduction, using a linear-based model, is presented in this paper. To estimate the 6 unknown parameters from the model, a variation of iterated conditional modes (ICM) optimisation is used, whereby each parameter is estimated in turn via the mode of their conditional probabilities. However, in the case of the clean image estimates, samples are drawn instead, with mean and variance obtained from the likelihood and prior models. Furthermore every ten iterations the mean and variance are obtained from a model that ignores the prior probabilities and two of the unknown variables. The proposed method is tested on various samples of high resolution (600dpi) scanned manuscript images, with different degrees of degradation. The performance of the method is visually compared with that of a recent user-assisted restoration method, and, though the user-assistance leads to better classification in regions where bleed-through is darker in intensity than foreground text, the proposed method is shown to produce better overall results.

8297-29, Poster Session

Clustering document fragments using background color and texture information

S. Chanda, K. Franke, Gjøvik Univ. College (Norway); U. Pal, Indian Statistical Institute (India)

Forensic analysis of questioned documents sometimes can be extensively data intensive. A forensic expert might need to analyze a heap of document fragments and in such cases to ensure reliability he/she should focus only on relevant evidences hidden in those document fragments. Relevant document retrieval needs finding of similar document fragments. One notion of obtaining such similar documents could be by using document fragment's physical characteristics like color, texture, etc. In this article we propose an automatic scheme to retrieve similar document fragments based on visual appearance of document paper and texture. Multispectral color characteristics using biologically inspired color differentiation techniques are implemented here. This is done by projecting document color characteristics to Lab color space. Gabor filter-based texture analysis is used to identify document texture. It is desired that document fragments from same source will have similar color and texture. For clustering similar document fragments of our test dataset we use a Self Organizing Map (SOM) of dimension 5 x 5, where the document color and texture information are being used as features. We obtained an encouraging accuracy of 97.17% from 1063 test images.

8297-30, Poster Session

Lecture video segmentation and indexing

D. Ma, G. Agam, Illinois Institute of Technology (United States)

Video structuring and indexing are two crucial processes for multi-media document understanding and informa- tion retrieval. This paper presents a novel approach in automatic structuring and indexing lecture videos for an educational video system. By structuring and indexing video content, we can support both topic indexing and semantic querying of multimedia documents. In this paper, our goal is to extract indices of topics and link them with their associated video and audio segments. Two main techniques used in our proposed approach are video image analysis and video text analysis. Using this approach, we obtain accuracy of over 90.0% on our test collection.

8297-31, Poster Session

Unsupervised categorization method of graphemes on handwritten manuscripts: application to style recognition

H. Daher, D. Gaceb, V. Eglin, S. Bres, N. Vincent, Institut National des Sciences Appliquées de Lyon (France)

We present in this paper a feature selection and weighting method for medieval handwriting images that relies on codebooks of shapes of small strokes of characters (graphemes that are issued from the decomposition of manuscripts). These codebooks are important to simplify the automation of the analysis, the manuscripts transcription and the recognition of styles or writers. Our approach provides a precise features weighting by genetic algorithms and a high-performance methodology for the categorization of the shapes of graphemes by using graph coloring into codebooks which are applied in turn on CBIR in a mixed handwriting database containing different pages from different writers, periods of the history and quality. We show how the coupling of these two mechanisms "features weighting - graphemes classification" can offer a better separation of the forms to be categorized by exploiting their grapho-morphological, their density and their significant orientations particularities.

8297-32, Poster Session

Retrieving handwriting by combining word spotting and manifold ranking

S. Peña Saldarriaga, Synchromedia (Canada); E. Morin, C. Viard-Gaudin, Univ. de Nantes (France)

Online handwritten data, produced with Tablet PCs or digital pens, consists in a sequence of points (x,y). As the amount of data available in this form increases, algorithms for retrieval of online data are needed. Word spotting is a common approach used for the retrieval of handwriting. However, from an information retrieval (IR) perspective, word spotting is a primitive keyword based matching and retrieval strategy. We propose a framework for handwriting retrieval where an arbitrary word spotting method is used, and then a manifold ranking algorithm is applied on the initial retrieval scores. Experimental results on a database of more than 2,000 handwritten newswires show that our method can improve the performances of a state-of-the-art word spotting system by more than 10%.

8297-33, Poster Session

The A2iA French handwriting recognition system at the Rimes-ICDAR2011 competition

F. Menasri, J. Louradour, A. Bianne-Bernard, C. Kermorvant, A2iA SA (France)

This paper describes the system for the recognition of French handwriting submitted by A2iA to the competition organized at ICDAR2011 using the Rimes database.

This system is composed of several recognizers based on three different recognition technologies, combined using a novel combination methods.

A framework multi-word recognition based on weighted finite state

transducers is presented, using an explicit word segmentation, a combination of isolated word recognizers and a language model.

The system was tested both for isolated word recognition and for multi-word line recognition and submitted to the RIMES-ICDAR2011 competition.

This system outperformed all previously proposed systems on these tasks.

8297-34, Poster Session

Using connected component decomposition to detect straight line segments in documents

X. Feng, A. Youssef, The George Washington Univ. (United States)

Straight line segment detection in digital documents has been studied extensively for the past few decades. One of the challenges is to detect line segments without priori information about document images and render good results without much parameter calibration. In this paper, we introduce a novel algorithm that is simple but effective in detecting straight line segments in scanned documents. Our Connected Component Decomposition (CCD) approach first decomposes the connected components based on the gradient direction of the edge contours, and then uses Chebyshev's inequality to statistically distinguish lines from characters, followed by a simple post processing step to examine straightness of remain segments. This CCD approach is simple to follow and fast in its implementation, and its high accuracy and usability are demonstrated empirically on a practical data set with large varieties.

8297-35, Poster Session

A synthetic document image dataset for developing and evaluating historical document processing methods

D. D. Walker IV, W. B. Lund, E. K. Ringger, Brigham Young Univ. (United States)

Document images accompanied by OCR output text and ground truth transcriptions are useful for developing and evaluating document recognition and processing methods for historical documents. Additionally, research into improving the performance of such methods often requires further annotation (e.g., topical document labels). However, transcribing and labeling historical documents is expensive. As a result, existing real-world datasets with such resources are rare and relatively small. We introduce synthetic document image datasets of varying levels of noise that have been created from standard text corpora using an existing document degradation model applied in a novel way. Included in the datasets is OCR output from the commercial ABBYY FineReader and the open-source Tesseract OCR engines. These synthetic datasets are designed to exhibit the characteristics of an example real-world document image dataset, the Eisenhower Communiques. The new datasets also benefit from additional metadata that exist due to the nature of their collection and prior labeling efforts. We demonstrate the usefulness of the synthetic datasets by training an OCR correction model on the synthetic data and then applying the model to reduce word error rates on historical documents. The synthetic datasets will be made available for use by other researchers.

8297-01, Session 1

Large scale visual semantic extraction

S. Bengio, Google Inc. (United States)

Image annotation is the task of providing textual semantic to new images, by ranking a large set of possible annotations according to how they correspond to a given image. In the large scale setting, there could be millions of images to process and hundreds of thousands of potential distinct annotations. In order to achieve such a task we propose to build a so-called "embedding space", into which both images and annotations can be automatically projected. In such a space, one can then find the nearest annotations to a given image, or annotations similar to a given annotation. One can even build a visio-semantic tree from these annotations, that corresponds to how concepts (annotations) are similar to each other with respect to their visual characteristics. Such a tree will be different from semantic-only trees, such as WordNet, which do not take into account the visual appearance of concepts. We propose a new learning-to-rank approach that can scale to such dataset and show some annotation results.

8297-02, Session 2

Graphical image classification combining an evolutionary algorithm and binary particle swarm optimization

B. Cheng, R. Wang, Missouri Univ. of Science and Technology (United States); S. K. Antani, National Library of Medicine (United States); R. J. Stanley, Missouri Univ. of Science and Technology (United States); G. R. Thoma, National Library of Medicine (United States)

Biomedical journal articles contain a variety of image types that can be broadly classified into two categories: regular images, and graphical images. Graphical images can be further classified into four classes: diagrams, statistical figures, flow charts, and tables. Automatic figure type identification is an important step toward improved multimodal (text + image) information retrieval and clinical decision support applications. This paper describes a feature-based learning approach to automatically identify these four graphical figure types. We apply Evolutionary Algorithm (EA), Binary Particle Swarm Optimization (BPSO) and a hybrid of EA and BPSO (EABPSO) methods to an optimal subset of extracted image features that are then classified using a Support Vector Machine (SVM) classifier. Evaluation performed on 1038 figure images extracted from ten BioMedCentral® journals with the features selected by EABPSO yielded classification accuracy as high as 87.5%.

8297-03, Session 2

Combining SVM classifiers to identify investigator name zones in biomedical articles

J. Kim, D. X. Le, G. R. Thoma, National Library of Medicine (United States)

This paper describes an automated system to label zones containing Investigator Names (IN) in biomedical articles, a key item in a MEDLINE® citation. The correct identification of these zones is necessary for the subsequent extraction of IN from these zones. A hierarchical classification model is proposed using two Support Vector Machine (SVM) classifiers. The first classifier is used to identify an IN zone with highest confidence, and the other classifier identifies the remaining IN zones. Eight sets of word lists are collected to train and test the classifiers, each set containing collections of words ranging from 100 to 1,200. Experiments based on a test set of 105 journal articles show a Precision of 0.88, 0.97 Recall, 0.92 F-Measure, and 0.99 Accuracy.

8297-04, Session 2

Comprehensive color segmentation system for noisy digitized documents to enhance text extraction

A. Ouji, Y. Leydier, F. LeBourgeois, Institut National des Sciences Appliquées de Lyon (France)

This paper presents a novel, general purpose and multi-applications color segmentation system providing optimal chromatic and achromatic layers and filtering the hue and illumination distortions, with minimal information

loss. A text extraction method based on the resulting segmentation is proposed to illustrate the usefulness of the method. The system is validated through the evaluation of a well-known commercial OCR line segmentation performances on the processed images.

8297-05, Session 2

Ensemble methods with simple features for document zone classification

T. Obafemi-Ajayi, G. Agam, B. Xie, Illinois Institute of Technology (United States)

Document layout analysis is of fundamental importance for document image understanding and information retrieval. It requires the identification of blocks extracted from a document image via features extraction and block classification.

In this paper, we focus on the classification of the extracted blocks into five classes: text (machine printed), handwriting, graphics,

images, and noise. We propose a new set of features for efficient classifications of these blocks. We present a comparative evaluation of three ensemble based classification algorithms (boosting, bagging, and combined model trees) in addition to other known learning algorithms.

Experimental results are demonstrated for a set of 36503 zones extracted from 416 document images which were randomly selected from the tobacco legacy document collection.

The results obtained verify the robustness and effectiveness of the proposed set of features in comparison to the commonly used Ocropus recognition features. When used in conjunction with the Ocropus feature set, we further improve the performance of the block classification system to obtain a classification accuracy of 99.21%.

8297-06, Session 3

A robust omnifont open-vocabulary arabic OCR system using pseudo-2D-HMM

A. M. Rashwan, Cairo Univ. (Egypt) and RDI (Egypt); M. A. Rashwan, S. Abdou, A. Abdel-Hameed, Cairo Univ. (Egypt)

Recognizing old documents is highly desirable since the demand for quickly searching millions of archived documents has recently increased. Using Hidden Markov Models (HMMs) have been proven to be a good solution to tackle the main problems of recognizing typewritten Arabic characters. These attempts however achieved a remarkable success for omnifont OCR under very favorable conditions, they didn't achieve the same performance in practical conditions, i.e. noisy documents. In this paper we present an omnifont, large-vocabulary Arabic OCR system using Pseudo Two Dimensional Hidden Markov Model (P2DHMM), which is a generalization of the HMM. P2DHMM offers more efficient way to model the Arabic characters, such model offer both minimal dependency on the font size/style (omnifont), and high level of robustness against noise. The evaluation results of this system are very promising compared to HMM system and best OCRs available in the market (Sakhr and NovoDynamics). The recognition accuracy of the P2DHMM classifier is measured against classic HMM classifier, the average word accuracy rates for P2DHMMand HMM classifiers are 79%and 66% respectively. The overall system accuracy is measured against Sakhr and NovoDynamics OCR systems, the average word accuracy rates for P2DHMM, NovoDynamics, and Sakhr are 74%, 71%, and 61% respectively.

8297-07, Session 3

Variable length and context-dependent HMM letter form models for Arabic handwritten word recognition

A. Bianne-Bernard, A2iA SA (France) and Telecom ParisTech (France); F. Menasri, A2iA SA (France); L. Likforman-Sulem, Telecom ParisTech (France); C. Mokbel, Univ. of Balamand (Lebanon); C. Kermorvant, A2iA SA (France)

We present in this paper an HMM-based recognizer for the recognition of unconstrained Arabic handwritten words. The recognizer is a context-dependent HMM which considers variable topology and contextual information for a better modeling of writing units. We propose an algorithm to adapt the topology of each HMM to the character to be modeled. For modeling the contextual units, a state-tying process based on decision tree clustering is introduced which significantly reduces the number of parameters. Decision trees are built according to a set of expert-based questions on how characters are written. Questions are divided into global questions yielding larger clusters and precise questions yielding smaller ones. We apply this modeling to the recognition of Arabic handwritten words. Experiments conducted on the OpenHaRT2010 database show that variable length topology and contextual information significantly improves the recognition rate.

8297-08, Session 3

Post-processing for offline Chinese handwritten character string recognition

Y. Wang, X. Ding, C. Liu, Tsinghua Univ. (China)

Offline Chinese handwritten character string recognition is one of the most important research fields in pattern recognition. Due to the free writing style, large variability in character shapes and different geometric characteristics, Chinese handwritten character string recognition is a challenging problem to deal with. However, among the current methods over-segmentation and merging method which integrates geometric information, character recognition information and contextual information, shows a promising result. It is found experimentally that a large part of errors are segmentation error and mainly occur around non-Chinese characters. In a Chinese character string, there are not only wide characters namely Chinese characters, but also narrow characters like digits and letters of the alphabet. The segmentation error is mainly caused by uniform geometric model imposed on all segmented candidate characters. To solve this problem, post processing is employed to improve recognition accuracy of narrow characters. On one hand, multi-geometric models are established for wide characters and narrow characters respectively. Under multi-geometric models narrow characters are not prone to be merged. On the other hand, top rank recognition results of candidate paths are integrated to boost final recognition of narrow characters. The post processing method is investigated on two datasets, in total 1405 handwritten address strings. The wide character recognition accuracy has been improved lightly and narrow character recognition accuracy has been increased up by 10.41% and 10.03% respectively. It indicates that the post processing method is effective.

8297-09, Session 3

Complexity reduction with recognition rate maintained for online handwritten Japanese text recognition

J. Gao, Tokyo Univ. Agriculture and Technology (Japan)

The paper presents complexity reduction of an on-line handwritten Japanese text recognition system by selecting an optimal off-line recognizer in combination with an on-line recognizer, geometric context evaluation and linguistic context evaluation. The result is that a surprisingly small off-line recognizer, which alone is weak, produces the nearly best recognition rate in combination with other evaluation factors in remarkably small space and time complexity. Generally speaking, lower dimensions with less principle components produce a smaller set of prototypes, which reduce memory-cost and time-cost. It degrades the recognition rate, however, so that we need to compromise them. In an evaluation function with the above-mentioned multiple factors combined, the configuration of only 50 dimensions with as little as 5 principle components for the off-line recognizer keeps almost the best accuracy 98.85% (the best accuracy 98.91%) for text recognition while it suppresses the total memory-cost from 99.4 MB down to 32 MB

and the average time-cost of character recognition for text recognition from 0.1621 ms to 0.1191 ms compared with the traditional off-line recognizer with 160 dimensions and 50 principle components.

8297-10, Session 3

Improving isolated and in-context classication of handwritten characters

V. Mazalov, S. M. Watt, The Univ. of Western Ontario (Canada)

Earlier work has shown how to recognize handwritten characters by representing coordinate functions or integral invariants as truncated orthogonal series. The series basis functions are orthogonal polynomials dened by a Legendre-Sobolev inner product. It has been shown that the free parameter in the inner product, the "jet scale", has an impact on recognition both using coordinate functions and integral invariants. This paper develops methods of improving series-based recognition. For isolated classication, the rst consideration is to identify optimal values for the jet scale in dierent settings. For coordinate functions, we nd the optimum to be in a small interval with the precise value not strongly correlated to the geometric complexity of the character. For integral invariants, used in orientation-independent recognition, we nd the optimal values of the jet scale are dierent for each invariant and conrm that each of these optima is independent of the orientation. Furthermore, we examine the optimal degree for the truncated series. For in-context classication, we develop a rotation-invariant algorithm that takes advantage of sequences of samples that are subject to similar distortion. The algorithm yields signicant improvement over orientation-independent isolated recognition and can be extended to shear and, more generally, ane transformations.

8297-11, Session 4

Using specic evaluation for comparing and combining competing algorithms: applying it to table column detection

A. C. C. Silva, Univ. do Porto (Portugal)

It is a commonly used evaluation strategy to run competing algorithms on a test dataset and state which performs better in average on the whole set. We call this textit{generic evaluation}. Although it is important, we believe this type of evaluation is incomplete.

In this paper, we shall proceed to propose a methodology for algorithm comparison, which we call textit{specific evaluation}. This approach does not limit itself to taking the average performance of two or more algorithms on a given unseen dataset, but instead attempts to identify subsets of the data where one algorithm is better than the other. We shall be applying specific evaluation to an experiment that aims at grouping pre-obtained table cells into columns; we demonstrate how it identifies a subset of data for which the on-average least good but faster algorithm is equivalent or better, and then manages to create a policy for combining the two complementary table column delimitation algorithms, thus exploiting the strengths of each.

8297-12, Session 4

Identification of embedded mathematical formulas in PDF documents using SVM

X. Lin, L. Gao, Z. Tang, Peking Univ. (China); X. Hu, BeiHang Univ. (China); X. Lin, Vobile, Inc. (United States)

With the tremendous popularity of PDF format, recognizing mathematical formulas in PDF documents becomes a new and important problem in document analysis field. In this paper, we present a method of embedded mathematical formula identification in PDF documents, based on Support Vector Machine (SVM). The method first segments text lines into words, and then classifies each word into two classes, namely formula or ordinary text. Various features of embedded formulas, including geometric layout, character and context content, are utilized to build a robust and adaptable SVM classifier. Embedded formulas are then extracted through merging the words labeled as formulas. Experimental results show good performance of the proposed method. Furthermore, the method has been successfully incorporated into a commercial software package for large-scale e-Book production.

8297-13, Session 4

Chemical structure recognition: a rule-based approach

N. M. Sadawi, A. P. Sexton, V. Sorge, The Univ. of Birmingham (United Kingdom)

In chemical literature much information is given in the form of diagrams depicting molecules. In order to access this information diagrams have to be recognised and translated into a processable format. We present an approach that models the principal recognition steps for molecule diagrams in a strictly rule based system, providing rules to identify the main components - atoms and bonds - as well as to resolve possible ambiguities. The result of the process is a translation into a graph representation that can be used for further processing. We show the effectiveness of our approach by describing its embedding into a full recognition system and present an experimental evaluation that demonstrates how our current implementation outperforms the leading open source system currently available.

8297-14, Session 4

Quantify spatial relations to discover handwritten graphical symbols

J. Li, H. Mouchère, C. Viard-Gaudin, Univ. de Nantes (France)

To model a handwritten graphical language, spatial relations describe how the strokes are positioned in the 2-dimensional space. Most of existing handwriting recognition systems make use of some predefined spatial relations. However, considering a complex graphical language, it is hard to express manually all the spatial relations. Another possibility would be to use a clustering technique to discover the spatial relations. In this paper, we discuss how to create a relational graph between strokes (nodes) labeled with graphemes in a graphical language. Then we vectorize spatial relations (edges) for clustering and quantization. As the targeted application, we extract the repetitive sub-graphs (graphical symbols) composed of graphemes and learned spatial relations. On two handwriting databases, a simple mathematical expression database and a complex flowchart database, the unsupervised spatial relations outperform the predefined spatial relations. In addition, we visualize the frequent patterns on two text-lines containing Chinese characters.

8297-15, Session 5

Language modeling for information retrieval

C. Manning, Stanford Univ. (United States)

No abstract available

8297-16, Session 6

Automatic indexing of scanned documents: a layout-based approach

D. Esser, D. Schuster, K. Muthmann, A. Schill, Technische Univ. Dresden (Germany)

Archiving official written documents such as invoices, reminders and account statements in business and private area gets more and more important. Creating appropriate index entries for document archives like sender's name, creation date or document number is a tedious

Conference 8297

manual work. We present a novel approach to handle automatic indexing of documents based on generic positional extraction of index terms. For this purpose we apply the knowledge of document templates stored in a common full text search index to find index positions that were successfully extracted in the past and use them to gain relevant information out of new documents. Therefore, we implemented a template detection and information extraction algorithm and combined both to a workflow for robust and fast information extraction using the layout of business documents. On a real world set of business correspondences, our layout-based method achieves extraction rates up to 95%, depending on the concrete index type. Moreover, even if the structure of documents changes over time, our approach guarantees high rates due to a fast learning user feedback mechanism.

8297-17, Session 6

Layout-based substitution tree indexing and retrieval for mathematical expressions

T. Schellenberg, R. Zanibbi, B. Yuan, Rochester Institute of Technology (United States)

We present a new method for indexing and retrieving mathematical expressions in a layout-based representation (LaTeX) using substitution trees. Substitution trees allow us to index expressions efficiently and group them based on similarity. Searching through the index returns identical expressions, expressions with similar symbols and symbol layout, and expressions with shared sub-expressions. In an experiment we test the search results from our novel system and show that it is comparable to an existing Lucene system.

8297-18, Session 7

Efficient cost-sensitive human-machine collaboration for off-line signature verification

J. Coetzer, J. Swanepoel, Stellenbosch Univ. (South Africa); R. Sabourin, Ecole de Technologie Supérieure (Canada)

We propose a novel strategy for the optimal combination of human and machine decisions in a cost-sensitive environment. The proposed algorithm should be especially beneficial to financial institutions where off-line signatures, each associated with a specific transaction value, require authentication. When presented with a

collection of genuine and fraudulent training signatures, produced by so-called guinea pig writers, the proficiency of a workforce of human employees and a score-generating machine can be estimated and represented in receiver operating characteristic (ROC) space. Using a set of Boolean fusion functions, the majority vote decision of the human workforce is combined with each threshold-specific machine-generated decision. The performance of the candidate ensembles is estimated and represented in ROC space, whereafter only the optimal ensembles and associated decision trees are retained. When presented with a questioned signature linked to an arbitrary writer, the system first uses the ROC-based cost gradient associated with the transaction value to select the ensemble that minimises the expected cost, and then uses the corresponding decision tree to authenticate the signature in question. We show that, when utilising the entire human workforce, the incorporation of a machine streamlines the authentication process and decreases the expected cost for all operating conditions.

8297-19, Session 7

Questioned document workflow for handwriting with automated tools

S. N. Srihari, K. Das, H. Srinivasan, Univ. at Buffalo (United States)

During the last few years many document recognition methods have been developed to determine whether a handwriting specimen can

be attributed to a known writer. However, in practice, the work-flow of the document examiner continues to be manual-intensive. Before a systematic or computational, approach can be developed, an articulation of the steps involved in handwriting comparison is needed. We describe the work flow of handwritten questioned document examination, as described in a standards manual, and the steps where existing automation tools can be used. A well-known ransom note case is considered as an example, where one encounters testing for multiple writers of the same document, determining whether the writing is disguised, known writing is formal while questioned writing is informal, etc. The findings for the particular ransom note case using the tools are given. Also observations are made for developing a more fully automated approach to handwriting examination.

8297-20, Session 7

Iterative analysis of document collections enables efficient human-initiated interaction

J. Chazalon, B. Coüasnon, Institut National des Sciences Appliquées de Rennes (France)

Document analysis and recognition systems often fail to produce results with a sufficient quality level when processing old and damaged documents sets, and require manual corrections to improve results. This paper presents how, using the iterative analysis of document pages we recently proposed, we can implement a spontaneous interaction model, suitable for mass document processing. It enables human operators to detect and correct errors made by the automatic system, and reintegrates the corrections they made into subsequent analysis steps of the iterative analysis process. Then, a page analyzer can reprocess erroneous parts and those which depend on them, avoiding the necessity to manually fix during post-processing all the consequences of errors made by the automated system. After presenting the global system architecture and a prototype implementation of our proposal, we show that document model can be simply enriched to enable the spontaneous interaction model we propose, and we use it in a practical example to correct under-segmentation issues during the localization of numbers in documents from the 18th century. Evaluations we conducted on the example case show, on 50 pages containing 1637 numbers to localize, that the interaction model we propose can reduce human workload while improving localization quality (+7.5% zones well detected) when compared to a manual post-processing.

8297-21, Session 7

VeriClick: an efficient tool for table format verification

G. Nagy, M. Tamhankar, Rensselaer Polytechnic Institute (United States)

The essential layout attributes of a visual table can be defined by the location of four critical grid cells. Although these critical cells can often be located by automated analysis, some means of human interaction is necessary for correcting residual errors. VeriClick is a macro-enabled spreadsheet interface that provides ground-truthing, confirmation, correction, and verification functions for CSV tables. All user actions are logged. Experimental results of seven subjects on one hundred tables suggest that VeriClick can provide a ten to twenty-fold speedup over performing the same functions with standard spreadsheet editing commands.

8297-22, Session 7

Asymptotic cost in document conversion

D. Blostein, Queen's Univ. (Canada); G. Nagy, Rensselaer Polytechnic Institute (United States)

In spite of a hundredfold decrease in the cost of relevant technologies, the role of document image processing systems is gradually declining

due to the transition to an on-line world. Nevertheless, in some high-volume applications, document image processing software still saves millions of dollars by accelerating workflow, and similarly large savings could be realized by more effective automation for the multitude of low-volume personal document conversions. While potential cost savings, based on estimates of costs and values, are a driving force for new developments, quantifying such savings is difficult. The most important trend is that the cost of computing resources for DIA is becoming insignificant compared to the associated labor costs. An econometric treatment of document processing complements traditional performance evaluation, which focuses on assessing the correctness of the results produced by document conversion software. Researchers should look beyond the error rate for advancing both production and personal document conversion.

8297-23, Session 8

Style comparisons in calligraphy

X. Zhang, Shanghai Maritime Institute (China); G. Nagy, Rensselaer Polytechnic Institute (United States)

Calligraphic style is considered, for this research, to be the visual attributes of images of calligraphic characters sampled randomly from a "work" created by a single artist. It is independent of page layout or textual content. An experimental design is developed to investigate to what extent the source of a single, or of a few pairs, of character images can be assigned to the either same work or to two different works. The experiments are conducted on the 13,571 segmented and labeled 600-dpi character images of the CADAL database. The classifier is not trained on the works tested, only on other works. Even when only a few samples of same-class pairs are available, the difference-vector of a few simple features extracted from each image of a pair yields over 80% classification accuracy for a same-work vs. different-work dichotomy. When many pairs of different classes are available for the each pair, the accuracy, using the same features, is almost the same. These style-verification experiments are part of research on the larger goal of style identification and forgery detection.

8297-24, Session 8

An Oracle-based co-training framework for writer identification in offline handwriting

U. Porwal, Univ. at Buffalo (United States); S. Rajan, Fujitsu Labs. of America, Inc. (United States); V. Govindaraju, Univ. at Buffalo (United States)

State of the art techniques for writer identification has been centered primarily on performance enhancement of the system. Machine learning algorithms have been extensively used to improve accuracy of the system assuming availability of the sufficient amount of data for the training purpose. Little attention has been paid to the prospect of harnessing the information tapped in the large amount of un-annotated data. This paper focuses on Co-training based framework that can be used for iterative labeling of the unlabeled data set exploiting the independence between the multiple views(features) of the data. This paradigm relaxes the assumption of sufficiency of the data available and tries to generate labeled data from unlabeled data set along with improving the accuracy of the system. However, performance of Co-training based framework is dependent on the effectiveness of the selection algorithm used for the selection of data points to be added in the labeled set. We propose an Oracle based approach for data selection that learns the patterns in the score distribution of classes for labeled data points and then predicts the labels(writers) of the unlabeled data point. This method for selection statistically learns the class distribution and predicts the most probable class unlike traditional selection algorithms which were based on heuristic approaches. We conducted experiments on publicly available IAM dataset and illustrate the efficacy of proposed approach.

8297-25, Session 8

Handwritten document age classification based on handwriting styles

C. Ramaiah, G. Kumar, V. Govindaraju, Univ. at Buffalo (United States)

Handwriting styles have shown an observable variance across centuries. In this paper, we approach the novel problem of estimating the age of historical handwritten documents using handwriting styles. This system will have far reaching applications in handwritten document processing engines where specialized processing techniques can be applied once the age of the document has been estimated. We propose to learn a distribution over styles across centuries using Topic Models and to apply a classifier over weights learned in order to estimate the age of the documents. We also present a comparison of different distance metrics such as Euclidean Distance and Hellinger Distance.

8297-26, Session 8

Handwriting individualization using distance and rarity

Y. Tang, S. N. Srihari, Univ. at Buffalo (United States); H. Srinivasan, Janya Inc. (United States)

The likelihood ratio (LR) is a quantitative measure of the degree of uncertainty in forensic individualization, in which the central problem is to determine whether the observed evidence came from a particular source/object. Evaluation of the LR based on the feature distribution needs a large number of parameters for the model, which is data demanding and computationally expensive. One simplification is to compute the ratio of the likelihoods of the distance. Although being simple to compute, there is a severe loss of information in going from a high-dimensional feature space to a one-dimensional distance. This paper proposed an intermediate solution, in which the LR is decomposed as the product of two factors, distance and rarity. The proposed method was evaluated using a data set of handwriting samples, and compared with the feature distribution method and distance methods. The results showed that the proposed method outperforms the distance methods and the feature distribution method in terms of accuracy, with acceptable efficiency.

8297-27, Session 8

Construction of language models for an handwritten mail reading system

O. Morillot, L. Likforman-Sulem, Telecom ParisTech (France); E. Grosicki, Direction Générale de L'armement (France)

This paper presents a system for the recognition of unconstrained handwritten mails. The main part of this system is an HMM recognizer which uses trigraphs to model contextual information. This recognition system does not require any segmentation into words or characters and directly works at line level. To take into account linguistic information and enhance performance, a language model is introduced. This language model is based on bigrams and built from training document transcriptions only. Different experiments with various vocabulary sizes and language models have been conducted. Word Error Rate and Perplexity values are compared to show the interest of specific language models, fit to handwritten mail recognition task.

Conference 8298: Sensors, Cameras, and Systems for Industrial/Scientific Applications XIII

Wednesday-Thursday 25-26 January 2012

Part of Proceedings of SPIE Vol. 8298 Sensors, Cameras, and Systems for Industrial and Scientific Applications XIII

8298-27, Poster Session

The embedded network infrared video monitoring system based on Linux OS

L. Liu, Nanjing Univ. of Science & Technology (China); N. Chen, Nanjing Normal Univ. (China); X. Zhou, T. Pan, Nanjing Univ. of Science & Technology (China)

The video monitoring system could provide remote video information. It is convenient and intuitive, so it could be widely used in industrial, transportation, finance, teaching and other fields. With the development of computer and network technology, network video monitoring system will replace the traditional monitoring system and play an important role in many fields.

This paper describes the way to design an embedded network infrared video monitoring system based on Linux OS. Firstly, we make a comparison of the hardware solution between some regular monitoring systems, and then design the hardware system that we needed. Our hardware system uses the i.mx27 processor with the ARM9 core. Secondly, the software platform is introduced in this paper. The Linux operate system is applied in our software solution. According to the characteristic of Linux OS, we download uboot to the demo board, transplant the Linux kernel and jffs2 file system to the embedded system, and briefly compile and download drivers. Finally, the application software design process is introduced in the paper. The system can be used to encode the picture captured from infrared CCD, and then send the picture to another same embedded system to decode the picture, and finally display it on the LCD and achieve the goal of the infrared video's remote monitoring. As the infrared CCD would not be affect by the dim light, this monitoring system could be used all day long.

8298-28, Poster Session

Motion blur-free time-of-flight range sensor

S. Lee, B. Kang, J. D. Kim, C. Kim, Samsung Advanced Institute of Technology (Korea, Republic of)

Time-of-flight depth sensor provides faster and easier way of 3D scene capturing and reconstruction. The depth sensor, however, suffers from motion blur caused by any movement of camera or subjects. In this manuscript, we propose a novel depth motion blur pixel detection and elimination method that can be implemented on the depth sensor with light memory and computation resources.

We propose an easy blur decision method using the plots of charge relations and detect blur pixel at each depth value calculation step only by checking the four charge values. Once we detect blur pixels, they are replaced by any closest normal pixel values. With this method, we eliminate blur pixel before we build the depth image with only few more calculations and memory addition.

8298-29, Poster Session

CMOS buried double junction (BDJ) photodiode for trichromatic sensing

L. Tu, S. V. Setlur Nagesh, Univ. at Buffalo (United States); Z. Fu, OmniVision Technologies, Inc. (United States); A. H. Titus, Univ. at Buffalo (United States)

A novel method for achieving trichromatic color detection using a single photodetector with less than three p-n junctions is presented.

This new method removes the constraint of color sensing in buried-double-junction (BDJ) photodiode [1], eliminates the need for a priori light source knowledge [2] or for changing color intensity [3]. After using a visible light optical filter to block light from outside the visible spectrum, the color detection is achieved by taking the difference of depletion region photocurrent generated by different reverse bias voltages. This "difference output" is effectively from the "third" optical wavelength specific depletion region required for trichromatic color sensing. This method is based on exploiting the relationship between photon absorption and photon penetration depth of silicon [5, 6], and the basic property of a p-n junction photodiode [7, 8] which states that only photons absorbed in depletion region generate photo-signals. Varying photodiode's reverse bias implies varying depletion region width, which further implies photon absorption of different wavelength/frequency. The theory is validated by experiment using a BDJ photodiode fabricated through MOSIS in the AMI-ABN 1.5um technology, a visible light filter, a narrow wavelength LED, and a current to voltage converter built using operational amplifiers. Trichromatic color detection with peak wavelength of 500, 600, and 700 um photon wavelength is achieved.

8298-30, Poster Session

On image sensor dynamic range utilized by security cameras

A. Johannesson, Axis Communications AB (Sweden)

The dynamic range is an important quantity used to describe an image sensor. Wide/High/Extended dynamic range is often brought forward as an important feature to compare one device to another. The dynamic range of an image sensor is normally given as a single number, which is often insufficient since a single number will not fully describe the dynamic capabilities of the sensor.

A camera is ideally based on a sensor that can cope with the dynamic range of the scene. Otherwise it has to sacrifice some part of the available data. For a security camera the latter may be critical since important objects might be hidden in the sacrificed part of the scene.

In this paper we compare the dynamic capabilities of some image sensors utilizing a visual tool. The comparison is based on the use case, common in surveillance, where low contrast objects may appear in any part of a scene that through its uneven illumination span a high dynamic range. The investigation is based on real sensor data that has been measured in our lab and a synthetic test scene is used to mimic the low contrast objects. With this technique it is possible to compare sensors with different intrinsic dynamic properties as well as some capture techniques used to create an effect of increased dynamic range.

8298-31, Poster Session

Design of low-noise output amplifier for p-channel: fully depleted charge-coupled devices

S. Haque, S. E. Holland, A. Karcher, W. Kolbe, N. Roe, Lawrence Berkeley National Lab. (United States); R. Groulx, R. Frost, F. Dion, Teledyne DALSA Semiconductor (Canada); G. Wang, Lawrence Berkeley National Lab. (United States)

We describe the design and development of low-noise, single-stage output amplifiers for p-channel charge-coupled devices (CCDs) used for scientific applications in astronomy and other fields. The CCDs are

fabricated on high-resistivity, > 4000 Ω-cm, n-type silicon substrates. Single-stage amplifiers with different output structure designs and technologies have been characterized. The standard output amplifier is designed with an n+ polysilicon gate that has a metal connection to the sense node. In an effort to lower the output amplifier readout noise by minimizing the capacitance seen at the sense node, buried contact technology has been investigated. In this case, the output transistor has a p+ polysilicon gate that connects directly to the p+ sense node. Output structures with buried contact holes as small as 2 um x 2 um are characterized. In addition, the size of the source follower transistor was varied. We report test results on the conversion gain and noise of the various amplifier structures. By optimizing the amplifier biases, resistive loads and test electronics design, a read noise of less than 2 electrons at 70 kpixels/sec and -140 °C is obtained, corresponding to a reduction in noise of 45% compared to the standard output amplifier configuration.

8298-32, Poster Session

S/N improvement for the optical-multiplex image-acquisition system

T. Narabu, Sony Corp. (Japan)

The optical-multiplex system is comprised of an image sensor, a multi-lens array and signal processing unit. The key feature of the optical-multiplex system is that each pixel of the image sensor captures multiple data of the object through multi-lenses and the object data is obtained by processing the raw data output from the optical-multiplex image sensor.

We are now able to report that our system can improve the signal-to-noise ratio of the image output from the optical-multiplex system by changing the shading characteristics of the multi-lenses in the optical-multiplex system. In a model of the system for simulation purposes, an optical-multiplex system with five lenses is used. The five lenses are located at the center, upper, lower, left and right above an image sensor. We calculate the signal-to-noise ratio of the image output from the optical-multiplex system by changing the shading characteristics of the four lenses located at the upper, lower, left and right. The best signal-to-noise ratio of this image output by the optical-multiplex system is 8.895 dB better than that of a camera with a single lens. This value is beyond the value of 3.764 dB which we reported in 2007.

8298-33, Poster Session

Fully integrated system-on-chip for pixel-based 3D depth and scene mapping

M. Popp, B. De Coi, M. Thalmann, R. Gancarz, P. Ferrat, M. Dürmüller, F. Britt, M. Annese, M. Ledergerber, G. Catregn, ESPROS Photonics AG (Switzerland)

We present for the first time a fully integrated System on Chip for pixel-wise 3D range detection suited for commercial applications. It is based on the time-of-flight (ToF) principle, i.e. measuring the phase difference of a reflected pulse train. The product epc600 is fabricated using a dedicated process flow, called Espros Photonic CMOS®. This integration makes it possible to achieve a Quantum Efficiency (QE) of >80% in the full wavelength band from 500nm up to 900nm as well as very high timing precision in the sub-ns range which is needed for exact detection of the phase delay. The SoC features 8x8 pixels and includes all necessary sub-components such as ToF pixel array, voltage generation and regulation, non-volatile memory for configuration, LED driver for active illumination, digital SPI interface for easy communication, column based 12bit ADC converters, PLL and digital data processing with temporary data storage. The system can be operated at up to 100 frames per second. Performance data of the system will be shown such as QE, timing capabilities, range resolution as well as tolerance versus ambient background light.

8298-01, Session 1

High-speed VGA resolution CMOS image sensor with global shutter

P. Willems, G. Vanhorebeek, C. Ma, CMOSIS nv (Belgium)

CMOSIS presents its new member in the CMV (CMOSIS Machine Vision) family. The CMV300 is a high speed CMOS image sensor with 644 by 484 pixels (1/3 optical inch) developed for machine vision applications. The image array consists of 7.4µm x 7.4µm pipelined global shutter pixels which allow exposure during read out, while performing CDS operation. The image sensor has four 8-, 10- or 12-bit digital LVDS outputs (serial) or one 10 bit parallel CMOS output. The image sensor also integrates a programmable gain amplifier and offset regulation. Each LVDS channel runs at 600 Mbps maximum which results in 600 fps frame rate at full resolution. Higher frame rates can be achieved in row-windowing mode or row-subsampling mode. These modes are all programmable using the SPI interface. All internal exposure and read out timings are generated by a programmable on-board sequencer. External triggering and exposure programming is also possible. Extended optical dynamic range can be achieved by multiple integrated high dynamic range modes. The packaged BGA sensor size is only 7mm x 7mm x 0.7mm which allows for very small and cost effective cameras. Architecture, building blocks, pixel and column ADC are described into detail in the paper.

8298-02, Session 1

High-speed global shutter CMOS machine vision sensor with high-dynamic range image acquisition and embedded intelligence

Á. B. Rodríguez-Vázquez, Univ. de Sevilla (Spain); R. Dominguez-Castro, F. Medeiro, F. Jimenez-Garrido, Anafocus (Spain)

High-speed imagers are required for industrial applications, traffic monitoring, robotics and unmanned vehicles, movies, etc. Many of these applications require also high spatial resolution, high sensitivity and the ability to detect images with large intra-frame dynamic range. This paper reports a CIS intelligent digital image sensor with 5.2Mpxl which delivers either 12-, 10-, or 8-bit fully-corrected images at 250Fps. The new sensor incorporates sophisticated functions on-chip, such as: windowing; pixel binning; sub-sampling; combined windowing-binning-sub-sampling modes; fixed-pattern noise correction; fine gain and offset control; and colour processing. These and other CIS functions are programmable through a simple four-wire serial port interface (SPI).

8298-03, Session 1

High-speed CMOS image sensor for high-throughput lensless microfluidic imaging system with point-of-care application

M. Yan, X. Huang, R. Nadipalli, Y. Shang, H. Yu, Nanyang Technological Univ. (Singapore); M. Je, A*STAR Institute of Microelectronics (Singapore); C. Xu, OmniVision Technologies, Inc. (United States); K. Yeo, Nanyang Technological Univ. (Singapore)

The miniaturized lensless microfluidic bio-imaging system has great potential to replace existing bulky and expensive clinic laboratory instruments for detecting, counting, and sorting biomedical samples. The primary challenge is to design compact and high-speed CMOS imager sensor for the system integration. This paper has demonstrated design of the first CMOS imager chip for high-throughput flow-cytometer bioimagining system. The chip is fabricated in 0.18um 1P6M CMOS process. Compared to off-the-shelf image sensor, the proposed novel image sensor has three unique features: (1) high sensitivity and better SNR achieved by optimizing 10um-size pixel with fill factor of 60%. Super-resolution imaging analysis algorithm overcome the pixel physical resolution limitation and be able to reconstruct 2um-pixel

image; (2) column-parallel readout architecture with DCDS technique to achieve both ultra-high speed (up to 1000frames/sec) and low noise at the same time; (3) a real-time on-chip imaging processing is implemented with system-on-a-chip approach to boost the sensor performance. In conclusion, this presented image sensor chip design will pave the way for implementation of commercialize autonomous, potable and low-cost microfluidic lab-on-chip platform, such as HIV diagnostics and monitoring, circulating tumor cell capturing, sepsis and bacterial pathogen detection etc.

8298-04, Session 2

Smart image sensor with adaptive correction of brightness

M. Paindavoine, A. Ngoua, O. Brousse, C. Clerc, Univ. de Bourgogne (France)

Today, intelligent image sensors require the integration in the focal plane (or near the focal plane) of complex image processing algorithms. Such devices must meet the constraints related to the quality of acquired images, speed and performance of embedded processing, as well as low power consumption. To achieve these objectives, analog pre-processing, on the one hand, improve the quality of the images making them usable whatever the light conditions, and secondly, detect regions interest (ROIs) in order to limit the amount of pixels to be transmitted to a digital processor performing the high-level processing such as feature extraction for pattern recognition. In this context, our goal was to design an intelligent image sensor prototype, built-in 130nm CMOS technology, and to achieve a triple analog function: image acquisition in multi-exposure mode, detection of image blocks with a constant average brightness and high dynamic range image generation.

8298-05, Session 2

Algorithm architecture co-design for ultra low-power image sensor

T. Laforest, A. Dupret, A. Verdant, D. Lattard, P. Villard, CEA-LETI (France)

Stand alone leftbehind image sensor are used to achieve the detection of events with very low power consumption and in a relevant way. Using a steady camera, motion detection algorithms based on static scene estimation to find regions in movement are simple to implement and computationally efficient. Their implementation had already been done on an integrated architecture with programmable analog processors. In order to extend the class of objects to be detected, we propose an original mixed mode architecture developed thanks to an algorithms and architecture co-design methodology. That programmable architecture is composed of a vector of SIMD processors which are able to process images high definition of 1920x1080 pixels. However, the background is estimated using a down sampled image formed of macropixels to reduce power consumption. A basic RISC architecture was optimized in order to implement motion detection algorithms with a dedicated set of 42 instructions. Definition of delta modulation as a calculation primitive had allowed to implement algorithms in a very compact way. Thereby, motion detection is achieved with the lowest power consumption per pixel known until now.

8298-06, Session 2

A CMOS imager using focal-plane pinhole effect for confocal multi-beam scanning microscopy

M. Seo, A. Wang, Z. Li, K. Yasutomi, K. Kagawa, S. Kawahito, Shizuoka Univ. (Japan)

A CMOS imager for confocal multi-beam scanning microscopy, where the pixel itself works as a pinhole, is proposed. This CMOS imager is suitable for building compact, low-power, and non-cooling confocal microscopes because the complex Nipkow disk with a precisely aligned pinhole array can be omitted. The CMOS imager is composed of an array of sub-imagers, and can detect multiple beams at the same time. To achieve a focal-plane pinhole effect, only one pixel in each sub-imager, which is at the conjugate position of a light spot, accumulates the photocurrent, and the other pixels are not read and reset. This operation is achieved by 2-stage vertical and horizontal shift registers. The proposed CMOS imager for the confocal multi-beam scanning microscope system was fabricated in 0.18-μm standard CMOS technology with a pinned photodiode option. The total area of the chip is 8.4mm×9.4mm. The number of effective pixels is 256(Horizontal)×256(Vertical). The pixel array consists of 32 (H)×32 (V) sub-imagers each of which has 8(H)×8(V) pixels. The pixel is an ordinary 4-transistor active pixel sensor using a pinned photodiode and the pixel size is 7.5μm×7.5μm. The basic operations such as normal image acquisition and selective pixel readout were experimentally confirmed.

8298-07, Session 2

Time-to-impact sensors in robot vision applications based on the near-sensor image processing concept

A. Astrom, Combitech AB (Sweden); R. Forchheimer, Linköping Univ. (Sweden)

The concept of optical flow has been known for more than 30 years. It is also known that this is a memory consuming and computationally demanding task requiring powerful processors. Recently, it has been described how to implement optical flow for time-to-impact, TTI, detection using the Near-Sensor Image Processing, NSIP concept, which was described for the first time 30 years ago, in which an optical sensor array and a specific low-level processing unit are tightly integrated into a hybrid analog-digital device. Despite its low overall complexity, numerous image processing operations can be performed at high speed competing favorably with state-of-art.

The resulting performance for the TTI algorithm on the NSIP concept would be in the order of 10 kHz of time-to-impact calculations. The reason for the high performance is that the TTI algorithm fits very well into the NSIP architecture. TTI is defined as the distance to the object divided by the speed towards the object.

In this paper we will present in more detail NSIP-based TTI applications for robot vision. We will show that TTI from a single sensor can be used for calculation of the angle between the object's surface, that we are approaching, and the travel direction. We will also show that TTI from two different sensors can be used to calculate the absolute distance. Finally, we will show how the accumulated information during the travelling time will give us more detailed information about the object. This can be described as "Shape-from-TTI".

8298-08, Session 3

A 176x144 148dB adaptive tone-mapping imager

S. Vargas-Sierra, G. Liñán-Cembrano, Á. B. Rodríguez-Vázquez, Univ. de Sevilla (Spain)

This work presents a QCIF HDR imager where visual information is simultaneously captured and adaptively compressed by means of an in-pixel tone mapping scheme. The tone mapping curve (TMC) is calculated from a subsampled nonlinear histogram of the previous image, which serves as a probability indicator of the distribution of illuminations within the present frame. The chip produces 7-bit/pixel images in a way that each pixel decides when to stop observing photocurrent integration. It is capable of compressing illuminations from 222μlux to 55392 lux in a single frame, with extreme values captured at 8s and 2.34μs respectively. It implies a photocurrent dynamic range of 168dB for SNR=1 and 148dB for SNR=10. Pixels use a 3x3μm2 Nwell-Psubstrate photodiode and an autozeroing technique for establishing the reset voltage, which cancels out most of the offset contributions created by the analog processing circuitry. Measured

sensitivity is 5.79 V/lux-s. Dark signal (10.8 mV/s) effects in the final image are attenuated by an automatic programming of the on-chip DAC top voltage. The chip has been designed in the 0.35μm OPTO technology from AMS.

8298-09, Session 3

A high-dynamic range (HDR) back-side illuminated (BSI) CMOS image sensor for extreme UV detection

X. Wang, CMOSIS nv (Belgium)

This paper describes a back-side illuminated 1 Megapixel CMOS image sensor made in 0.18 um CMOS process for EUV detection. First, the design of the front-side illuminated sensor are described, including the low noise, high dynamic range (HDR) pixel architecture, the column gain stage and readout circuitary. Secondly, the back-side illumination post-processing steps are explained, including the wafer-to-wafer bonding, thinning, buried-oxide (BOX) removal and pad opening. In the end, sensor characterization result is shown, including general electrical optical performance and EUV detection characterization.

8298-10, Session 3

A low-noise, 15μm pixel-pitch, 640x512 hybrid InGaAs image sensor for night vision

F. Guellec, S. Dubois, E. de Borniol, P. Castelein, S. Martin, R. Guiguet, M. Tchagaspanian, A. Rouvie, P. Bois, CEA-LETI (France)

Hybrid InGaAs focal plane arrays are very interesting for night vision because they can benefit from the nightglow in the Short Wave Infrared band. Through a collaboration between III-5 Lab and CEA-Leti, a 15μm pixel-pitch, 640x512 InGaAs image sensor with 15μm pixel pitch has been developed. The readout IC (ROIC) design in a standard CMOS 0.18μm technology is presented. The pixel circuit is based on a capacitive transimpedance amplifier (CTIA) stage and offers a selectable gain. The input stage is optimized to deliver low noise performance at low light levels in the high gain mode. In this mode, the charge-to-voltage conversion factor is around 17.5μV/electron. The exposure time can be maximized up to the frame period thanks to a rolling shutter approach. The frame rate can be up to 120fps or 60fps if the Correlated Double Sampling (CDS) capability of the circuit is enabled. The first results show a dark noise of 90 electrons for a full-well capacity above 105e3 electrons. They are obtained at room temperature with a reverse photodiode bias voltage of 100mV, the maximum exposure time (16ms at 60fps) and CDS. To our best knowledge, this set of results establishes state-of-the-art performance for this kind of detector.

8298-11, Session 3

High-dynamic range, 4 megapixel CMOS image sensor for scientific applications

P. Vu, B. A. Fowler, C. Liu, S. Mims, P. Bartkovjak, H. Do, W. Li, J. Appelbaum, A. Lopez, BAE Systems Imaging Solutions (United States)

As bio-technology transitions from research and development to high volume production, dramatic improvements in image sensor performance will be required to support the throughput and cost requirements of this market. This includes higher resolution, higher frame rates, higher quantum efficiencies, increased system integration, lower read-noise, and lower device costs. We present the performance of a recently developed low noise 2048(H) x 2048(V) CMOS image sensor optimized for scientific applications such as life science imaging, microscopy, as well as industrial inspection applications. The sensor architecture consists of two identical halves which can be operated independently and the imaging array consists of 4T pixels with pinned photodiodes on a 6.5μm pitch with integrated micro-lens.

The operation of the sensor is programmable through a SPI interface. The measured peak quantum efficiency of the sensor is greater than 72% at 600nm, and the read noise is about 1.1e- RMS at 100 fps data rate. The sensor features dual gain column parallel ouput amplifiers with 11-bit single slope ADCs. The full well capacity is greater than 36ke-, the dark current is less than 6pA/cm2 at 20ºC. The sensor achieves an intra-scene linear dynamic range of greater than 91dB (36000:1) at room temperature.

8298-36, Session 3

Diffusion dark current in front-illuminated CCDs and CMOS image sensors

M. M. Blouke, Portland State Univ. (United States)

Historically the diffusion related component of the dark current has been estimated from the diffusion related current of a diode with an infinite substrate. This paper explores the effect of a substrate of finite extent beneath the collecting volume of the pixel for a front-illuminated device and develops a corrected expression for the diffusion related dark current. Calculations show that the diffusion dark current can be much less, by a couple orders of magnitude, than that predicted by the standard model.

8298-12, Session 4

Projecting the rate of in-field pixel defects based on pixel size, sensor area, and ISO

G. H. Chapman, J. Leung, A. Namburete, Simon Fraser Univ. (Canada); I. Koren, Z. Koren, Univ. of Massachusetts Amherst (United States)

Image sensors continuously develop in-field permanent hot pixel defects over time. Experimental measurements of DSLR, point and shoot, and cell phone cameras, show that the rate of these defects depends on the technology (APS or CCD) and on design parameters like imager area, pixel size, and gain (ISO). Expanding image sensitivity (ISO) enhances defects appearance and sometimes creates saturation. 40% of defects are partially stuck hot pixels, with an offset independent of exposure time, and are particularly affected by ISO changes. Comparing different sensor sizes with similar pixel sizes showed that defect rates scale linearly with sensor area suggesting the metric of defects/year/sq mm. Plotting this rate for different pixel sizes (7.5 to 1.5 microns) shows that defect rates grow rapidly as pixel size shrinks. Curve fitting shows an empirical power law with defect rates proportional to the pixel size to the power of -2 for CCD and to the power of -3.3 for CMOS. At 7um pixels the CCD defect rate is ~2.5 greater than for CMOS, but for 2um pixels the rates are equal. Extending our empirical formula to include ISO allows us to predict the expected defect development rate for a wide set of sensor parameters.

8298-13, Session 4

Dynamic CCD pixel depletion edge model and the effects on dark current production

J. C. Dunlap, M. M. Blouke, E. Bodegom, R. Widenhorn, Portland State Univ. (United States)

The depletion edge in Charge-Coupled Devices (CCD) pixels is dependent upon the amount of signal charge located within the depletion region. A model is presented that describes the movement of the depletion edge with increasing signal charge. This dynamic depletion edge is shown to have an effect on the amount of dark current produced by some pixels. Modeling the dark current behavior of pixels both with and without impurities over an entire imager demonstrates that this moving depletion edge has a significant effect on a subset of the pixels. Dark current collected by these pixels is shown to behave nonlinearly with respect to exposure time and additionally the dark current is affected by the presence of illumination. The model successfully predicts unexplained aspects of dark current behavior previously observed in some CCD sensors.

8298-14, Session 4

Characterizing the response of charge-couple device digital color cameras

V. Slavkovikj, J. Y. Hardeberg, Gjøvik Univ. College (Norway); A. Eichhorn, Simula Research Lab. (Norway)

The advance and rapid development of electronic imaging technology has lead the way to production of imaging sensors capable of acquiring good quality digital images with a high resolution. At the same time the cost and size of imaging devices have reduced. This has incited an increasing research interest for techniques that use images obtained by multiple camera arrays. Use of multi-camera arrays is attractive because it allows capturing of multi-view images of dynamic scenes, enabling the creation of novel computer vision and computer graphics applications, as well as next generation video and television systems.

There are additional challenges when using a multi-camera array, however. Due to inconsistencies in the fabrication process of imaging sensors and filters, multi-camera arrays exhibit inter-camera color response variations. For the majority of applications, which use multi-view images obtained from multi-camera arrays, it is insufficient to assume that the different camera's response can be considered the same without prior verification. Therefore, it is necessary to characterize the response of the different cameras in the array.

8298-15, Session 4

Implementing and using the EMVA1288 standard

A. Darmont, J. Lemaitre, J. Chahiba, Aphesa SPRL (Belgium)

The European Machine Vision Association took in the last years the initiative of developing a measurement and reporting standard for industrial image sensors and cameras called EMVA1288.

Aphesa offers camera and sensor measurement services and test equipment according to this EMVA1288 standard. We have measured cameras of various kinds on our self-made test equipment. This implementation and all the measurement sets require to go in the details of the standard and also show us how good it can be but also how difficult it can be.

The purpose of this paper is to give feedback on the standard based on our experience of implementers and users. We will see that some measurements are easily reproducible and can be easily implemented while others require more research on hardware, software and procedures and that the results can sometimes have very little meaning.

Our conclusion will be that the EMVA1288 standard is good and well suited for the measurement and characterization of image sensors and cameras for the image processing applications but that it is hard for a newcomer to understand the produced data and properly use a test equipment. Developing a complete and compliant test equipment is also a difficult task.

8298-34, Session 4

Image sensor for defence applications

E. Belhaire, J. L. Fontanella, Thales Optronique S.A. (France)

The defense market makes intensive use of image sensors. The equipments integrate more and more image sensors to fulfill the primary observations missions: detection, recognition and identification. These tasks has to be fulfilled in more and more complex situations: during night and under daylight, in severe urban environment, in front of complex backgrounds, at long distance, using small size displays under conditions of heavy parasitic lighting and in stressful situations. The traditional infrared longwave image sensors are now completed or replaced by image sensors in other atmospheric transmissions bands. Among those CMOS and CDD image sensors are primary considered, either in their traditional customer configuration or, more frequently with adaptations. Those adaptations are realized on standard image sensor technology and are limited to the strict minimum required by the applications in order to reduce the cost.

The use of technologies, stated as "dual use", is primary considered as the defense market cannot support the important research and development costs of all the technologies required by defense applications. In consequence, the defense market is mainly concentrating its R&D funding on technologies without large scale applications in the customer market.

The different wavebands of interest are presented. The main technologies for each band are cited. Then, the image sensor specification adaptations, required by the defense market, are reviewed and discussed in this paper. Some example of typical future defense applications will be presented. When possible, application counterpart in the customer market will be pointed.

8298-35, Session 4

An overview of the European patent system with particular emphasis on IP issues for imaging devices

M. Boero, A. Cabrita, European Patent Office (Netherlands)

In this article we give a comprehensive review of the European Patent System with focus on the procedure, its typical duration, the requirements that must be met at the various stages in order to obtain an European Patent and its related costs. All the options available to the applicant are discussed in detail, potential pitfalls are highlighted, and the differences between the European and US Patent Systems are analysed.

Furthermore, an in-depth and very informative analysis of applications and granted patents in the field of imaging devices is presented including a study of their evolution during the last 10 years together with an analysis of the countries and companies that are most active in the field of imagers.

8298-16, Session 5

Development of high-transmittance, back-illuminated, silicon-on-sapphire substrates thinned below 25 micrometers and bonded to fused silica for high-quantum efficiency and high-resolution avalanche photodiode imaging arrays

A. G. Stern, AG Stern, LLC (United States)

There is a growing need in scientific and industrial applications for dual-mode, passive and active 2D and 3D LADAR imaging methods. To fill this need, solid-state, single photon sensitive silicon avalanche photodiode (APD) detector arrays offer high sensitivity and the possibility to operate with wide dynamic range in dual linear and Geiger-mode for passive and active imaging. To support the fabrication of large scale, high quantum efficiency and high resolution silicon avalanche photodiode arrays and other advanced solid-state optoelectronics, a novel, high transmittance, back-illuminated silicon-on-sapphire substrate has been developed incorporating a single crystal, epitaxially grown aluminum nitride (AlN) antireflective layer between silicon and R-plane sapphire, that provides refractive index matching to improve the optical transmittance into silicon from sapphire. A one quarter wavelength magnesium fluoride antireflective layer deposited on the back-side of the sapphire improves optical transmittance from the ambient into the sapphire. The magnesium fluoride plane of the Si-(AlN)-sapphire substrate is bonded to a fused silica wafer that provides mechanical support, allowing the sapphire to be thinned below 25 micrometers to improve back-illuminated optical transmittance, while suppressing indirect optical crosstalk from APD emitted light undergoing reflections in the sapphire, to enable high quantum efficiency and high resolution detector arrays.

Conayerence 8298

8298-17, Session 5

29 mp, 35 mm format interline CCD image sensor

E. J. Meisenzahl, D. A. Carpenter, J. E. Doran, R. P. Fabinski, S. L. Kosman, J. P. McCarten, Truesense Imaging, Inc. (United States)

This paper describes the design and performance of a new high-resolution 35 mm format CCD image sensor using an advanced 5.5 um interline pixel. The pixels are arranged in a 6576 (H) × 4384 (V) format to support a 3:2 aspect ratio. This device is part of a family of devices that share a common architecture, pixel performance, and packaging arrangement. Unique to this device in the family is the implementation of a fast line dump structure and horizontal CCD lateral overflow drain.

8298-18, Session 5

Photodiode dopant structure with atomically flat Si surface for high-sensitivity and stability to UV light

T. Nakazawa, R. Kuroda, Y. Koda, S. Sugawa, Tohoku Univ. (Japan)

Ultraviolet light (UV-light) sensing, especially in about 200-300nm wavelength, is required widely. Recently, the technology to fabricate photodiode with almost 100% internal quantum efficiency and high stability to UV-light by forming the surface photo-generated carrier drift layer uniformly by using an atomically flat Si surface has been reported [1].

In this work, by using this atomically flat Si surface, we investigate the relationships between the surface drift layer profiles and sensitivity and stability to UV-light. Thus, the sensitivity of n+pn-type photodiodes with various n+ layer profiles as well as, fixed charges (Qf) and interface states (Dit) were measured before and after UV-light exposure. Finally, we clarify the mechanism of photodiode's characteristics, and propose a guideline to achieve high stability to UV-light.

Based on the measurement results, 3 types of photodiodes' characteristics were explained by one mechanism.

In conclusion, to achieve high sensitivity and stability to UV-light, suppressing the change in Qf and Dit by UV-light and suppressing the change of drift field caused by the change in Qf are necessary. The combination of an atomically flat Si surface and a surface drift layer with sufficiently high concentration was demonstrated to be effective for them.

[1] R.Kuroda et al., "High Ultraviolet Light Sensitive and Highly Reliable Photodiode with Atomically Flat Si Surface" International Image Sensor Workshop, pp.38-41, 2011

8298-19, Session 5

New smart readout technique performing edge detection designed to control vision sensors dataflow

H. Amhaz, G. Sicard, TIMA Lab. (France)

Integrating CMOS active pixel sensors (APS) with carefully chosen signal processing units has nowadays become a trend in the design of camera-on-chip systems. In this paper we present a new readout technique designed to control the excessive dataflow generated by high resolution image sensors. This new readout method achieves edge detection without touching the pixel area since the processing circuitry is implemented in the periphery of the matrix. This point is very essential and important from the industrial point of view. By using this readout scheme we can benefit from the reduction of the power consumption and the increase of the readout speed. This technique presents several advantages and liberty degrees that permit several functionality of the same sensor. MAtlab modeling of the readout scheme and the electrical simulations of the analogue blocks will also be presented. What we essentially consider as an innovation is the fact of not touching the in-pixel circuitry and so the pixel area and Fill

Factor. Several researches have already been studied but none of them presents such an independency vis a vis the pixel circuitry and area.

8298-20, Session 5

Characterization of orthogonal transfer array CCDs for the WIYN one-degree imager

M. P. Lesser, D. Ouellette, The Univ. of Arizona (United States); T. Boroson, National Optical Astronomy Observatory (United States); D. Harbeck, P. Martin, WIYN Observatory (United States); G. Jacoby, Carnegie Observatories (United States); J. Cavin, Univ. of Wisconsin-Madison (United States); D. Sawyer, National Optical Astronomy Observatory (United States); K. Boggs, R. Bredthauer, Semiconductor Technology Associates Inc. (United States)

The WIYN One Degree Imager (ODI) will provide a one degree field of view for the WIYN 3.5 m telescope located on Kitt Peak near Tucson, Arizona. Its focal plane consists of an 8x8 grid of Orthogonal Transfer Array (OTA) CCD detectors. These detectors are the STA2200 OTA CCDs designed and fabricated by Semiconductor Technology Associates, Inc. and backside processed at the University of Arizona Imaging Technology Laboratory. Several lot runs of the STA2200 detectors have been fabricated. We have backside processed devices from these different lots and provide detector performance characterization, including noise, CTE, cosmetics, quantum efficiency, and some orthogonal transfer characteristics. We discuss the performance differences for the devices with different silicon thickness and resistivity. A fully buttable custom detector package has been developed for this project which allows hybridization of the silicon detectors directly onto an aluminum nitride substrate with an embedded pin grid array. This package is mounted on a silicon-aluminum alloy which provides a flat imaging surface of less than 20 microns peak-valley at the -100 C operating temperature. Characterization of the package performance including low temperature profilometry is described in this paper.

8298-21, Session 6

Multispectral imaging device for help in diagnosis

C. Delporte, M. Ben Chouikha, Univ. Pierre et Marie Curie (France); S. Sautrot, Univ. Denis Diderot (France); F. Viénot, Muséum National d'Histoire Naturelle (France); G. Alquié, Univ. Pierre et Marie Curie (France)

This work focuses on development of a multispectral imaging system for intraoperative diagnosis aid purpose. The system should be able to identify and differentiate biological tissues in an operative scene, by measurement of their spectral reflectance in the visible and the near infrared wavelength range. Multispectral images in the near infrared may be used to see in depth and to identify buried tissues.

The paper introduces a multispectral imaging device we have developed and validated by in vitro measurements. In vivo tests need specific measurements protocol and should be carried later. The imaging system is composed by one calibrated CCD camera, by one optical device with three holographic gratings and by halogen light source. Multispectral images produced by this system are obtained according to the 0°/45° CIE standard geometry.

In order to validate this system to evaluate measurements fluctuations several tests were made on different muscle slices. This study showed a fluctuation of 20% to 30% on the same slice of a given tissue. This large error rises from the intrinsic, complex and non homogeneous nature of biological tissues. It represents a great limitation to this spectral measurements technique and affects its ability in terms of tissues spectral reflectance factor evaluation.

8298-22, Session 6

Development of a driving method suitable to ultra-high-speed shooting in 2M-fps, 300k-pixel single-chip color camera

J. Yonai, T. Arai, Japan Broadcasting Corp. (Japan); T. Hayashida, NHK Engineering Services, Inc. (Japan); H. Ohtake, Japan Broadcasting Corp. (Japan); T. G. Etoh, Kinki Univ. (Japan); T. Yoshida, J. Namiki, Hitachi Kokusai Electric Inc. (Japan)

We have previously manufactured a 1M-fps high-speed camera for shooting TV programs. Using this camera has revealed certain problems such as decrease in image quality and overheating of the camera at speeds higher than 200k fps. To circumvent these problems, we developed a new ultra-high-speed camera that enabled practical camera speeds up to2M fps by changing the CCD layout and improving the diving method.

In the previous model, image quality decreased with increase in speed at more than 200k-fps due to distortion of the CCD driving pulse and reduction in dynamic range. The CCD was redesigned to reduce interconnection resistance, and simulations confirmed that reducing the RC time constant solved the above problems. Also, the increase in current when shooting at high speeds resulted in overheating of the drive circuit board. Shooting continuously for one minute at 1M fps abruptly increased the drive circuit board temperature from 27 degC to around 80 degC. To prevent excessive heating, we devised a method for driving the CCD only within the minimum time required during high-speed shooting. We experimentally manufactured an ultra-high-speed camera using this method and confirmed that heating did not cause problems even at 2M fps.

8298-24, Session 6

Accurate color with increased sensitivity using IR

A. Enge, Truesense Imaging, Inc. (United States)

Many applications require accurate color captures in daylight conditions and increased sensitivity for low light conditions. These captures are often accomplished by using a mechanical switch to remove the IR cut filter. The sensitivity is increased at the expense of color accuracy, and a mechanical part is required in the camera. A sparse color filter pattern offers an opportunity to obtain increased sensitivity-using the IR region-while still maintaining color accuracy. A 2x increase in sensitivity can be achieved over the current KODAK TRUESENSE Color Filter Pattern, which uses an IR cut filter.

8298-25, Session 6

Computational color constancy using chromagenic filters in color filter arrays

R. Shrestha, J. Y. Hardeberg, Gjøvik Univ. College (Norway)

We have proposed a new illuminant estimation technique based on extension of chromagenic based color constancy in this paper. Basic chromagenic illuminant estimation method takes two shots of a scene, one without and one with a specially chosen color filter in front of the camera lens. Here, we introduce chromagenic filters on top of R, G or B filters in place of one of the two green filters in the Bayer's pattern. Introduction of chromagenic filters allow to obtain two images of the same scene via demosaicking, a normal RGB image, and a chromagenic image, equivalent of RGB image with a chromagenic filter. The illuminant can then be estimated using chromagenic based illumination estimation algorithms. The method, therefore, does not require two shots and no registration issues involved unlike as in the basic chromagenic filter based color constancy, making it more practical and useful computational color constancy method in many applications.

Conference 8299: Digital Photography VIII

Monday-Tuesday 23-24 January 2012

Part of Proceedings of SPIE Vol. 8299 Digital Photography VIII

8299-01, Session 1

An objective protocol for comparing the noise performance of silver halide film and digital sensor

F. Cao, R. Tessières, G. Frédéric, H. Hornung, DxO Labs (France)

Digital sensors have obviously invaded the market. However, some photographers with very high expectancy still use silver halide film. Are they only nostalgic reluctant to technology or is there more than meets the eye? The answer is not so easy if we remark that, at the end of the golden age, films were actually scanned before development. Nowadays film users have adopted digital technology and scan their film to take advantage from digital processing afterwards. Therefore, it is legitimate to evaluate silver halide film "with a digital eye", with the assumption that processing can be applied as for a digital camera.

The article will describe in details the operations we need to consider the film as a RAW digital sensor. In particular, we have to account for the film characteristic curve, the autocorrelation of the noise (related to film grain) and the sampling of the digital sensor (related to Bayer filter array). We also describe the protocol that was set, from shooting to scanning. We then present and interpret the results of sensor response, signal to noise ratio and dynamic range.

8299-02, Session 1

Sensor defect probability estimation and yield

H. Peng, B. Keelan, Aptina Imaging Corp. (United States)

Sensor yield is directly related to the probability of defective pixel occurrence and the screening criteria. Assuming a spatially independent distribution of single pixel defects, effective on-the-fly correction of single-pixel defects in a color plane, and effective correction of two-pixel defects in a color plane (couplets) through a defect map, sensor yield can be computed based on the occurrence of three adjacent defective pixels in a color plane (triplets). Closed-form equations are derived for calculating the probability of occurrence of couplets and triplets as a function of the probability of a single pixel being defective. If a maximum of one triplet is allowed in a 5-megapixel sensor, to obtain a 98% yield, the probability of a pixel being defective (p) must not exceed 1.3E-3 (6500 defective pixels). For an 8-megapixel sensor, the corresponding requirement would be $p < 1.1E-3$ (8900 defective pixels). Simulation experiments have confirmed the accuracy of the derived equations. These equations can also be applied to compute reliability of any two dimensional spatially-distributed device network.

8299-03, Session 1

Optimum spectral sensitivity functions for single-sensor color imaging

Z. Sadeghipoor, Ecole Polytechnique Fédérale de Lausanne (Switzerland); Y. Lu, Harvard School of Engineering and Applied Sciences (United States); S. Süsstrunk, Ecole Polytechnique Fédérale de Lausanne (Switzerland)

A cost-effective and convenient approach for color imaging is to use a single sensor and mount a color filter array (CFA) in front of it, such that at each spatial position the scene information in one color channel is captured. To estimate the missing colors at each pixel, a demosaicing algorithm is applied to the CFA samples. Besides the filter arrangement and the demosaicing method, the spectral sensitivity functions of the CFA filters considerably affect the quality of the demosaiced image. In this paper, we extend the algorithm proposed by Lu and Vetterli, originally proposed for designing the optimum CFA, to compute the optimum spectral sensitivities. The proposed algorithm solves a constrained optimization problem to find optimum spectral sensitivities and the corresponding linear demosaicing method. An important constraint of this problem is the smoothness of spectral sensitivities, which is imposed by modeling these functions as a linear combination of several smooth kernels. Simulation results verify the effectiveness of the proposed algorithm in finding optimal spectral sensitivity functions that outperform measured camera sensitivity functions.

8299-04, Session 1

A metric for the evaluation of wide dynamic-range cameras

P. W. Wong, Pixim Inc. (United States); Y. H. Lu, Ministry of Public Security (China)

We propose a multi-component metric for the evaluation of digital or video cameras under wide dynamic range (WDR) scenes. The method is based on a single image capture using a specifically designed WDR test chart and light box. Test patterns on the WDR test chart include gray ramps, color patches, arrays of gray patches, white bars, and a relatively dark gray background. The WDR test chart is professionally made using 3 layers of transparencies to produce a contrast ratio of approximately 110 dB for WDR testing. A light box is designed to provide a uniform surface with light level at about 80K to 100K lux, which is typical of a sunny outdoor scene.

From a captured image, 9 image quality component scores are calculated. The components include number of resolvable gray steps, dynamic range, linearity of tone response, grayness of gray ramp, number of distinguishable color patches, smearing resistance, edge contrast, grid clarity, and weighted signal-to-noise ratio. From the 9 component scores, a composite score is calculated that reflects the comprehensive image quality in cameras under WDR scenes. Experimental results have demonstrated that the multi-component metric corresponds very well to subjective evaluation of wide dynamic range behavior of cameras.

8299-05, Session 1

Active pixels of transverse field detector based on a charge preamplifier

G. Langfelder, C. Buffa, A. F. Longoni, A. Pelamatti, F. Zaraga, Politecnico di Milano (Italy)

The Transverse Field Detector (TFD), a filter-less and tunable color sensitive pixel, is based on the generation of specific electric field configurations within a depleted Silicon volume. Like in other tunable sensors, each field configuration determines a set of specific spectral responses that can be used for direct color acquisition. In order to avoid changes of the electric field configuration during the single image capture, a specific active pixel (AP) has been designed. In this AP the dark- and photo-generated charge is not integrated directly on the junction capacitance, but, for each color, it is integrated on the feedback capacitance of a single-transistor charge pre-amplifier. The AP further includes a bias transistor, a reset transistor and a follower.

In this work (i) the design of such a pixel, suitable for a tunable detector, is discussed focusing on several design constraint and in particular on the optimization of the pixel-level loop gain; and (ii) the

experimental results obtained on a small matrix of these active pixels are analyzed in terms of dark current, linearity, reset noise, dynamic range and fixed pattern noise.

8299-06, Session 1

Digital focusing and re-focusing with thin multi-aperture cameras

A. Oberdörster, A. Brückner, F. C. Wippermann, A. Bräuer, Fraunhofer-Institut für Angewandte Optik und Feinmechanik (Germany); H. P. A. Lensch, Univ. Ulm (Germany)

For small camera modules in consumer applications, such as mobile phones or webcams, size and cost are important constraints. An autofocus system increases both size and cost and can also introduce additional aberrations due to misalignment. Therefore, a monolithic optical system with a fixed focus is preferable for these applications. On the other hand, the optical system of the camera has to exhibit a very large depth of field, as it is expected to deliver sharp images for all typical working distances. The depth of field of a camera system can be increased by closing the aperture, but also by reducing focal length.

Multi-aperture systems use multiple optical channels, each of them with a potentially smaller focal length than a comparable single-aperture system. Accordingly, each of the channels has a large depth of field. However, as the channels are displaced laterally, parallax becomes noticeable for close objects. Therefore, the channel images have to be shifted accordingly when recombining them into a complete image.

We demonstrate an algorithm that compensates for parallax as well as chromatic aberration and geometric distortion. We present a very flat camera system that is capable of taking photos or video at a wide range of distances, without moving parts. On the demonstration system, object distance can be adjusted in real time, from 4 mm to infinity. Focus can be selected during capture or after the images were taken.

8299-07, Session 1

The multi-focus plenoptic camera

T. G. Georgiev, Adobe Systems Inc. (United States); A. Lumsdaine, Indiana Univ. (United States) and Adobe Systems Inc. (United States)

The focused plenoptic camera is based on the Lippmann sensor: An array of microlenses focused on the pixels of a conventional image sensor. This device samples the radiance, or plenoptic function, as an array of cameras with extended depth of field, focused at a certain plane in front of the microlenses. For the purpose of digital focusing after the fact (which is one of the important applications), the depth of field needs to be as large as possible. There are fundamental opticalcal limitations to this, related to aperture and diffraction.

A solution of the above problem is to use and array of interleaved microlenses of different focal lengths, focused at two or more different planes. In this way focused microimages would be captured at multiple depths, and a really wide range of digital refocusing would be possible.

This paper presents our theory and results of implementing such camera. Real world images are demonstrating the extended capabilities. Limitations are discussed.

8299-08, Session 1

Spatial domain analysis of discrete plenoptic sampling

A. Lumsdaine, Indiana Univ. (United States); T. G. Georgiev, Adobe Systems Inc. (United States)

Plenoptic cameras are intended to fully capture the light rays in a scene. Using this information, optical elements can be applied to a scene computationally rather than physically---allowing an infinite

variety of pictures to be rendered after the fact from the same plenoptic data. Practical plenoptic cameras necessarily capture discrete samples of the plenoptic function, which together with the overall camera design, can constrain the variety and quality of rendered images. In this paper we specifically analyze the nature of the discrete data that plenoptic cameras capture, in a manner that unifies the traditional and focused plenoptic camera designs. We further present a resolution analysis for plenoptic cameras and develop design guidelines for maximizing resolution. A generalized rendering algorithm is presented that minimizes artifacts resulting from the lower resolution angular sampling that accompanies high-resolution spatial sampling. Experimental results using a real-time GPU implementation of our algorithms demonstrates the effectiveness of our approach.

8299-09, Session 1

Design framework for a spectral mask for a plenoptic camera

K. Berkner, S. A. Shroff, Ricoh Innovations, Inc. (United States)

Plenoptic cameras are designed to capture different combinations of light rays from a scene, sampling the lighfield of a scene. Most of these camera designs capture directional ray information enable applications such as digital refocusing, rotation, or depth estimation only few address capturing spectral information of the scene. It has been demonstrated that by modifying a plenoptic camera with a filter array containing different spectral filters inserted in the pupil plane of the main lens, sampling of in the spectral dimension of the plenoptic function is performed. As a result, the plenoptic camera is turned into a single-snapshot multispectral imaging system that trades-off spatial with spectral information captured with a single sensor. Little work has been performed so far on analyzing diffraction effects and aberrations of the optical system on the performance of the spectral imager. In this paper we demonstrate simulation of a spectrally-coded plenoptic camera optical system, evaluate quality of the spectral measurements captured at the detector plane, and demonstrate opportunities for optimization of the spectral mask for a few sample applications.

8299-10, Session 2

Detection thresholds of structured noise in the presence of shot noise

F. Li, B. W. Keelan, A. Dokoutchaev, R. Jenkin, Aptina Imaging Corp. (United States)

As a growing number of consumers become aware of the tradeoff between mega pixel count and camera performance, image quality, especially at low light weights more when a user purchases a camera. Structured noise, namely row or column fixed pattern noise (FPN), is one of the main challenges in improving image quality for imaging sensors. The present paper studied the detectability of the structured noise at various noise levels. The results may be used to guide the pixel and algorithm designs for improving structured noise in imaging sensors.

8299-11, Session 2

Reduced-reference image quality assessment based on statistics of edge patterns

Y. Chen, X. Mou, Xi'an Jiaotong Univ. (China)

Recently, research of Objective Image Quality Assessment (IQA) has gained much attention due to its wide application prospect. Among them, the Reduced-Reference (RR) methods estimate perceptual quality of distorted images with partial information from the reference images. This paper proposes a novel universal RR-IQA metric based on the statistics of edge patterns. Firstly, the binary edge maps of the reference and distorted images are created by the LOG operator and zero-crossing detection. Based on them, 15 groups of typical edge

patterns are extracted and then their statistical distributions calculated respectively for the reference and distortion images. The proposed RR-IQA metric is achieved by computing the L-1 Minkowski distance between those two distributions. We have evaluated this metric on six publicly accessible subjective IQA databases. Experiments show that the proposed metric works better than its competing RR-IQA metrics across the different databases.

8299-12, Session 2

Joint chromatic aberration correction and demosaicking

T. Singh, M. Singh, Image Algorithmics (United States)

Chromatic Aberration of lenses is becoming increasingly visible with the rise of sensor resolution, and methods to algorithmically correct it are becoming increasingly common in commercial systems. A popular algorithm is to undo the geometric distortions after demosaicking. Since most demosaickers require high frequency correlation of primary colors to work effectively, the result is artifact-ridden as Chromatic Aberration destroys this correlation. The other existing approach of undistorting primary color images before demosaicking requires resampling of sub-sampled primary color images and is prone to aliasing. Furthermore, this algorithm cannot be applied to panchromatic CFAs. We propose a joint demosaicking and chromatic aberration correction algorithm that is applicable to both panchromatic and primary color CFAs and suffers from none of the above problems. Our algorithm treats the mosaicing process as a linear transform that is invertible if luminance and chrominance are appropriately bandlimited. We develop and incorporate Chromatic Aberration corrections to this model of the mosaicing process without altering its linearity or invertibility. This correction works for both space variant linear filter demosaicking and the more aggressive compressive sensing reconstruction.

8299-13, Session 2

Optimal defocus estimates from individual images for autofocusing a digital camera

J. Burge, W. S. Geisler, The Univ. of Texas at Austin (United States)

Recently, we developed a method for optimally estimating defocus blur given a set of natural scenes, a wave-optics model of the lens, a sensor array, and a specification of measurement noise. The method is based on first principles and can be tailored to any vision system for which these properties can be characterized. Here, the method is used to estimate defocus in local areas of images (64x64 pixels) formed in a Nikon D700 digital camera fitted with a 50mm Sigma prime lens. Performance is excellent. Defocus magnitude and sign can be estimated with high precision and accuracy over a wide range. The method offers an alternative to both phase detection and contrast measurement auto-focus techniques. Like phase-detection, the method provides point estimates of defocus (magnitude and sign), yet it does not require specialized hardware. Like contrast measurement, the method is image-based and can operate in "Live View" mode, yet it does not require an iterative search for best focus. Thus, this new method has the advantages of both phase-detection and contrast measurement auto-focus techniques, without their disadvantages. The approach can be used to develop improved autofocus and depth estimation algorithms for computational vision systems.

8299-14, Session 2

Quality versus color saturation and noise

B. W. Keelan, R. B. Jenkin, E. W. Jin, Aptina Imaging Corp. (United States)

A softcopy quality ruler study involving 12 scenes and 34 observers was performed to quantify the dependence of quality on color saturation, in the absence of noise, with saturation measured using Imatest software. Quality falls off symmetrically with deviation of color saturation from the preferred value of about 110%, with a 20% change in saturation reducing quality by about two just noticeable differences (JNDs). Optimization of noise versus color saturation was investigated using (1) the aforementioned transform of color saturation to JNDs of quality; (2) a previously published objective metric and JND transform for isotropic noise; and (3) the multivariate formalism, for combining JNDs from independent attributes into a overall quality JNDs. As noise increases and signal to noise ratio (SNR) decreases, the optimal color saturation decreases from the 110% position, so that there is less noise amplification by the color correction matrix. A quality contour plot is presented, showing a region of plausible color saturation values, as a function of SNR, for a representative use case. One example of a reasonable strategy is to provide 80% color saturation at SNR = 5, 90% at SNR = 10, 100% at SNR = 20, and 110% at SNR 50.

8299-15, Session 2

Bio-inspired framework for automatic image quality enhancement

F. Gasparini, S. Raimondo, C. Andrea, M. Fabrizio, Univ. degli Studi di Milano-Bicocca (Italy)

We propose a bio-inspired framework for automatic image quality enhancement. Restoration algorithms usually have fixed parameters whose values are not easily settable. We show that it is possible to correlate no-reference visual quality values to specific parameter settings such that the quality of an image could be effectively enhanced through the restoration algorithm. Given a generic degraded image a no-reference metric evaluates its visual quality. This value is correlated to a tuple of optimized algorithm parameters determined a priori and gathered in a table. This table is the outcome of an offline processing. For a given artifact, this processing involves a restoring algorithm, a full-reference metric and a database of reference images that have been degraded with a wide range of distortion. Each degraded image is processed by the restoration algorithm that, by means of the given parameters produces an enhanced version of the image. This, along with its original reference, is evaluated by the full-reference quality metric. These restoration and evaluation operations are repeated within a genetic optimization algorithm. When the evolution is over, the best so-found parameters are associated to a no-reference quality measure of the initial degraded image. In this paper, we chose JPEG blockiness distortion as a case study. As for the restoration algorithm, we used either a bilateral filter, or a total variation denoising detexturer. The experimental results on the LIVE database will be reported. These results will demonstrate that a better visual quality is achieved through the optimized parameters over the entire range of compression, with respect to the algorithm default parameters.

8299-16, Session 2

An efficient, multiple-exposure image fusion in JPEG domain

R. S. M. Hebbalaguppe, Dublin City Univ. (Ireland); R. Kakarala, Nanyang Technological Univ. (Singapore)

An efficient method to fuse multiple images taken with varying exposure times in the JPEG domain is presented. The algorithm uses the spatial frequency analysis provided by the DCT within JPEG to combine the uniform regions of the longest-exposure image with the detailed regions of the short-exposure images, thereby reducing noise while providing sharp details. Advantages of the proposed method are great reduction in processing time, improved memory management, and efficient ghost removal in obtaining reasonably good quality HDR images.

Experiments show both quantitative and qualitative improvement over the short-long exposed images. Qualitatively, the fused image looks sharp with better colors than the long-short images. Quantitatively, the fused image shows improvement in SNR over the shortest exposed image and the sharpness (obtained by blur metric) over the longest exposed image.

To summarize our method, we use a single pass sigmoidal boosting on the shorter exposed images implemented as LUT, unlike other published methods which require two or more passes. Reuse of edge detection which is a part of JPEG for removal of artifacts further optimizes the algorithm. Lastly, the method requires no more than a single macro block to be kept in memory, because the image fusion is performed essentially in the JPEG file and rendered only on decoding the image.

8299-17, Session 2

A controllable anti-aliasing filter for digital film cameras

B. Petljanski, Panavision Inc. (United States)

In this paper, the theoretical foundation and practical implementation of a controllable anti-aliasing filter for digital film cameras is presented.

A prototype of an optical anti-aliasing filter that is based on moving a parallel optical window was designed and built to demonstrate the ability to control the spatial frequency response of an acquisition system. During the image exposure, four spring preloaded voice coils are rapidly changing the pitch and yaw of the parallel window resulting in a displacement of the image content that is projected onto the sensor. The image content displacement during the exposure results in altering the frequency response of the scene that is captured by the sensor.

Specifically, during the exposure time, a carefully controlled movement of the parallel optical window results in a circular trajectory of the image content that is projected onto the sensor. By increasing or decreasing the radius of the circular trajectory, the spatial cut-off frequency of the system is dynamically modified. In addition to the circular path, this paper shows theoretical justification and demonstrates the use of more complex trajectories, such as the double circle, elliptical, rose and others. These trajectories improve the suppression of aliased components in the acquired image.

8299-01, Session 3

Development of the I3A CPIQ spatial metrics

H. Eliasson, Sony Ericsson Mobile Communications AB (Sweden); D. Baxter, STMicroelectronics (R&D) Ltd. (United Kingdom); F. Cao, DxO Labs (France); J. Phillips, Eastman Kodak Co. (United States)

The I3A Camera Phone Image Quality (CPIQ) initiative is aiming at providing a consumer-oriented overall image quality metric for mobile phone cameras. In order to achieve this goal, a set of subjectively correlated image quality metrics has been developed. This paper describes the development of a specific group within this set of metrics, the spatial metrics. Contained in this group are the edge acutance, visual noise and texture acutance metrics. A common feature is that they are all dependent on the spatial content of the specific scene being analyzed. Therefore, the measurement results of the metrics are weighted by a contrast sensitivity function (CSF) and, thus, the conditions under which a particular image is viewed must be specified. This leads to the establishment of a common framework consisting of three components shared by all spatial metrics. First, the RGB image is transformed to a color opponent space, separating the luminance channel from two chrominance channels. Second, associated with this color space are three contrast sensitivity functions for each individual opponent channel. Finally, the specific viewing conditions, comprising both digital displays as well as printouts, are supported through two distinct MTFs.

8299-18, Session 3

A functional-design approach to lens shading correction issues on mobile camera system

S. Yoo, SAMSUNG Electronics Co., Ltd. (Korea, Republic of); T. Kim, SAMSUNG Electronics Co., Ltd. (United States)

This paper proposes the lens shading correction system through combination of functional factors based on recent issues related to mobile camera system. The main issue is designing of the shading algorithm. The second issue is correction of color temperature. The third issue is correction by optical zoom lens. The fourth issue is related with mass production of camera module. The final issue in mobile devices is reducing size of camera module. Previous lens shading correction that situate to stable function through rich hardware resource in DSLR and compact camera for recently the several years is becoming issues again being admitted to part of system of mobile camera that resource is limited. As days go by, size reduction of image sensor and module is essential due to request of marketplace along with the development of mobile devices. The proposed lens shading correction system provides minimizing of accuracy errors in shading evaluation, as well as size reduction of memory and module in hardware. Finally, the proposed system for lens shading correction solves raised issues in mobile environment and guides design of algorithm and hardware for lens shading correction in mobile camera module.

8299-19, Session 3

Rethinking user interfaces for cameraphones

S. A. Brewster, C. McAdam, Univ. of Glasgow (United Kingdom)

Cameraphones are very common but have usability issues that affect their use and reduce the quality of images captured. Users often spend little time post-processing photos, often uploading them immediately to social networking sites. Problems may occur because users look through the LCD to frame shots and often miss icons displayed around the edges of the screen that present important information about camera status. This may lead to shots being missed, blurred or poorly exposed. Most cameraphones do not take full advantage of the features of the underlying phone platform (such as rich sensing and powerful audio) to try to solve these problems. We created an Android cameraphone application that featured novel interactions and made use of the features of the platform to provide a rich variety of information in more usable forms, such as: sonification of the luminance histogram to ensure better exposure before a picture was taken; measuring phone movement to ensure the phone was being held steady; and the detection of motion in the image to support panning. We evaluated these interactions with users in real settings and showed they could solve some of the most common cameraphone problems and improve the picture taking process.

8299-02, Session 4

Calibration and adaptation of ISO visual noise for camera phone image quality assessment

D. J. Baxter, A. Murray, STMicroelectronics (R&D) Ltd. (United Kingdom)

The I3A Camera Phone Image Quality (CPIQ) visual noise metric described is a core image quality attribute of the wider I3A CPIQ consumer orientated, camera image quality score. This paper describes the motivation behind the choice of a visual noise metric, why the characteristics of cell phone cameras necessitate the adaptation of the ISO 15739 visual noise protocol and the calibration of the adapted visual noise metric in terms of Quality Loss JNDs. Visual noise metrics such as ISO 15739 visual noise, S-CIELAB and vSNR are shown to have the important property of being able to discriminate between different NPS shapes. The optical non-uniformities prevalent in cell phone cameras and higher noise levels pose significant challenges to the ISO 15739 visual noise protocol.

The non- uniformities are addressed using a frequency based high pass filter. Secondly, the data clipping at high noise levels is avoided by use of the Johnson and Fairchild frequency based SCIELAB

Luminance CSF. The final result is a visually based noise metric calibrated in terms of Quality Loss JNDs using Aptina Imaging's subjectively calibrated image set.

Conference 8299

8299-03, Session 4

An objective method for evaluating the texture-preserving capability of digital camcorders

K. Zhu, S. Li, D. Saupe, Univ. Konstanz (Germany)

This paper presents a method for evaluating performance of camcorders in terms of texture preservation, taking human visual system (HVS) into account. A metric we call PSDM (power spectral density metric) is the outcome of the method as a quantitative indicator. The relationships between the PSDM value and the motion speed and the compression bitrate were studies. The dead leaves chart was used in our experiments to simulate a scene with textures of different scales. The dead leaves charts is known as a good target invariant to scale, shift, rotation and contrast (exposure) and its radial PSD follows a power law. Experimental results on five camcorders from three different vendors have shown that 1) the PSDM value has a monotonic relationship with the motion speed for all tested camcorders; 2) the PSDM value has a monotonic relationship with the compression bitrates for three tested camcorders but not for the other two. We attribute the mismatch between the PSDM value and the compression bitrate for some camcorders to the intrinsic nonlinearity of many components in the whole video capturing/recording process. Our study has confirmed PSDM as a useful indicator for measuring a camcorder's performance in terms of preserving textures.

8299-04, Session 4

Improving texture loss measurement: spatial frequency response based on a colored target

U. Artmann, D. Wueller, Image Engineering GmbH & Co. KG (Germany)

The pixel race in the digital camera industry and for mobile phone imaging modules have made noise reduction to a significant part in the signal processing. Depending on the used algorithms and the underlying amount of noise that has to be removed, noise reduction leads to a loss of low contrast fine details, also know as texture loss. The description of these effects became an important part of the objective image quality evaluation in the last years, as the established methods for noise and resolution measurement fail to do so. Different methods have been developed and presented in the last years, but could not fully satisfy the requested stability and correlation with subjective tests. In our paper, we present our experience with the current approaches for texture loss measurement. We have found a critical issue within these methods: the used targets are color neutral. We could show that the test-lab result do not match the real live experience with the cameras under test. We present an approach using a colored target and our experience with this method, based on a huge variety of digital cameras.

8299-20, Session 4

On the performances of computer vision algorithms on mobile platforms

S. Battiato, G. M. Farinella, E. Messina, G. Puglisi, D. Ravì, Univ. degli Studi di Catania (Italy); A. Capra, V. Tomaselli, STMicroelectronics (Italy)

In recent years there is a growing interest in new technology to be employed in the context of mobile devices. Despite today's mobile devices (e.g., smartphone, tablet, etc.) are still limited in terms of resources (e.g., processor speed, available RAM, etc.), novel Computational Photography solutions are available to build appealing imaging applications that cannot be performed before. The main idea is to overcome the limitation of traditional imaging devices by using computational methods which can exploit the different inputs offered by a mobile devices (e.g., from low level data, such as Bayern pattern,

GPS position, etc.). Since different cameras are usually embedded in devices of new generation, computer vision algorithms will be extremely useful in many applications of next future. For example, visual tracking can be exploited to interact with video games, or the recognition of the visual content could help in building new applications in the context of cultural heritage (e.g., giving back information on a recognized archeological site).

The main contribution of this work is related to the porting and testing of some classic computer vision algorithms on mobile platforms. Specifically, a few algorithms covering the main tasks of Computer Vision have been considered: keypoint extraction, face detection, image segmentation. The porting has been performed considering the following operating systems: Maemo, typically used in Nokia N900, and Android employed in LG Optimus one, Samsung Galaxy SII. These operating systems have been considered because they can be easily extended with customized libraries and/or programs and provide a standardized and fairly widespread API (Application Program Interface).

It is worth noting that the aforementioned algorithms should be optimized to properly work on low resources devices. For instance, the FCAM library available for N900 Nokia smart phone, allows to interact with the low level algorithms (e.g., demosaicing, white balancing, denoising, etc.) and data (Bayer pattern) involved in the imaging pipeline. In this way a better design of computer vision algorithms for constrained resources devices can be done.

Finally, comparative tests, conducted on standard datasets, quantitatively and qualitatively evaluate the performance of the algorithms on mobile devices.

8299-22, Session 5

Spectral sensitivity evaluation considering color constancy

H. Kuniba, Nikon Corp. (Japan)

The human visual system automatically adjusts the white point and the color of an object remains almost constant. This is called "color constancy." Thus digital cameras should adjust the white point according to the illuminant used (white balance correction). Hung (2001) evaluated color constancy error of several spectral sensitivities and showed that the color constancy error was reduced if the white balance correction was conducted in an appropriate color space. Three types of sensor models were used for the simulation, then the color constancy errors seems to be large for the sensor model whose overlaps between channels are large and channel widths are large. Thus two new metric which evaluated the overlaps and the widths were defined and used to optimize the conversion and the color constancy errors were evaluated. Though the reduction rate was not as large as the former results, the color constancy errors was reduced by reducing the overlap between channels and narrowing the widths. The narrower widths and smaller overlaps of RGB channels gave less accurate colorimetric reproduction and less noisy images. In addition to this, it gave smaller color constancy errors. In view of consumer digital cameras, colorimetric reproduction accuracy is not the first priority because the objective of their reproduction is "preferred colour reproduction." They are used under wide spectrum of illuminations and scenes but users just want to press the button and let the camera do the rest. Thus the narrower widths and smaller overlaps of RGB channels is suitable because it can give less noisy images and consistent color reproduction with simple white balance processing.

8299-23, Session 5

Multispectral demosaicking using guided filter

Y. Monno, M. Tanaka, M. Okutomi, Tokyo Institute of Technology (Japan)

A multispectral imaging is highly demanded for precise color reproduction and for various computer vision applications. Unfortunately, the multispectral imaging system is still impractical compared to commonly used consumer RGB cameras. The

multispectral imaging with a multispectral color filter array (MCFA), in which more than three spectral component are arrayed, could be a simple, low-cost, and practical system. A challenge of the multispectral imaging system with the MCFA is a multispectral demosaicking because each spectral component of the MCFA is severely undersampled. In this paper, we propose a novel multispectral demosaicking algorithm using a guided filter. The guided filter is recently proposed as an excellent structure-preserving filter. The guided filter requires so-called a guide image. The output of the guided filter can be represented as the linear transform of the guide image, thus it nicely resembles the guide image. In the proposed algorithm, we first generate a guide image from the most densely sampled spectral component in the MCFA. Then, ohter spectral components are interpolated with the guided filter. Experimental comparisons demonstrate that the proposed algorithm outperforms other existing multispectral demosaicking algorithms both visually and quantitatively.

8299-24, Session 5

An LED-based lighting system for acquiring multispectral scenes

M. Parmar, S. Lansel, J. E. Farrell, Stanford Univ. (United States)

Multispectral scene information is useful for the simulation of the elements of the digital imaging pipeline and to analyze their effect on final image quality 1. Certain scenes with animate subjects, e.g., humans, pets, etc., are of particular interest to consumer camera manufacturers because of their ubiquity in common images, and the importance of maintaining colorimetric fidelity for skin. Typical multispectral acquisition methods rely on techniques that use multiple acquisitions of a scene with a number of different optical filters2-6 or illuminants7. Such schemes require long acquisition times and are best suited for still scenes. In scenes where animate objects are present, movement leads to problems with registration and methods with shorter acquisition times are needed.

To address the need for shorter image acquisition times, we developed a multispectral imaging system that captures multiple acquisitions during a rapid sequence of differently colored LED lights. In this paper, we describe the design of the LED-based lighting system and report results of our experiments capturing scenes with human subjects.

8299-25, Poster Session

Fast in-plane translation and rotation estimation for multi-image registration

X. Jiang, H. Wang, Qualcomm Inc. (United States)

This document considers the planar motions of camera, that is, the rotation, and horizontal and vertical translations. The approach based on projection including both Cartesian coordinate system and polar coordinate system can estimate the three parameters comparably quickly with simple calcuation. The potential applications cover motion deblurring, noise reduction, super-resolution, image fusion, high dyanmic range image processing, EDOF, 3D imaging or those techniques which require global or local registration.

8299-26, Poster Session

Multispectral filter wheel cameras: modeling aberrations for filters in front of lens

J. Klein, T. Aach, RWTH Aachen (Germany)

Aberrations occur in multispectral cameras featuring filter wheels because of color filters with different optical properties being present in the ray path. In order to ensure an exact compensation of these aberrations, a mathematical model of the distortions has to be developed and its parameters have to be calculated using the measured data. Such a model already exists for optical filters placed between the sensor and the lens, but not for bandpass filters placed in front of the lens. For this configuration, the rays are first distorted by the filters and then by the lens. In this paper, we derive a model for aberrations caused by filters placed in front of the lens in multispectral cameras. We compare this model with distortions obtained with simulations as well as with distortions measured during real multispectral acquisitions. In both cases, the difference between modeled and measured aberrations remains low, which corroborates the physical model. Multispectral acquisitions with filters placed between the sensor and the lens or in front of the lens are compared: the latter exhibit smaller distortions and the aberrations in both images can be compensated using the same algorithm.

8299-27, Poster Session

Correcting saturated pixels in images

J. Fu, W. Ji, X. Mou, Xi'an Jiaotong Univ. (China)

This paper proposes a novel method to correct saturated pixels in images. This method is based on the YCbCr color space and separately corrects the chrominance and the luminance of saturated pixels. Dynamic thresholds are adopted to identify saturated pixels, i.e. the thresholds for different images and different color channels are different. So our method can correct not only RAW images but also processed images. Once the saturated pixels are identified, there are three kinds of saturated pixels: 1-channel saturated pixels, 2-channel saturated pixels and 3-channel saturated pixels. They are denoted as $\Omega1$, $\Omega2$ and $\Omega3$ respectively. Different strategies are implemented to these three kinds of regions. The color of saturated pixels in $\Omega1$ is corrected according to their original color and the color of their neighborhood. And the color of saturated pixels in $\Omega2$ and $\Omega3$ is corrected only according to the color of their neighborhood. The luminance of saturated pixels is corrected using the model proposed in this paper. Experiment results show that our method is effective in correcting saturated pixels of RAW images and processed images.

8299-28, Poster Session

Real-time, multi-directional 2D fast wavelet transform and its denoised sharpening application

B. Baek, T. Kim, SAMSUNG Electronics Co., Ltd. (Korea, Republic of)

The new 2D fast wavelet transform and real-time wavelet approach have been presented through the paper. The proposed wavelet transform includes multi-directional decomposition/reconstruction concept designed to show better performance near diagonal edges than two-direction based fast wavelet. The real-time wavelet architecture for hardware cost-effective is also presented and shows blocking artifact-free performance. Using a small size of wavelet gives rise to the use of a significantly reduced amount of line memory and much lower latency and using overlapped windows instead of exclusive macro block windows results in avoiding blocking artifact that easily happens in macro block based transform. This architecture corresponds well to the raster scan order image sensors as well. In addition, the concept of power-scalable method is suggested in brief as a way to control the tradeoff between power consumption and blocking-artifact-free performance. Experiments show the result that the proposed wavelet and real-time approach has much lower blocking and directional artifacts. As a result, the proposed wavelet and real-time system implementation approach make the use of a wavelet more feasible for real-time application with much less artifacts by significant reduction of hardware resources. This enables the new transform to have many potential application areas.

8299-29, Poster Session

Color transfer using semantic image annotation

F. Gasparini, R. Schettini, C. Cusano, Univ. degli Studi di Milano-Bicocca (Italy)

In this work we present an automatic local color transfer method based on semantic image annotation. With this annotation, images are segmented into homogeneous regions, assigned to seven different classes (vegetation, snow, water, ground, street, and sand). Our method permits to automatically transfer the color distribution from regions of the source and target images annotated with the same class (for example the class "sky"). It is also possible to select the color distributions corresponding to regions of different classes, starting from a set of annotated images (for example the class "sky" from one annotated source image and the class "vegetation" from a different source image). These color distributions will then be transferred to the corresponding regions of the target image.

In the final paper we will provide more details of the developed method and further examples.

8299-30, Poster Session

Adaptive directional color image sharpening with overshoot control

T. Tajbakhsh, Technische Univ. Hamburg-Harburg (Germany)

The paper presents an algorithm for contrast enhancement of noisy color images based on adaptive unsharp masking. Color images are processed by transforming the gamma encoded signal to an opponent color space and applying edge sharpening to the luminance signal. The adaptive unsharp masking algorithm fusions five key ideas: the adaptive Wiener filter as proposed by Lee to adapt the filtering strength locally, its directional refinement, the overshoot control as proposed by Bruna et al., the regulation of the amount of contrast enhancement based on the edge strength as proposed by Polesel et al., and the temporal noise estimation technique proposed by Foi et al. The results are edge emphasized images without unwanted overshoot artifacts and virtually no noise amplification in smooth areas and reduced noise amplification at edge transitions.

Conference 8300: Image Processing: Machine Vision Applications V

Wednesday 25 January 2012

Part of Proceedings of SPIE Vol. 8300 Image Processing: Machine Vision Applications V

8300-16, Poster Session

Efficient local approximation of perceptual color differences for color inspection

R. Huber-Mörk, Austrian Institute of Technology (Austria)

We suggest a local approximation of perceptual color differences in a device dependent color space, e.g. the RGB space. The approximation is efficiently computed from measuring Euclidean color distance in the device dependent color space combined by with an associate memory data structure. Established measures of color difference are considered. The suggested approach is suited for industrial color inspection where small tolerance levels for color differences are known in advance. No on-line color space transformation or expensive evaluation of advanced color metrics is necessary. The computation time is independent from the used metric. Results for approximation of small perceptual color differences in a color inspection setup are given.

8300-17, Poster Session

Modified fuzzy c-means applied to a Bragg-grating-based spectral imager for material clustering

A. Rodríguez, J. L. Nieves, E. Valero, Univ. de Granada (Spain); E. Garrote, TECNALIA (Spain); J. Hernández-Andrés, J. Romero, Univ. de Granada (Spain)

An adaptation of fuzzy c-means algorithm has been used for segmentation of hyperspectral urban images provided by a Bragg-based spectral imager.

Classical fuzzy c-means algorithm uses Euclidean distance for computing sample membership to each cluster. We have introduced a different distance metric, Spectral Similarity Value (SSV), in order to have a more convenient similarity measure for reflectance information. SSV distance metric considers both magnitude difference (by the use of Euclidean distance) and spectral shape (by the use of Pearson correlation). The combination of these terms makes it a more convenient metric for creating spectrally similar clusters.

Two sets of images have been used to test the algorithm performance. First set was a collection of hyperspectral urban images acquired by a Bragg grating based imager. Results using these images show that clusters are spectrally more homogeneous than clusters created with the original fuzzy c-means algorithm.

The second experiment was carried out with a set of 100 hyperspectral synthetic images. The proposed method improves the segmentation performance and assigns highest membership value to pixels corresponding to highest purity reflectance.

8300-18, Poster Session

Robust recognition of 1D barcodes using hough transform

J. Dwinell, L. X. Bian, SICK, Inc. (United States); P. Bian, Microsoft Corp. (China)

Automated package handling systems require a robust solution for decoding barcodes. Typical applications process several thousand packages per hour at high speeds. Throughputs are typically 2 to 3 packages per second. There is only one opportunity to acquire a package image and many factors contribute to poor quality of the barcode.

An efficient and powerful barcode decode algorithm is needed to analyze low quality or damaged codes and is able to handle the high throughput. A multi-level algorithm is shown that meets this challenge. At the low level, initial analysis is performed in the gray-scale domain. The middle level introduces a Hough transform domain. The traditional Hough transform is optimized for barcode decode providing a very fast yet powerful algorithm for badly damaged regions of the barcode. Finally, a top level guides the lower level processing utilizing various global information.

The traditional Hough transform for a straight line is:

$$\rho = x \cos \theta + y \sin \theta$$

Calculating in entire Hough plane is computationally intensive. The parameters in the Hough domain are the angle and offset for each point. The optimization allows the Hough transform to be only the offset. The offset is critical as each line influences the corresponding bar and space width determination. An optimized Hough transform provides a robust algorithm for determining the offset of each line when working with noisy data.

8300-19, Poster Session

Estimating the coordinates of pillars and posts in the parking lots for intelligent parking assist system

J. Choi, J. G. Kuk, N. I. Cho, Seoul National Univ. (Korea, Republic of)

This paper proposes a vision algorithm for the detection of pillars or posts in the video captured by a single front view camera implemented on the fore side of a room mirror in a car. The main purpose of this algorithm is to complement the weakness of current ultrasonic IPAS, which does not well find the exact position or pillars or cannot recognize narrow posts. The proposed algorithm is consisted of three steps: straight line detection, line tracking and estimation of 3D position of pillars. In the first step, we find the strongest line based Hough transform. Second step is the combination of detecting and tracking, and the third is the calculation of 3D position of the line by the analysis of trajectory of relative positions and the parameters of camera. Experiments on synthetic and real images show that the proposed method successfully locate and track the position of pillars, which helps the ultrasonic system to correctly locate the edges of pillars. It is believed that the proposed algorithm can also be employed as a basic element for vision based autonomous driving system.

8300-21, Poster Session

Recognizing human gestures using a novel SVM tree

H. Jain, A. Chatterjee, S. Kumar, B. Raman, Indian Institute of Technology Roorkee (India)

In this paper, a novel support vector machine (SVM) tree is proposed for gesture recognition in a large environment having complex background. A skeleton based strategy is adopted to extract the features from a video sequence representing any human gesture. The background image for a large scene is obtained by using mosaic of several images captured by a pan-tilt-zoom (PTZ) camera with different orientations. This mosaic image is used to compute the background frame in real time for any arbitrary pan and tilt setting. In our binary tree implementation of SVM, the number of binary classifiers required is reduced since, instead of grouping different classes together in order to train a global classifier, we select two classes for training at every node of the tree and use probability theory to classify the remaining points based on their similarities and differences to the two classes used for training. This process is carried on, randomly selecting two classes

for training at a node, thus creating two child nodes and subsequently assigning the classes to the nodes derived. In the classification phase, we start out at the root node. At each node of the tree, a binary decision is made regarding the assignment of the input data point to either of the group represented by the left and right sub-tree of the node which may contain multiple classes. This is repeated recursively downward until we reach a leaf node that represents the class to which the input data point belongs to. Finally, the proposed framework is tested on various data sets to check its efficiency. Encouraging results are achieved in terms of classification accuracy.

8300-22, Poster Session

Fabric defect detection using the wavelet transform in an ARM processor

J. A. Fernandez Gallego, Antonio Nariño Univ. (Colombia); S. A. Orjuela Vargas, Univ. Gent (Belgium); J. Alvarez, Antonio Nariño Univ. (Colombia); W. Philips, Univ. Gent (Belgium)

Nowadays, advances in technology offers portable devices for our day life with entertainment and communication applications as well as customizable capabilities. The system architectures of these devices are a powerful tool to develop friendly visual environments for industrial applications such as inspection, control or monitoring tasks. Furthermore, these systems operate in reduced working spaces and offer portability capabilities that are useful for visual inspection tasks among others.

Currently, platforms with direct access to the instructions set of the processor are preferred for real time applications because they permit a fast data processing.

We present in this paper an example of the use of an embedded system, the Zeus Epic 520 single board computer, for an image processing application in defect detection by implementing the wavelet transform using the Embbeded Visual C++ 4.0 compiler for Windows CE 5.

The main difficulty in implementing the algorithm was the fact that the ARM processor use Fixed Point Arithmetic calculations. The algorithm was tested for defect detection using images of fabrics with six types of defects. An average of 80% in terms of correct defect detection was obtained, achieving a similar performance than using processors with float point arithmetic calculations.

8300-23, Poster Session

Orthophotoplan segmentation based on regions merging for roof detection

Y. El Merabet, C. Meurie, Y. Ruichek, Univ. de Technologie de Belfort-Montbéliard (France); A. Sbihi, Ecole Nationale des Sciences Appliquées de Tanger (Morocco); R. Touahni, Univ. Ibn Tofail (Morocco)

In this paper, we propose a strategy of regions merging for roof detection which is made on pre-segmentation results. It is based on a 2D modeling of the roof ridges and region features. The preliminary segmentation is obtained by the watershed algorithm with an optimal colorimetric invariant and color gradient. The choice of an appropriate couple invariant/gradient permits to limit illuminations changes (shadows, brightness, etc) present on several roofs and increases the segmentation results. The watershed algorithm offers satisfactory results but produces an over-segmentation due to many germs (ie. local minima). This effect is reduced by using an appropriate selection of germs but can also be improved with a post-treatment based on regions merging. The proposed merging criteria is based on the 2D modeling of roof ridges (number of segments modeling the common boundary between two regions candidates to the fusion) and on the region features (contrast on boundary of two common regions, average color of region). The proposed strategy is evaluated on 100 real roof images with the Vinet criteria using a ground truth in order to demonstrate the effectiveness and the reliability of the proposed approach.

8300-24, Poster Session

Motion cue analysis for Parkinsonian gait recognition

T. M. Khan, J. Westin, M. Dougherty, Dalarna Univ. (Sweden)

This paper presents a vision-based marker-free method for gait impairment detection in the patients with Parkinson's disease (PWP). The system is based on the idea that a normal human body attains equilibrium during the gait by aligning the body posture with the axis of gravity using feet as the base of support. Due to the rigidity in muscular tone, the PWP fail to align their body with the axis of gravity. The leaned posture of PWP appears to fall forward. Whereas a normal gait exhibits a constant erect posture throughout the gait. PWP walk with shortened stride angle (less than 25 degrees on average) between the legs with high variability between the stride cycles. A normal gait exhibits periodic stride cycles with stride angle of 45 degrees on average between the legs. In order to analyze Parkinsonian Gait (PG), the subjects were videotaped for several gait cycles. The subject's body is segmented in each frame of the gait video using color segmentation method based on thresholding to form a silhouette. The centre of gravity of this silhouette is calculated. The silhouette is skeletonized for the motion cues extraction. Two motion cues analyzed were stride cycles (based on the cyclic leg motion of the skeleton) and the posture lean (based on the angle between the leaned torso of the skeleton and the axis of gravity). High variations were recorded in the stride intervals of PG whereas constant stride intervals were recorded in case of normal gait. Cosine Similarity measurements between an imaginary perfect gait pattern and the subject's gait pattern produced 100% recognition rate of PG for 7 subjects. Results suggested that the method is a promising tool to be used for the remote assessment of gait by the clinicians.

8300-01, Session 1

An introduction to omnidirectional vision: theory and applications

D. Fofi, S. Mosaddegh, A. E. R. Shabayek, O. Morel, Univ. de Bourgogne (France)

This paper presents an overview of the omnidirectional vision techniques and their potential applications in different fields of machine vision: robotics, industrial inspection, medical imaging, videosurveillance and cultural heritage. After a brief introduction, a taxonomy of the omnidirectional cameras is proposed (special lenses, polydioptric, catadioptric, central or non-central) with the scientific challenges they give rise to. A list of applications that can be found in the literature is also reported.

This paper does not pretend to propose new material in the field of omnidirectional vision, but to gather the main results in a comprehensive overview, to highlight the advantages and drawbacks, and to suggest the use of omnidirectional sensors for machine vision (and related) applications.

8300-02, Session 1

Sensor placement optimization in buildings

S. Bianco, R. Schettini, F. Tisato, Univ. degli Studi di Milano-Bicocca (Italy)

In this work we address the problem of optimal sensor placement for a given region and task. The problem of sensor placement arises in several areas of computer vision, computer graphics, and robotics, such as localization, tracking, surveillance, object or environment reconstruction, and image based rendering.

An important issue in designing sensor arrays is the appropriate placement of the sensors such that they achieve a predefined goal.

We consider four different problems: maximizing coverage subject to a given number of sensors (a) or a maximum total price of the sensor array (b), optimizing sensor poses given fixed locations (c),

and minimizing the cost of a sensor array given a minimally required percentage of coverage (d).

To solve the four problems considered, we propose two different optimization-based algorithms: the former for discrete problem spaces, the latter for continuous ones.

The different algorithms proposed are experimentally evaluated and compared to state of the art algorithms for camera placement and experimental results are presented. The results show that the algorithms work well and are suited for different practical applications.

8300-03, Session 1

Optical feature extraction with illumination-encoded linear functions

R. Gruna, Karlsruher Institut für Technologie (Germany); J. Beyerer, Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (Germany)

The choice of an appropriate illumination design is one of the most important steps in creating successful machine vision systems for automated inspection tasks. In many applications it is difficult or even impossible to find a single optimal illumination condition and therefore images series under angularly-varying illumination directions are captured and analyzed. In this paper we present an illumination technique which reduces the effort for capturing inspection images for each angular reflectance feature by using a combination of point light sources over the hemisphere as illumination pattern. The key idea is to encoding linear functions for feature extraction as angular-dependent illumination patterns, and thereby to compute linear features from the scenes reflectance field directly in the optical domain. Due to the linearity of light transport, the illumination-encoded linear functions for feature extraction can be computed offline using labeled training data. In our experiments we apply the proposed illumination technique for supervised defect detection and material-based segmentation and use hand-labeled ground data for evaluation. For creating arbitrary angularly-dependent illumination patterns, we present an illumination device which uses a digital projector as programmable light source and a parabolic reflector to direct the emitted illumination patterns. Since feature extraction is transferred form images-space to the optical domain, the number of images needed to evaluate reflectance features in the inspection process is minimized.

8300-04, Session 2

An illumination-invariant phase-shift algorithm in three-dimensional profilometry

F. Deng, The Univ. of Hong Kong (Hong Kong, China) and ASM Assembly Automation Ltd. (Hong Kong, China); E. Y. Lam, The Univ. of Hong Kong (Hong Kong, China); W. Sze, J. Deng, K. S. Fung, W. Leung, C. Liu, ASM Assembly Automation Ltd. (Hong Kong, China)

Since uneven illumination is common in real optical system for machine vision applications and causes major error when using phase-shift algorithm (PSA) for surface reconstruction, we study the uneven illumination effect on phase-measuring profilometry, then propose a novel illumination-reflectivity-focus (IRF) model to handle uneven illumination effect.

Under this model, we separate the illumination factor effectively and formulate the phase reconstruction as an optimization problem. To simplify the optimization process, we calibrate the uneven illumination factor beforehand and then use the calibrated illumination information during surface profilometry. After calibration, the degrees of freedom are reduced. Accordingly, we develop a novel illumination-invariant phase-shift algorithm (II-PSA) to reconstruct the surface of a moving object under uneven illumination environment. Experimental results show that our algorithm can improve the reconstruction quality visually and numerically. Therefore, using this IRF model and the corresponding II-PSA, we can not only handle uneven illumination effect in real optical system with large filed of view, but also develop a robust and efficient method for phase-measuring profilometry. And this novel method can

be applied to real-time, high-precision machine vision applications such as inspection of semiconductor components.

8300-05, Session 2

Fusing shape and texture features for pose-robust face recognition

T. Gernoth, R. Grigat, Technische Univ. Hamburg-Harburg (Germany)

Unconstrained environments with variable ambient illumination and changes of head pose are still challenging for many face recognition systems. To recognize a person independent of pose, we separate shape from texture information using an active appearance model. We do not directly use the texture information from the active appearance model for recognition. Instead we extract local texture features from a shape and pose free representation of facial images. We use a smooth warp function to transform the images. We compensate also the shape information for head pose changes and fuse the results of separate classifiers for shape features and local texture features at the decision level. We analyze the influence of the individual contributions of shape and texture information on the recognition performance. We show that fusing shape and texture information at the decision level considerably boosts the recognition performance in an access control scenario. Furthermore, our experimental results show a significant improvement in face recognition performance on faces with pose variations when compared with a traditional appearance based face recognition system.

8300-06, Session 2

Automated inspection of tubular material based on magnetic particle inspection

A. Mahendra, C. Stolz, F. Meriaudeau, Univ. de Bourgogne (France); S. Petit, A. Noel, F. Degoutin, Vallourec S.A. (France)

Automatic industrial surface inspection methodology based on Magnetic Particle Inspection is developed from image acquisition to defect classification. First the acquisition system is optimized, then tubular material images are acquired, reconstructed then stored. The characteristics of the crack-like defect with respect to its geometric model and curvature are used as a priory knowledge for mathematical morphology and linear filtering. After the segmentation and binarization of the image, vast amount of defect candidates exists. Finally classification is performed with decision tree learning algorithm due to its robustness and speed. The parameters for mathematical morphology, linear filtering and classification are analyzed and optimized with Design Of Experiments based on Taguchi approach. The most significant parameter obtained can be analyzed and tuned further. The experiment is performed on tubular materials and evaluated by its accuracy and robustness by comparing the ground truth and testing images. The result is promising with 97 % True Positive and 0.01 % False Positive rate on the testing set.

8300-07, Session 2

Intermediate-level segmentation of color images through perception and geometry-based contour completions and shape cuts

J. Grazzini, L. Prasad, Los Alamos National Lab. (United States)

In this paper, we present a hybrid segmentation method that provides an intermediate-level representation of images into perceptually meaningful entities without any prior knowledge of the image content, nor any object model. It partitions images into generic parts made of closed regions with accurately positioned boundaries using region and boundary information. This approach is consistent with classical computational vision models, while its implementation is motivated by practical rather than theoretical considerations. It benefits from low-level feature detection, image triangulation and shape decomposition

based on the understanding of higher-level perceptual behavior. Namely, contour fragments are first extracted to approximate the actual boundaries of the image structures. Then, a triangulation is computed over subsampled contours to provide with a natural decomposition into simple building blocks. Finally, perception and geometric-based filters are applied on the interior edges of the triangulation to choose among them a set of optimal contour completions and shape cuts delineating objects. The novelty of this approach revolves around the shape decomposition technique adapted for the design of those filters. While the resulting partition does not necessarily correspond to the natural objects in the image, it can be used for an accurate detection of the real salient objects.

8300-08, Session 3

Runway hazard detection in poor visibility conditions

B. Jiang, National Institute of Aerospace (United States); Z. Rahman, Old Dominion Univ. (United States)

More recently, research on enhancing the situational awareness of pilots, especially in poor visibility flight conditions, gains more and more interests. Since pilots may not be able to spot the runway clearly in poor visibility conditions, such as fog, smoke, haze or dim lighting conditions, aviation landing problem can occur due to the (unexpected) presense of objects on the runway. Complicated and trivial instruments, switches, bottoms, plus sudden happenings are enough for the pilots to take care of during landing approach. Therefore, an automatic hazard detection approach that utilizes smart image enhancement, edge detection and image analysis is proposed. To detect the presence of objects on the runway, the existing cannoical edge representations of the runway are compard with enhanced edge representations of the geometrically corrected approach image. The effect of applying the enhancement method is to make the image of the runway almost independent the atmospheric conditions. The following edge detection process can reduce the storing space, the comparison and retrieval time, and the effect of sensor noise. After analyzing the features existing in the edge differences, the potential hazard will be localized and labeled. Experimental results show that the proposed approach is effective in runway hazard detection.

8300-09, Session 3

Application of image processing to track twin boundary motion in magnetic shape memory alloys

A. Rothenbuhler, E. H. Barney Smith, P. Müllner, Boise State Univ. (United States)

We propose an experimental method to facilitate magnetic shape memory alloy research through the use of image processing. Current magneto-mechanical experiments only measure strain in one direction. Furthermore, in situ visualization of twin boundaries are difficult as they are only visible under particular lighting conditions. The experimental method proposed here enables researchers to get strain as well as twin boundary information in a simplified experiment. Strain in Ni2MnGa shape memory alloys is optically measured by applying an edge detection algorithm. This algorithm consists of thresholding the image since the background is significantly darker than the sample itself. Through a region of interest, the stationary sample holder is cut out. A grid overlay is used to measure the detected sample's dimensions. This information is then used to calculate strain over multiple image frames. Since twin boundaries cause some parts of the sample to move, a motion detection algorithm was developed to find the moving parts of the sample. The sample's motion is analyzed with a Hough transform to extract the location of twin boundaries (they appear as straight lines). The Hough accumulator provides all necessary information about the location of all twin boundaries location and is used to visualize them on the captured image.

8300-10, Session 3

A new point process model for trajectory-based events annotation

N. Ballas, CEA LIST (France) and Mines ParisTech (France); B. Delezoide, CEA LIST (France); F. Prêteux, Mines ParisTech (France)

Human actions annotation in videos has received an increase attention from the scientific community these last years mainly due to its large implication for many computer vision applications. The current leading paradigm to perform human actions annotation is based on local features. Local features robust to geometric transformations and occlusion are extracted from a video and aggregated to obtain a global video representation. Current aggregation schemes such as Bag-of-Words or spatio-temporal grids have no or limited information about the local features spatio-temporal localization. However, it has been shown that local features localization can be discriminative for detecting a concept or action. In this work we improve on the aggregation step by embedding local features spatio-temporal information in the final video representation by introducing a point process model.

We propose an event recognition system involving two main steps:

(1) local features extraction based on robust point trajectories, and (2) a global action representation capturing the spatio-temporal context information through an innovative events-based point process clustering. A point process provides indeed a well-defined formalism to characterize local features localization along with their interaction information. It also benefits from the classification context, building clusters of trajectorie constrained by the action to detect. Results are evaluated on three standard datasets (KTH, YouTube,Hollywood2) showing a significant improvement over the state-of-art.

8300-11, Session 3

Face detection and eyeglasses detection for thermal face recognition

Y. Zheng, Alcorn State Univ. (United States)

Thermal face recognition becomes an active research direction in human identification because it does not rely on illumination condition. Face detection and eyeglasses detection are necessary steps prior to face recognition using thermal images. Infrared light cannot go through glasses and thus glasses will appear as dark areas in a thermal image. One possible solution is to detect eyeglasses and to exclude the eyeglasses areas before face matching. In thermal face detection, a projection profile analysis algorithm is proposed, where region growing and morphology operations are used to segment the body of a subject; then the derivatives of two projections (horizontal and vertical) are calculated and analyzed to locate a minimal rectangle of containing the face area. Of course, the searching region of a pair of eyeglasses is within the detected face area. The eyeglasses detection algorithm should produce either a binary mask if eyeglasses present, or an empty set if no eyeglasses at all. In the proposed eyeglasses detection algorithm, block processing, priori knowledge (i.e., low mean and variance within glasses area), region merging, and region growing are employed. The results of face recognition and eyeglasses detection will be presented, and quantitatively analyzed.

8300-12, Session 4

Combining spatial and spectral information to improve crop/weed discrimination algorithms

G. Jones, S. Villette, J. Paoli, C. Gée, AgroSup Dijon (France)

The reduction of herbicide spraying is an important key to environmentally and economically improve weed management. To achieve this, remote sensors such as imaging systems are commonly used to detect weed plants in field. We developed several spatial algorithms that detect the crop rows to discriminate crop from weeds. These algorithms have been thoroughly tested and provided robust

and accurate results without needing any learning process. Their major limitation is the fact that they only discriminate weeds in inter-row areas. Crop/Weed discrimination using spectral information is able to detect intra-row weeds but generally needs a prior learning process.

We propose a new method based on spatial and spectral information to enhance the discrimination and overcome the limitations of both algorithms. Three methods are developed and combine the advantages of spatial and spectral methods without their drawbacks.

With this new approach we are able to improve the range of weed detection in the entire field (inter and intra-row). To test the efficiency of these algorithms, a relevant database of virtual images combined to LOPEX93 spectral database has been used.

The three methods are evaluated in this paper and allow an improvement from 70% of weed detection to more than 90%.

8300-13, Session 4

Automated parasite detection in clams by transillumination imaging and pattern classification

M. E. Soto, P. A. Coelho, S. N. Torres, D. G. Sbarbaro, Univ. de Concepción (Chile)

Quality control of clams considers the detection of foreign objects like shell pieces, sand and even parasites. Particularly, Mulinia edulis clams are highly susceptible to have a parasite infection caused by the isopoda Edotea magellanica, which represents a serious commercial problem commonly addressed by manual inspection.

In this paper a novel machine vision system capable of automatically solve this need is presented. The parasite visualization inside the clam is achieved by an optoelectronic imaging system based on an innovative transillumination technique. Furthermore, automatic parasite detection in the clam's image is accomplished by a pattern recognition system designed to describe quantitatively the regions that constitute the clam.

The extracted features are used to predict the parasite presence by means of a decision tree classifier. 192 shell-off cooked Mulinia edulis clams from the Chilean south pacific coasts are used to generate a real sample dataset. More than 155,000 clam's regions patterns are employed to test the proposed detection system.

The classifier performance is evaluated using cross-validation. Primary results have shown a mean parasite detection rate of 85% and a mean general classification rate of 87%, which represent a substantive improvement to the existing solutions.

8300-14, Session 4

Vision-based, in-line fabric defect detection using yarn-specific shape features

D. Schneider, T. Aach, RWTH Aachen (Germany)

We develop a methodology for automatic in-line flaw detection in industrial woven fabrics. Where state of the art detection algorithms apply texture analysis methods to operate on low-resolved (~200ppi) image data, we describe here a process flow to segment single yarns in high-resolved (~1000ppi) textile images. Four yarn shape features are extracted, allowing a precise detection and measurement of defects. The degree of precision reached, allows a classification of detected defects according to their nature, providing an innovation in the challenge of automatic fabric flaw detection. The design has been carried out to meet real time requirements and face adverse conditions caused by loom vibrations and dirt. The entire process flow is discussed followed by an evaluation using a database with real-life industrial fabric images. This work pertains to the construction of an on-loom defect detection system to be used in the manufacturing practice.

8300-15, Session 4

3D temperature mapping of turboshaft components using thermal paints and color recognition

S. Guérin, Turbomeca SA (France) and ONERA, The French Aerospace Lab (France); C. Lempereur, ONERA, The French Aerospace Lab. (France); P. Brevet, Turbomeca SA (France)

In order to enhance turboshaft lifespan, aeronautical engine manufacturers develop specific measurement techniques. Particularly, surface temperature of engine components is a critical aspect of the design. Usual temperature measurement techniques reach their limit in this harsh environment so thermal paints have been used since several years associated with skilled operator observations. This article describes a system capable of robust color recognition applied to thermal paints color variation. The relevance of the described method is to outperform temperature resolution and robustness and to make more reliable comparison between numerical simulation and bench test measurement.

A 3D non-contact digitizer captures in the same plane color images and depth map which facilitates the projection of the measured temperature map on the 3D mesh. Temperature map is processed by means of a color recognition algorithm associated with a color to temperature database. The database is composed of a set of metallic samples heated in a laboratory kiln in 10°C steps. Different colorimetric distances are used to compare each pixel to the database and find the best matching sample. Finally, the system gives a robust measurement, which is currently used in the engine development process at Turbomeca.

8300-25, Session 4

Strain analysis by regularized non-rigid registration

A. Badshah, P. L. O'Leary, M. J. Harker, Montan Univ. Leoben (Austria)

This paper investigates a new regularization scheme which can be used in conjunction with non-rigid registration to perform optical material strain analysis. Much work has been performed on optical strain analysis, see for example,1{4 these all use some form of elastic registration to perform non-contact strain measurement. However, the issue of regularization is not systematically dealt with: for example, Haile and Ifuj4 use a locally weighted mean transform. This is in no way related to the physics of the problem at hand. Koljonen et al.5 investigated di®erent measures for con¯dence and adapted the template size to achieve a higher degree of con¯dence. This paper investigates using a set of tensor polynomials to implement regularization. The tensor polynomials correspond to basis functions which form solutions to the Navier-Stokes partial di®erential equation,

@u

@t

+ u ¢ ru =

rP

½

+ vr2u (1)

whereby, u is the velocity vector, P is the pressure vector, ½ the °uid density and v is the kinematic viscosity.

This equation describes the °ow of viscous media, i.e. the physics underlying this measurement problem. The coe±cients for the tensor polynomials are determined via a direct least squares approach with local entropy weighting. The non-rigid registration method utalized is based on a hierarchical decomposition with decimation. The procedure has been described in detail by the authors in.6 The procedure is very stable due to the entropy weighting, i.e. each patch in the image is by it relative entropy during the least squares approximation. The new direct ¯tting procedure is su±ciently fast so that it can be used in real time. The universal testing machine used can be seen in Figure 1 and the sample held in the jaws of the machine in Figure 2. The results of the

first test on the extension of a sample is shown in Figure 3: the sample prior to and after extension are shown together with the last two layers of the hierarchical registration are shown. At each layer in the hierarchy the patch size is halved in size and the number of nodes in the support of the corresponding tensor product is increased. The color of the patch is proportional to the entropy in the corresponding image patch and is used as the weighting during fitting. The entropy weighting ensures that patched with only a little information do not contribute strongly to the least squares solution. It can be seen from this figure that the correct extension and shearing of the material has bee detected.
In the second test a porous compressible medium is compressed, see Figure 4 and 5. The typical barrel distortion with flat tops and bottoms can be observer. The area of the original grid can be compared with the area of the compressed grid to determine the volumetric compression of the material. In this work it has been shown that the new techniques based on tensor polynomial product solutions to the Navier-Stokes equation can be used to implement regularization in optical strain measurements. The use of entropy weighting leads to a robust solution. This abstract shown that there is justification for further investigation of this technique.

Conference 8301: Intelligent Robots and Computer Vision XXIX: Algorithms and Techniques

Monday-Tuesday 23-24 January 2012

Part of Proceedings of SPIE Vol. 8301 Intelligent Robots and Computer Vision XXIX: Algorithms and Techniques

8301-01, Session 1

Software-based, neural-network-assisted movement compensation for nanoresolution piezo actuators

J. Röning, M. Kauppinen, Univ. of Oulu (Finland)

This paper presents a software based controller implementation utilizing neural networks for high precision positioning of a piezoelectric actuator. The controller developed can be used for controlling nanopositioning piezo actuators when sufficiently accurate feedback information is available.

Piezo actuators exhibit complex hysteresis dynamics that need to be taken into account when designing an accurate control system. For inverse modelling purposes of the hysteresis related phenomena, a static hysteresis operator and a new developed dynamic creep operator is presented to be used in conjunction with a Feed Forward type neural network.

The controller utilizing the neural network inverse hybrid model is implemented as a software component for existing Scalable Modular Control framework (SMC). Using the SMC framework and off-the-shelf components, a measurement and control system for the nanopositioning actuator is constructed and tested using two different capacitive sensors operating on y- and z-axes of the actuator.

Using the developed controller, piezo actuator related hysteresis phenomena were successfully reduced making the nanometer range positioning of the actuator axes possible. Also the effect of using a lower accuracy position sensor with more noise to control accuracy is briefly discussed.

8301-02, Session 1

Traffic monitoring with distributed smart cameras

O. Sidla, M. Rosner, SLR Engineering OG (Austria); M. Ulm, Austrian Institute of Technology (Austria)

The observation and monitoring of traffic with smart visions systems will become increasingly important in the years to come. In this paper we present a system which is designed to monitor traffic at a relatively busy pedestrian crossing is described. The system observes the interaction between vehicles and pedestrians at the crossing and tries to detect incidents in which the interaction of pedestrians and vehicles might lead to dangerous or even critical encounters.

The proposed system consists of a cluster of 3 networked smart cameras, based on compact PC hardware:

Camera 1,2: Vehicle detection and tracking based on a state-of-the-art cascaded HOG detector. The tracker includes optical flow computation with a modified KLT point tracker.

Camera 3: Pedestrian tracking with a tracking-by-detection approach based on a cascaded HOG detector.

We describe the processing pipeline of the object (pedestrian, vehicle) tracking system which combines

- a cascaded HOG detector

- a modified and improved KLT point tracker

- feature analysis to enhance tracking and remove false tracking occurrences

The well known KLT tracking algorithm has been modified in order to include state-of-the-art keypoint detection so that the quality of feature points which are selected for tracking improves significantly compared to the original algorithm.

After detection and tracking all 3 smart cameras project their video detection results to ground plane. Based on the location and velocity of all objects detected in real-world co-ordinates, the master camera then executes a set of rules in order to decide whether a critical event is currently being observed.

The remaining paper concentrates on the system architecture and describes results of our experiments during extensive trials and tests in an outdoor environment in the City of Vienna, Austria.

8301-03, Session 1

The 19th annual intelligent ground vehicle competition: student-built autonomous ground vehicles

B. L. Theisen, U.S. Army Tank Automotive Research, Development and Engineering Ctr. (United States)

The Intelligent Ground Vehicle Competition (IGVC) is one of four, unmanned systems, student competitions that were founded by the Association for Unmanned Vehicle Systems International (AUVSI). The IGVC is a multidisciplinary exercise in product realization that challenges college engineering student teams to integrate advanced control theory, machine vision, vehicular electronics and mobile platform fundamentals to design and build an unmanned system. Teams from around the world focus on developing a suite of dual-use technologies to equip ground vehicles of the future with intelligent driving capabilities. Over the past 19 years, the competition has challenged undergraduate, graduate and Ph.D. students with real world applications in intelligent transportation systems, the military and manufacturing automation. To date, teams from almost 80 universities and colleges have participated. This paper describes some of the applications of the technologies required by this competition and discusses the educational benefits. The primary goal of the IGVC is to advance engineering education in intelligent vehicles and related technologies. The employment and professional networking opportunities created for students and industrial sponsors through a series of technical events over the four-day competition are highlighted. Finally, an assessment of the competition based on participation is presented.

8301-04, Session 2

Accurate, dense 3D reconstruction of moving and still objects from dynamic color image and depth image sequences based on temporal modified-RANSAC and feature-cut

N. Tatematsu, J. Ohya, Waseda Univ. (Japan)

The technology for mobile robots to automatically reconstruct 3D structure of the moving and still objects is increasing its significance. We proposed a Temporal modified-RANSAC (TMR) based method that can track moving objects from moving stereo cameras and can reconstruct 3D structure of the tracked objects. However, the TMR based method has the following problems: lack of accuracy of segmenting 3D points and lack of density of 3D reconstructed points.

To solve the problem of our conventional method, this paper proposes a new 3D segmentation method that utilizes Graph-cut, which is frequently used for segmentation tasks. More specifically, our

Conference 8301

proposed method utilizes pre-defined energy functions that consist of color probability and spatial structure probability as well as a-priori probabilities based on Feature-Cut, which enables to segment 3D points accurately based on the segmentation results in the previous frame, even if edge feature points of the tracked object are not successfully found.

Experimental results using real scenes that contain dynamic and static objects demonstrate that this algorithm can reconstruct dense 3D points of the moving objects and background. In addition, we show experimental results that prove that the accuracy of segmentation is higher than only Feature-Cut.

8301-05, Session 2

Efficient hybrid monocular-stereo approach to on-board, video-based traffic sign detection and tracking

J. Marinas, L. Salgado, J. Arróspide, M. Camplani, Univ. Politécnica de Madrid (Spain)

In this paper we propose an innovative method for the automatic detection and tracking of road traffic signs using an on-board stereo camera. It exploits a combination of monocular and stereo analysis strategies to increase the reliability of the detections such that it can boost the performance of any traffic sign recognition scheme. Firstly, an adaptive color and appearance based detection is applied at single camera level, followed by a SURF based matching strategy to perform a 3D reconstruction of candidate road signs: the best-fitting plane to the 3D cloud of points previously detected is estimated using a RANSAC based approach to improve robustness to outliers. Temporal consistency of the 3D information is incorporated through a Kalman based tracking stage, and tracking results are used twofold: to adapt the previously mentioned color-based detector, thus improving detection accuracy, and to reduce the need to compute stereo analysis for every incoming image, thus increasing algorithm efficiency. The proposed solution has been tested with real sequences under several illumination conditions and in both urban areas and highways, achieving very high detection rates in challenging scenarios, including rapid motion and significant perspective distortion.

8301-06, Session 2

A general model and calibration method for spherical stereoscopic vision

W. Feng, Tianjin Univ. (China); J. Röning, Univ. of Oulu (Finland); X. Zong, Tianjin Univ. of Technology (China); B. Zhang, Tianjin Univ. (China); J. Kannala, Univ. of Oulu (Finland)

In geometrical stereoscopic vision calibration the objective is to determine a set of parameters which describe the mapping from 3D reference coordinates to 2D image coordinates, and indicate the geometric relationships between the cameras. While various methods for ordinary lenses stereoscopic vision can be found from the literature, stereoscopic vision with extremely wide angle lenses calibration have been much less discussed. Spherical stereoscopic vision is more and more convenient in computer vision applications. However, it use for 3D measurement purposes is limited by the lack of an accurate, general, and easy-to-use calibration procedure. Hence, we present a geometric model for spherical stereoscopic vision equipped by extremely wide angle lenses. Then, a corresponding generic mathematical model is built. Method for calibration the parameters of the mathematical model is proposed. This paper shows practical results from the calibration of two high quality panomorph lenses mounted on double cameras with 2048x1536 resolutions. Here, the stereoscopic vision system is flexibly, the position and orientation of the cameras can be adjusted randomly. The calibration results include interior orientation, exterior orientation and the geometric relationships between two cameras. The achieved level of calibration accuracy is very satisfying.

8301-07, Session 2

An approach to stereo-point cloud registration using image homographies

S. D. Fox, D. M. Lyons, Fordham Univ. (United States)

A mobile robot equipped with a stereo camera can measure both the video image of a scene and the visual disparity in the scene. The disparity image can be used to generate a collection of points, each representing the location of a surface in the visual scene as a 3D point with respect to the location of the stereo camera: a point cloud. If the stereo camera is moving, e.g., mounted on a moving robot, aligning these scans becomes a difficult, and computationally expensive problem. Many finely tuned versions of the iterative closest point algorithm (ICP) have been used throughout robotics for registration of these sets of scans. However, ICP relies on theoretical convergence to the nearest local minimum of the dynamical system: there is no guarantee that ICP will accurately align the scans. In order to address two problems with ICP, convergence time and accuracy of convergence, we have developed an improvement by using salient keypoints from successive video images to calculate an affine transformation estimate of the camera location. This transformation, when applied to the target point cloud, provides ICP an initial guess to reduce the computational time required for point cloud registration and improve the quality of registration. We report ICP convergence times with and without image information for a set of stereo data point clouds to demonstrate the effectiveness of the approach.

8301-08, Session 2

Hazardous sign detection for safety applications in traffic monitoring

W. Benesova, Slovak Univ. of Technology (Slovakia); O. Sidla, SLR Engineering OG (Austria); M. Kottman, Slovak Univ. of Technology (Slovakia); E. Sikudova, Z. Cernekova, Comenius Univ. in Bratislava (Slovakia)

The transportation of hazardous goods in public streets systems can pose severe safety threats in case of accidents.

One of the solutions for these problems is an automatic detection and registration of vehicles which are marked with dangerous goods signs. In this way, traffic authorities can be made aware of the presence of dangerous loads on their streets.

Towards such an automated registration of special signs we present a prototype system which can detect a trained set of signs (typically a set of approx 50 different logos) under real-world conditions in real-time.

This paper compares two different approaches for the detection:

Brute force matching with keypoint detection and feature matching and a hybrid approach based on a code-book of visual words and bag of word procedure.

The results of an extended series of experiments are provided in this paper. The experiments are designed to verify the usability of the two proposed approaches in a real-world scenario.

All methods are tested and evaluated using two types of test data: experiments with printed logos and artificial sequences and detection of signs on actual traffic on roads.

8301-09, Session 2

PRoViScout: a planetary scouting rover demonstrator

G. M. Paar, JOANNEUM RESEARCH Forschungsgesellschaft mbH (Austria); M. Woods, SciSys Ltd. (United Kingdom); C. Gimkiewicz, Ctr. Suisse d'Electronique et de Microtechnique SA (Switzerland); F. Labrosse, Aberystwyth Univ. (United Kingdom); A. Medina, GMV S.A. (Spain)

Mobile systems exploring Planetary surfaces in future will require more autonomy than today. The EU FP7-SPACE Project ProViScout (2010-

2012) establishes the building blocks of such autonomous exploration systems in terms of robotics vision by a decision - based combination of navigation and scientific target selection, and integrates them into a framework ready for and exposed to field demonstration.

The PRoViScout on-board system consists of Mission Management components such as an Executive, a Mars Mission On-Board Planner and Scheduler, a science assessment module, and navigation & vision processing modules. The Platform HW consists of the rover with the sensors and pointing devices.

We report on the major building blocks and their functions & interfaces, emphasizing on the computer vision parts such as Image acquisition (using a novel zoomed 3D-TOF & RGB camera), Mapping from 3D-TOF data, panoramic image & stereo reconstruction, hazard and slope maps, Visual Odometry and the maintenance of a global map.

The paper will emphasize on the systems engineering and computer vision aspects as well as the benefits of the 3D-TOF concept.

Future activities will address the applicability in real environment at a field test on the island of Tenerife during summer 2012.

8301-10, Session 3

Red-light traffic enforcement at railway crossings

O. Sidla, M. Rosner, SLR Engineering OG (Austria)

Railway crossings, even when secured by automated barriers, pose a threat to drivers day and night. Especially drivers which try to cross the railway in the last moments before a barrier closes have a high probability to cause severe accidents.

This work presents an automated system which is designed to detect and record vehicles which move over the railway crossing after the red light has been activated just before the automated barrier closes. The system consists of a smart camera which observes one street lane as well as the red light at the crossing. Triggered by an electrical signal from the railway, the camera begins to observe the crossing. The detection of moving vehicles is based on a custom trained cascaded HOG detector in combination with motion analysis. Specifically we implement a tracker which is based on

- a cascaded HOG detector

- a modified and improved KLT point tracker

- feature analysis to enhance tracking and remove false tracking/ detection occurrences

We demonstrate that the combination of object detection and optical flow can result in robust monitoring systems which operate day and night and are robust enough in order to be usable in real-world situations.

The remaining sections of the paper concentrate on the system architecture and describes results of our experiments during tests at two installation sites in Austria.

8301-11, Session 3

Image projection clues for improved real-time vehicle tracking in tunnels

V. Jelaca, J. O. Niño-Castaneda, A. Pizurica, W. Philips, Univ. Gent (Belgium)

Vehicle tracking is of great importance for tunnel safety. To detect incidents or disturbances in traffic flow it is necessary to reliably track vehicles in real-time. The tracking is a challenging task due to poor lighting conditions in tunnels and frequent light reflections from tunnel walls, the road and the vehicles themselves. In this paper we propose a multi-clue tracking approach combining foreground blobs, optical flow of Shi-Tomasi features and image projection profiles in a Kalman filter with a constant velocity model. The main novelty of our approach lies in using vertical and horizontal image projection profiles (so-called vehicle signatures) as additional measurements to overcome the problems of inconsistent foreground and optical flow

clues in cases of severe lighting changes. These signatures consist of Radon-transform like projections along each image column and row. We compare the signatures from two successive video frames to find their alignment and to correct predicted vehicle position and size. We tested our approach on several tunnel sequences. The results show an improvement in the accuracy of the tracker and less target losses when image projection clues are used. Furthermore, calculation and comparison of image projections is computationally efficient so the tracker keeps real-time performance (25 fps, on a single 1.86 GHz processor).

8301-12, Session 3

Decentralized tracking of humans using a camera network

S. Gruenwedel, V. Jelaca, J. O. Niño-Castañeda, P. Van Hese, D. Van Cauwelaert, P. Veelaert, W. Philips, Univ. Gent (Belgium)

Real-time tracking of people has many applications in computer vision and typically requires multiple cameras, for instance for surveillance, domotics, elderly-care and video conferencing. The problem is challenging because of the need to deal with frequent occlusions and environmental changes. Another challenge is to develop solutions which scale well with the size of the camera network. Such solutions need to carefully restrict overall communication in the network and often involve dis-tributed processing. In this paper we present a distributed person tracker addressing the aforementioned issues. Real-time processing is achieved by distributing tasks between the cameras and a central server. The latter fuses only high level data based on low-bandwidth input streams from the cameras. This is achieved by performing tracking first on the image plane of each camera followed by sending only meta data to a local fusion center. We designed the proposed system with respect to a low communication load and towards robustness of the system. We evaluate the performance of the tracker in meeting scenarios where persons are often occluded by other persons and/or furniture. We present experimental results which show that our tracking approach is accurate even in cases of severe occlusions in some of the views.

8301-13, Session 3

Real-time detection of traffic events using smart cameras

M. M. Macesic, Tehnomobil-Protech (Serbia); V. Jelaca, J. O. Niño-Castaneda, Univ. Gent (Belgium); N. Prodanovic, M. Panic, Univ. of Novi Sad (Serbia); A. Pizurica, Univ. Gent (Belgium); V. Crnojevic, Univ. of Novi Sad (Serbia); W. Philips, Univ. Gent (Belgium)

With rapid increase of number of vehicles on roads it is necessary to maintain close monitoring of traffic. For this purpose many surveillance cameras are placed along roads and on crossroads, creating a huge communication load. Therefore, the data needs to be processed on site and transferred to the monitoring centers in form of metadata or as a set of selected images. For this purpose it is necessary to detect events of interest already on the camera side, which implies using smart cameras as visual sensors.

In this paper we propose a method for tracking of vehicles and analysis of vehicle trajectories to detect the different traffic events. Kalman filtering was used for tracking, combining foreground and optical flow measurements.

Obtained vehicle trajectories are used to detect different traffic events. Every new trajectory is compared with collection of normal routes and clustered accordingly. If the observed trajectory differs from all normal routes more than a predefined threshold, it is marked as abnormal and the alarm is raised.

System was developed and tested on Texas Instruments OMAP platform. Testing was done on four different locations, two locations in the city and two locations on the open road.

8301-14, Session 3

Mixed road traffic: data acquisition, optical tracking, and microscopic modeling

R. Schönauer, Technische Univ. Graz (Austria); Y. Lypetskyy, SLR Engineering OG (Austria)

In the last years mixed traffic designs for urban roads are applied in increasing numbers. Important research tasks arise with this trend: To study and model the interaction processes between traffic participants as well as the effects of urban infrastructural parameters and layouts.

This paper extends agent based microscopic modeling for handling interaction processes within motorized and non-motorized traffic and infrastructural elements. The work in this paper presents dynamic vehicle models for bicycles and cars representing their limited movement capabilities.

To cover a large and versatile amount of real world data for calibration and validation processes this paper proposes semi-automated data acquisition by video analysis. This work concentrates mainly on the aspects of a semi-automatic annotation tool applied to create trajectories of traffic participants over space and time. Based on manual identification of a vehicle or pedestrian the trajectory is continued with the help of object tracking algorithms.

The acquired data is then applied to calibrate a single track model which navigates through a road's surface and interacts with its environment. The results of optimized control and preview parameters show stable behavior and a good spatial and temporal correlation with the data.

8301-15, Session 4

AR.Drone: security threat analysis and exemplary attack to track objects or persons

F. Samland, J. Fruth, M. Hildebrandt, T. Hoppe, J. Dittmann, Otto-von-Guericke-Univ. Magdeburg (Germany)

In this article we illustrate an approach of a security threat analysis of a quadrocopter AR.Drone, a toy for augmented reality (AR) games. The technical properties of the drone can be misused for attacks, which may relate security and/or privacy aspects. Our aim is to sensitise for the possibility of misuses and the motivation for realisation of improved security mechanisms of the quadrocopter. We focus on primary obvious security vulnerabilities (e.g. communication over unencrypted WLAN, usage of UDP, life video streaming via unencrypted WLAN to the control device) of this quadrocopter. We could practically verify that this can be misused by unauthorised persons for several attacks, e.g. the eavesdropping of AR.Drone's unprotected video streams, high-jacking the drone. Amongst other aspects, our current research focuses on the realisation of the attack of tracking persons and objects with the drone. Besides the realisation of attacks, we want to evaluate the potential of this particular drone for a "safe-landing" function. Additionally, in future we plan an automatic tracking of persons or objects without the need of human interactions.

8301-16, Session 4

Detection of unknown targets from aerial camera and extraction of simple object fingerprints for the purpose of target reacquisition

T. N. Mundhenk, K. Ni, Y. Chen, K. Kim, Y. Owechko, HRL Labs., LLC (United States)

An aerial multiple camera tracking paradigm needs to not only spot unknown targets and track them, but also needs to know how to handle target reacquisition as well as target handoff to other cameras in the operating theater. Here we discuss such a system which is designed to spot unknown targets, track them, segment the useful features and then create a signature fingerprint for the object so that it can be reacquired or handed off to another camera. The tracking system spots unknown objects by subtracting background motion from observed motion allowing it to find targets in motion, even if the camera platform itself is moving. The area of motion is then matched to segmented regions returned by the EDISON mean shift segmentation tool. Whole segments which have common motion and which are contiguous to each other are grouped into a master object. Once master objects are formed, we have a tight bound on which to extract features for the purpose of forming a fingerprint. In videos which contain multiple vehicle occlusions and vehicles of highly similar appearance we obtain a reacquisition rate for automobiles of over 80% using the simple single Gaussian model compared with the null hypothesis of 25%.

8301-18, Session 4

Superresolution terrain map enhancement for navigation based on satellite imagery

J. Straub, Jacksonville State Univ. (United States)

The utility of satellite imagery for many applications is limited by its resolution. Resolution can be improved via enhancing the satellite imaging hardware; however, software enhancement can be used to overcome hardware technical and cost limitations. Super resolution is a set of image enhancement techniques which produce a higher resolution image from one or more lower-resolution ones. Conventional super resolution techniques are designed to work on imagery. However, some super resolution techniques can be applied directly to non-visual data.

This paper looks at the enhancement of digital elevation model (DEM) terrain height data using a database based inference technique. NASA/NGIA SRTM data collected from the Spaceborne Imaging Radar-C (SIR-C) was used to train the engine and alternate SRTM data was presented for enhancement. The application of super resolution to aircraft (specifically UAV) navigation path-planning was considered. Minimum safe distance constraints were defined and planning performed using the super-resolved data was compared to the up-scaled low-resolution data. Paths that were created were evaluated based on route optimality and constraint-violation-avoidance.

8301-19, Session 5

3D positional control of magnetic levitation system using adaptive control: improvement of positioning control in horizontal plane

T. Nishino, N. Kato, Mie Univ. (Japan); N. Tsuda, Wakayama National College of Technology (Japan); Y. Nomura, Mie Univ. (Japan)

The objective of this paper is to establish a technique that levitates and conveys a hand, a kind of micro-robot, by applying magnetic forces: the hand is assumed to have a function of holding and detaching the objects.

The equipment to be used in our experiments consists of four pole-pieces of electromagnets, and is expected to work as a 4DOF drive unit within some restricted range of 3D space: the three DOF are corresponding to 3D positional control and the remaining one DOF, rotational oscillation damping control. Having used the same equipment, Khamesee et al. had manipulated the impressed voltages on the four electric magnetics by a PID controller by the use of the feedback signal of the hand's 3D position, the controlled variable. However, in this system, there were some problems remaining: in the horizontal direction, when translating the hand out of restricted region, positional control performance was suddenly degraded.

The authors propose a method to apply an adaptive control to the horizontal directional control. It is expected that the technique to be presented in this paper contributes not only to the improvement of the response characteristic but also to widening the applicable range in the horizontal directional control.

Conference 8301

8301-20, Session 5

Robots in agriculture: an I-weed robot for a specific spraying

G. Salis, C. Gée, S. Villette, J. Paoli, G. Jones, AgroSup Dijon (France)

Considering the spatial variability of a field, it consists in adapting cultures and treatments to each area of the field rather than to apply an established model based on average information over the entire field. The development of new technologies (i.e. remote sensing, image processing, computer science, robotics), was an important step towards sustainable agriculture. In the early 1980s, first robots were developed in labs and they were dedicated to particular task: harvesting, planting, spraying...

Currently, we are developing an I-Weed robot (Intelligent Weed Robot) that is a chemical weeding robot. It is guided by a high precision GPS signal (Real Time Kinetic) with a precision of about 2 or 3 centimeters. Our robot is able to move between the crop's rows thanks to a route previously recorded. A camera positioned in the front of our device is equipped with a near infrared filter. It allows to identify and to locate the position of weeds. At the back of the robot, a spray boom system triggers at the right time the right nozzle in order to spray only on weed plants.

8301-21, Session 5

The magic glove: a gesture-based remote controller for intelligent mobile robots

C. Luo, Y. Chen, M. Krishnan, M. Paulik, Univ. of Detroit Mercy (United States)

This paper describes the design of a gesture-based Human Robot Interface (HRI) for an autonomous mobile robot entered in the Intelligent 2010 Ground Vehicle Competition (IGVC). While the robot is meant to operate autonomously in the various Challenges of the competition, an HRI is useful in moving the robot to the starting position.

 In this project a gesture-based embedded system called the Magic Glove is developed for remote control of a robot. The system consists of a microcontroller and sensors that is worn by the operator as a glove and is capable of recognizing hand signals. These are then transmitted through wireless communication to the robot. The design of the Magic Glove included contributions on two fronts: hardware configuration and algorithm development. A triple axis accelerometer used to detect hand orientation passes the information to a microcontroller, which interprets the corresponding vehicle control command. A Bluetooth device interfaced to the microcontroller then transmits the information to the vehicle, which acts accordingly.

 The Magic Glove was successfully demonstrated first in a Player/ Stage simulation environment. The gesture-based functionality was then also successfully verified on an actual robot and demonstrated to judges at the 2010 IGVC.

8301-22, Session 5

Way-point navigation for a skid steer vehicle in unknown environments

P. Chen, A. Das, P. Mukherjee, S. Waslander, Univ. of Waterloo (Canada)

This paper proposes a combined mapping, path planning, and control solution that will allow a skid-steer UGV to navigate safely to a desired location in an unknown environment. An Extended Kalman Filter (EKF) is employed using a non-linear model of the skid steer vehicle which assumes the vehicle operates on the condition of constant slippage. The proposed solution generates a 3D point cloud by nodding the planar LIDAR, and then uses an innovative algorithm to construct a 2D map of the traversable space. The map data is efficiently stored in memory using an octree structure with real-time pruning. Using this 2D map, the proposed path planner will generate dynamically feasible and collision-free paths to the desired location, and will re-plan these paths as necessary when new obstacles are detected. The path planner is based on the probabilistic roadmap (PRM) algorithm using a dynamic A* (D*) search, and considers the kinodynamic constraints of the vehicle. Finally, a non-linear tracking controller is modified for the skid-steer UGV to allow the vehicle to track the path. All of the algorithms are computationally efficient and allow for online processing on-board the robot, and were successfully implemented on a custom made skid steer vehicle.

8301-23, Session 6

Integrated field testing of planetary robotics vision processing: the PRoVisG campaign in Tenerife 2011

G. M. Paar, JOANNEUM RESEARCH Forschungsgesellschaft mbH (Austria); L. Waugh, EADS Astrium Ltd. (United Kingdom); D. P. Barnes, Aberystwyth Univ. (United Kingdom); T. Pajdla, Czech Technical Univ. in Prague (Czech Republic); M. Woods, SciSys Ltd. (United Kingdom); H. Graf, Ctr. Suisse d'Electronique et de Microtechnique SA (Switzerland); Y. Gao, Univ. of Surrey (United Kingdom); K. Willner, Technische Univ. Berlin (Germany); J. A. Muller, Univ. College London (United Kingdom); R. Li, The Ohio State Univ. (United States); M. Maurette, Ctr. National d'Études Spatiales (France)

In order to maximize the use of a robotic probe during its limited lifetime, scientists immediately have to be provided the best achievable visual quality of 3D data products. The EU FP7-SPACE Project PRoVisG (2008-2012) develops technology for the rapid processing and effective representation of visual data by improving ground processing facilities. Its ambition is to collect a tool set and integrate a versatile and flexible processing chain easily to be adapted to various tasks.

PRoVisG is holding a field test in September 2011 on the Caldera of Tenerife to verify the implemented 3D Vision processing mechanisms and to collect various sets of reference data in representative environment.

The paper covers the preparation work for such a test and highlights the test elements that include standard operations- and science-related components but also data capture to verify specific processing functions.

The campaign is supported by the Astrium UK Rover Bridget as a representative platform which allows simultaneous on-board mounting and powering of various vision sensors such as the ExoMars PanCam Emulator by Aberystwyth University (AUPE).

The final manuscript will contain a description of the captured data and the processing results, as well as a summary of the experience gathered during the test.

8301-24, Session 6

Hierarchical loop detection for mobile outdoor robots

D. Lang, C. Winkens, M. Häselich, D. W. Paulus, Univ. Koblenz-Landau (Germany)

Loop closing is a fundamental part of 3D simultaneous localization and mapping (SLAM) that can greatly enhance the quality of long-term mapping. It is essential for the creation of globally consistent maps.

Conceptually, loop closing is divided into detection and optimization.

Recent approaches depend on a single sensor to recognize previously visited places in the loop detection stage. In this study, we combine data of multiple sensors such as GPS, vision, and laser range data to enhance detection results in repetitively changing environments that are not sufficiently explained by a single sensor. We present a fast and robust hierarchical loop detection algorithm for outdoor robots to achieve a reliable environment representation even if one or more sensors fail.

Conference 8301

8301-25, Session 6

A novel margin-based, linear embedding technique for visual object recognition

F. Dornaika, Univ. del País Vasco (Spain)

Linear Dimensionality Reduction (LDR) techniques have been increasingly important in computer vision and pattern recognition since they permit a relatively simple mapping of data onto a lower dimensional subspace, leading to simple and computationally efficient classification strategies. Recently, a linear method called Average Neighbors Margin Maximization (ANMM) was proposed and shown to have powerful discrimination properties. Although good results were obtained with this method, it suffers from two shortcomings: i) it requires the setting of two parameters-the neighbors sizes for homogeneous and heterogeneous samples, and ii) it can be very sensitive to the label outliers since a margin average is used. In this paper, we propose a novel margin maximization that is based on the nearest hit and the nearest miss samples only. The proposed approach finds the projection directions such that the sum of local margins is maximized. Our proposed approach has been applied to the problem of appearance-based face recognition. Experimental results performed on three face databases show that the proposed approach can give better generalization performance than the ANMM method. Our proposed method is much less sensitive to outliers than the ANMM method. The proposed approach could also be applied to other category of objects characterized by large variations in their appearance.

8301-26, Session 6

Real-time, two-level foreground detection and person-silhouette extraction enhanced by body-parts tracking

R. Deeb, É. Desserée, S. Bouakaz, Univ. Claude Bernard Lyon 1 (France)

We propose a real-time, two-level foreground detection, enhanced by body parts tracking, designed to efficiently extract person silhouette for monocular video-based human motion analysis systems. We aim to find solutions for non-controlled environment challenges, such as light changes, camouflage and occlusions, which make the detection and the tracking of a moving person a hard task to accomplish. On the first level, we propose an enhanced Mixture of Gaussians, built on both chrominance-luminance and chrominance-only spaces, which handles global illumination changes. On the second level, we improve segmentation results, in interesting areas, by using statistical foreground models updated by a high-level tracking of body parts. Each body part is represented with a set of templates characterized by a feature vector built in an initialization phase. Then, a high level tracking is achieved by finding blob-template correspondences via distance minimization in feature space. Correspondences are then used to update foreground statistical models, then a Maximum A Posteriori framework is used to refine segmentation. We were able to extract a refined silhouette in the presence of light changes, noise and camouflage. Moreover, the tracking approach allowed us to infer information about the presence and the location of body parts even in the case of partial occlusion.

8301-27, Session 6

Activity recognition from video using layered approach

C. A. McPherson, J. M. Irvine, M. Young, Draper Lab. (United States); A. Stefanidis, George Mason Univ. (United States)

The adversary in current threat situations can no longer be identified by what they are, but by what they are doing. This has lead to a large increase in the use of video surveillance systems for security and defense applications. With the quantity of video surveillance at the disposal of organizations responsible for protecting military and civilian lives comes issues regarding the storage and screening the data for events and activities of interest.

Activity recognition from video for such applications seeks to develop automated screening of video based upon the recognition of activities of interest rather than merely the presence of specific persons or vehicle classes developed for the Cold War problem of "Find the T72 Tank". This paper explores numerous approaches to activity recognition, all of which examine heuristic, semantic, and syntactic methods based upon tokens derived from the video.

The proposed architecture discussed uses a multi-level approach that divides the problem into three or more tiers of recognition, each employing different techniques according to their appropriateness to strengths at each tier using heuristics, syntactic recognition, and HMM's of token strings to form higher level interpretations.

8301-28, Session 7

Method for fast detecting the intersection of a plane and a cube in an octree structure to find point sets within a convex region

K. Fujimoto, N. Kimura, T. Moriya, Hitachi, Ltd. (Japan)

Performing efficient view frustum culling is a fundamental problem in computer graphics. In general, an octree is used for view frustum culling. The culling checks the intersection of each octree node (cube) against the planes of the view frustum. However, this involves many calculations. We propose a method for fast detecting the intersection of a plane and a cube in an octree structure. When we check which child of the octree node intersects a plane, we compare the coordinates of the corner of the node and the plane. Using an octree, we calculate the vertices of the child node by using the vertices of the parent node. To find points within a convex region, a visibility test is performed by AND operation with the result of three or more planes. In experiments, we tested the problem of searching for the visible point with a camera. The method was two times faster than the conventional method, which detects a visible octree node by using the inner product of the plane and each corner of the node.

8301-30, Session 7

Lucas-Kanade image registration using camera motions

S. Cho, H. Cho, Pohang Univ. of Science and Technology (Korea, Republic of); Y. S. Moon, J. Cho, S. Lee, Samsung Electronics Co., Ltd. (Korea, Republic of); S. Lee, Pohang Univ. of Science and Technology (Korea, Republic of)

Lucas-Kanade based image registration methods are one of the most widely used algorithms in computer vision. The methods have been successfully used in numerous works, which include image registration as a component in the process. In this paper, we propose Lucas-Kanade based image registration methods using camera motions. We decompose a homography into camera intrinsic and extrinsic parameters, and estimate only extrinsic parameters, assuming that intrinsic parameters are given, e.g., from the EXIF information of a photograph. As the intrinsic and extrinsic parameters are based on the actual configuration of the camera, the proposed method can perform image registration more reliably. In addition, as the number of extrinsic parameters is smaller than the number of homography elements, the proposed method runs faster than the previous Lucas-Kanade based registration method that estimates a homography itself.

8301-31, Session 7

Object tracking with adaptive HOG detector and adaptive Rao-Blackwellised particle filter

S. Rosa, M. Paleari, P. Ariano, Istituto Italiano di Tecnologia (Italy); B. Bona, Politecnico di Torino (Italy)

Scenarios for a manned mission to the Moon or Mars call for astronaut teams to be accompanied by semi-autonomous robots. A prerequisite for human-robot interaction is the capability of successfully tracking humans and objects in the environment.

In this paper we present a system for real-time visual object tracking in 2D images for mobile robotic systems. The proposed algorithm is able to specialize to individual objects and to adapt to substantial changes in illumination and object appearance during tracking.

The algorithm is composed by two main blocks: a detector based on Histogram of Oriented Gradient (HOG) descriptors and linear Support Vector Machines (SVM), and a tracker which is implemented by an adaptive Rao-Blackwellised particle filter (RBPF). The SVM is re-trained online on new samples taken from previous predicted positions. We use the effective sample size to decide when the classifier needs to be re-trained.

Position hypotheses for the tracked object are the result of a clustering procedure applied on the set of particles.

The algorithm has been tested on challenging video sequences presenting strong changes in object appearance, illumination, and occlusion. Experimental tests show that the presented method is able to achieve near real-time performances with a good precision on standard video sequences.

8301-32, Session 7

A modular real-time vision system for humanoid robots

A. L. Trifan, A. J. Neves, B. Cunha, N. Lau, Univ. de Aveiro (Portugal)

For both humans and robots vision is a very important sense that has the task of interpreting spatial data, indexed by more than one dimension. For a humanoid robot, a robust vision system should be able to provide accurate information about the environment and a precise description of the objects of interest. We present an implementation of a vision system for a humanoid robot designed to perform in color-coded environments. From acquiring images, processing them and detecting the objects of interest, all the algorithms have been tested on the NAO soccer playing humanoid robot as well as on a Bioloid robot designed to navigate through a maze with the help of color-coded clues. For these robots the world is simplified to a number of colors that are meaningful in the mentioned contexts. The vision system that we propose can perform in real time and it has proven its practical efficiency. Moreover, we present an innovative algorithm for self-calibration of the most important intrinsic parameters of the camera, as well as two external applications developed for debugging and color calibration. One of the most important features of the vision system that we propose is its modularity, which allows its use with a wide range of robotic platforms.

8301-33, Session 8

Radial polar histogram approach to obstacle avoidance and path planning for robotic cognition and motion control

P. Wang, N. R. Keyawa, C. Euler, C. T. Lin, California State Univ., Northridge (United States)

The most fundamental element of autonomous mobile robotics is the robot's ability to detect and avoid obstacles while traveling to desired locations. Since ground vehicles frequently maneuver in curvilinear motions, an obstacle avoidance algorithm that can provide a desired instantaneous radius and velocity results in more accurate motion execution. This type of obstacle avoidance algorithm, which has been implemented in California State University Northridge's Intelligent Ground Vehicle (IGV), is known as Radial Polar Histogram (RPH). RPH utilizes raw data in a polar histogram that is read from a laser range finder and a camera. This data is filtered for obstacle grouping and edge detection. A concavity function is utilized to eliminate paths that may cause the robot to collide with obstacles. Within the concavity

function, several open blocks are defined and a cost function is utilized to determine a desired heading. This heading, along with concavity and navigational waypoints, determines the best open block. The left and right radial edges of the open block are determined and put into a step by step debug function to output a range of possible radial paths. The radial path optimization, velocity, and acceleration functions are implemented to determine a desired radius and velocity.

8301-34, Session 8

Optimising a mobile robot control system with GPU acceleration

N. Tuck, M. E. McGuinness, F. Martin, Univ. of Massachusetts Lowell (United States)

The International Ground Vehicle Competition (IGVC) Autonomous Challenge requires a control program that performs a number of different computationally intensive tasks. Our initial control program design can not run comfortably in real time on a standard multi-core processor in our budgeted power envelope. This paper describes the process and results of profiling this control program and porting appropriate computations to run accelerated on an AMD Radeon graphics processor. In order to GPU-accelerate some of these computations, we used the Bacon programming system, a tool developed by one of the authors for building compute kernels to run on OpenCL-compatible graphics processors. OpenCL-based GPU acceleration can provide an order of magnitude speedup over traditionally optimised code. Compared to using OpenCL code directly, Bacon is significantly easier to use and can provide as much as a factor of two additional speedup. This paper describes the accelerated implementation of the control software built on the Robot Operating System (ROS) running in the Gazebo 3D simulator. Speedups and, in some cases, quality improvements of the vision processing are demonstrated compared to the routines available in OpenCV. The performance improvements of the complete system on our simulation of the IGVC Autonomous Challenge are described.

8301-35, Session 8

Design and realization of an intelligent ground vehicle with modular payloads

M. A. Akmanalp, R. M. Doherty, J. Gorges, P. Kalauskas, E. Peterson, F. Polido, S. S. Nestinger, T. Padir, Worcester Polytechnic Institute (United States)

In June 2011, Worcester Polytechnic Institute's (WPI) unmanned ground vehicle participated in the 8th Annual Robotic Lawnmower and 19th Annual Intelligent Ground Vehicle Competitions back-to-back. This paper details the two-year design and development cycle for WPI's intelligent ground vehicle, Prometheus. The on-board intelligence algorithms include lane detection, obstacle avoidance, path planning, world representation and waypoint navigation. The authors present experimental results and discuss practical implementations of the intelligence algorithms used on the robot.

8301-36, Session 8

Navigating a path delineated by colored flags: an approach for an IGVC 2011 requirement

A. Szmatula, M. Parrish, M. Krishnan, M. Paulik, U. Mohammad, C. Luo, Univ. of Detroit Mercy (United States)

A new requirement for the Autonomous Challenge portion of the 2011 Intelligent Ground Vehicle Competition (IGVC) is to navigate between red and green colored flags that form a lane within a lane within the course. This addition makes for an interesting challenge in that it forces teams to incorporate color-based image characteristic extraction combined with possible LADAR detection of a very small feature in

the scene. Furthermore, since the flags are small and flexible, and thus susceptible to the wind, they present a variable cross section to the camera system making false positives and false negative identifications common. Thus in addition to the image processing challenge associated with identifying and classifying flags, the teams must address navigation algorithm updates necessary to provide smooth and reliable robot motion in the presence of higher obstacle uncertainty. This work then discusses an algorithm for detection and localization of colored flags followed by modification of a Vector Field Histogram (VFH) algorithm to navigate the robot appropriately. Results from simulation and field test studies will be presented and performance discussed.

8301-37, Session 8

Navigating with VFH: a strategy to avoid traps

C. Luo, M. Krishnan, M. Paulik, U. Mohammad, Univ. of Detroit Mercy (United States)

The IGVC Navigation Challenge course configuration has evolved in complexity to a point where use of a simple reactive local navigation algorithm presents problems in course completion. A commonly used local navigation algorithm, the Vector Field Histogram (VFH), is relatively fast and thus suitable when computational capabilities on a robot are limited. One of the attendant disadvantages of this algorithm is that a robot can get trapped when attempting to get past a concave obstacle structure. The Navigation Challenge course now has several such structures, including some that partially surround waypoints. Elaborate heuristics are needed to make VFH viable in such a situation and their tuning is arduous.

An alternate approach that avoids the use of heuristics is to combine a dynamic path planning algorithm with VFH. In this work, the D*Lite path planning algorithm is used to provide VFH with intermediate goals, which the latter then uses as stepping stones to its final destination. Results from simulation studies as well as field deployment are used to illustrate the benefits of using the local navigator in conjunction with a path planner.

8301-38, Poster Session

Measurement of noises and modulation transfer function of cameras used in optical-digital correlators

N. N. Evtikhiev, S. N. Starikov, P. A. Cheryomkhin, V. V. Krasnov, National Research Nuclear Univ. MEPhI (Russian Federation)

Hybrid optical-digital systems based on diffractive correlator are being actively developed. Such systems can be used for pattern recognition, image encryption, imaging with aberrations correction and extending depth of filed. For wider propagation of this kind of systems, the possibility of using inexpensive photo and video cameras is important. To correctly estimate capabilities of application of cameras of different types in optical-digital correlation systems the knowledge of modulation transfer function (MTF) and light depended temporal and spatial noises is required.

The method for measurement of 2D MTF is presented. The method based on random target method but instead of a random target the specially created target with flat power spectrum is used. It allows to measure MTF without averaging 1D Fourier spectra over rows or columns as is in the random target method and to achieve all values of 2D MTF instead of just two orthogonal cross-sections.

The simple method for measuring the dependence of camera temporal noise on light signal value by shooting a single scene is described. Light spatial noise is characterized by photo-response non-uniformity (PRNU) of photo sensor. Procedure for obtaining camera's light spatial noise portrait (array of PRNU values for all photo sensor pixels) is presented.

Results on measurements of MTF and light depended temporal and spatial noises for consumer photo camera, machine vision camera and video-surveillance camera are presented.

8301-39, Poster Session

A phase-space approach for detection and removal of rain in video

V. Santhaseelan, K. V. Asari, Univ. of Dayton (United States)

Nowadays, the widespread use of computer vision algorithms in surveillance systems and autonomous robots has increased the demand for video enhancement algorithms. In this paper, we propose an algorithm based on phase congruency features to detect and remove rain and thus improve the quality of video. We make use of the following characteristics of rain streaks in video in order to detect them: (1) rain streaks do not occlude the scene at all instances, (2) all the rain streaks in a frame are oriented in a single direction, and (3) presence of rain streak at a particular pixel causes a positive change in intensity. Combining all these properties we are able to detect rain streaks in a particular frame using phase congruency features. The pixels in a frame which are identified as rain streaks are then replaced using the pixel information of its spatial and temporal neighbors which are not affected by rain. Using this method we are able to remove rain of medium density from videos even when complex camera movement is involved. We also propose an adaptive compensation mechanism for scene restoration where the number of frames required for compensation changes automatically based on the density of rain.

8301-40, Poster Session

Intelligence algorithms for autonomous navigation in a ground vehicle

S. J. Petkovsek, A. Norton, Trinity College (United States)

"Q" is Trinity College Robot Study Team's (TCRST) entry in the Intelligent Ground Vehicle Competition (IGVC) held at Oakland University in Auburn Hills, Michigan. Trinity has participated in the competition for eleven years, while Q is in its sixth iteration. IGVC has four areas in which to compete, including the Design challenge, the Navigation challenge, the Autonomous challenge, and the JAUS Interoperability challenge. In each of these challenges, except for the Design challenge where the system itself is judged, Q must be able to autonomously navigate an outdoor environment. To successfully navigate the course, Q must avoid painted lines, traverse ramps and sand pits, avoid obstacles, and reach waypoints as quickly as possible, as each run is timed. For the JAUS challenge, Q must obey commands from the judge's server using Joint Architecture for Unmanned Systems protocol. This paper will outline Q's intelligence algorithms and image processing techniques for completing the course. The methods Q uses for path planning and path history will be discussed, as special considerations are taken to navigate successfully through dense obstacle regions where failures such as getting stuck in a corner or turning back the wrong direction on the course are more likely. The manner in which Q distinguishes impassable white lines from passable, bright surfaces such as a painted ramp or sand pit will also be examined. Last, the process by which Q navigates to GPS waypoints will be explained.

8301-41, Poster Session

Hierarchical, multi-level image mosaicing for autonomous navigation of UAV

S. Park, D. Ghosh, N. Kaabouch, R. Fevig, W. Semke, The Univ. of North Dakota (United States)

A novel algorithm for hierarchical multi-level image mosaicing for autonomous navigation of UAV is proposed. The main contribution of the proposed system is the minimization of the error accumulation propagated along the frames, by incrementally building a long-duration mosaic on the fly which is hierarchically composed of short-duration mosaics. The proposed algorithm fulfills the real-time processing requirements in autonomous navigation as follows. 1) Causality: the current output of the mosaicing system depends only on the current and/or previous input frames, contrary to existing offline mosaic

algorithms that depend on future input frames as well. 2) Learnability: the algorithm autonomously analyzes/learns the scene characteristics and classifies them into the 'manmade structural area' type vs. the 'natural vegetation field' type by the Bayesian learning and inference. 3) Adaptability: the system automatically adapts itself to the scene change and chooses the proper methods for feature selection (i.e., the fast but unreliable LKT vs. the slow but robust SIFT) and geometric image transform (i.e., the restrictive affine transform vs. the more relaxed perspective transform.)

The evaluation of our algorithm with the extensive field test data involving several thousand airborne images shows the significant improvement in processing time, robustness and accuracy of the proposed algorithm.

8301-42, Poster Session

A diffraction-limited 10-mm-aperture adaptive lens

R. Batchko, Holochip Corp. (United States)

We present an electrically-actuated adaptive fluidic lens having a 10-mm clear aperture and 6-diopter range in focal power. The lens employs biconvex membranes encasing a fluid. A piezoelectric actuator provides a 150-to-1000-mm focal length range while drawing less than 1 mW. The lens housing has an outer diameter of 1-inch and lens center-thickness is less than 1 mm. Polychromatic MTF exceeds 90% of the diffraction limit over the full 10-mm aperture, outperforming commercial spherical glass lenses over much of its dioptric range.

Conference 8302: Imaging and Printing in a Web 2.0 World III

Wednesday 25 January 2012

Part of Proceedings of SPIE Vol. 8302 Imaging and Printing in a Web 2.0 World III

8302-01, Session 1

Organizing visual moments for sharing and reflections: VisR

R. C. Jain, Univ. of California, Irvine (United States)

In this presentation we will discuss concepts and architecture of the system that we are building towards organizing all personal media for sharing and reflections. Our event-based organization of data enables access to large volumes of media data that is becoming common. We will also demonstrate our working system.

8302-02, Session 1

The role of digital presses, Web 2.0, and mobile in enabling the internet of things

T. F. Rodriguez, Digimarc Corp. (United States)

No abstract available

8302-03, Session 2

Mobile image processing for fashion marketplace

M. Chandra, G. Golwala, C. Pungaliya, Poshmark, Inc. (United States)

No abstract available.

8302-04, Session 2

Mobile capture: the end of the photocopier?

M. J. Gormish, Ricoh Innovations, Inc. (United States)

We investigated mobile phone usage for image capture in the workplace via surverys and interviews. Our surveys show smart mobile phones are being used increasingly for image capture and other non-phone call activities. We focus on information capture with mobile devices and the image processing necessary to make those camera images as usable as images from flat bed scanners. Algorithms to clean up whiteboards and paper documents can save printing costs and proper connectors can reduce the need to print entirely. In addition we capture information not available to scanners and make that information useful for business purposes and for printing. Ricoh has released multiple mobile phone applications and announced an eWriter tablet all of which change imaging and printing via cloud services. Some of these will be demonstrated.

8302-05, Session 2

From scenes to screens: connected digital imaging

Z. Gillat, Eye-Fi (United States)

No abstract available.

8302-06, Session 3

Text documents as social networks

H. Balinsky, Hewlett-Packard Labs. (United Kingdom); A. Balinsky, Cardiff Univ. (United Kingdom); S. J. Simske, Hewlett-Packard Co. (United States)

The extraction of keywords and features is a fundamental problem in text data mining. Document processing applications directly depend on the quality and speed of the identification of salient terms and phrases. Applications as disparate as automatic document classification, information visualization, filtering and security policy enforcement all rely on the quality of automatically extracted keywords.

Recently, a novel approach to rapid change detection in data streams and documents has been developed. It is based on ideas from image processing and in particular on the Helmholtz Principle from the Gestalt Theory of human perception. By modelling a document as a one-parameter family of graphs with its sentences or paragraphs defining the vertex set and with edges defined by Helmholtz's principle, we demonstrated that for some range of the parameters, the resulting graph becomes a small-world network.

In this article we investigate the natural orientation of edges in such small world networks. For two connected sentences, we can say which one is the first and which one is the second, according to their position in a document. This will make such a graph look like small WWW-type network and PageRank type algorithms will produce interesting ranking of nodes in such a document.

8302-08, Session 3

HP2.ME URL shorten service

H. Zhang, W. X. Wu, Y. Zhang, C. Zheng, Hewlett-Packard China Co., Ltd. (China); Q. Lin, J. Liu, Hewlett-Packard Labs. (United States)

HP2.ME is an url shorten service which provide by HP. Different than other existed URL shorten services, HP2.me not only provides URL shorten but also helps to extract valuable contents from any kind web page and return them back in the right format which can be properly view in different mobile devices or PC . Through browser engine webkit and Dom tree analysis, we get the original HTML page from URL and analysis the DOM elements 'geometric info to figure out the valuable contents in the page. Re layout the extract content to fit the request device screens. In this paper, we present the primary methods and algorithm to calculate the main content in web page and how we layout the extract contents to fit the different screen size.

8302-09, Session 3

HP Smartprint

H. Zhang, Z. Liu, Y. Yuan, G. J. Liao, Hewlett-Packard China Co., Ltd. (China); Q. Lin, J. Liu, Hewlett-Packard Labs. (United States)

In HP new web browser plug-in print tool, Smart Print, we apply novel methods to address complex web print problems. Through DOM element geometric calculation, Smartprint can auto extract value content from any kind of web pages. For keeping the output layout in a good way, we apply the original CSS to help re layout the content elements which have been marked as valuable content. Other than them, we also allow user to manual easily crop or remove any contents they want from the original web pages through drag and drop. During the dragging or dropping, user can directly see what content will be cover or removed through real time affinity calculation. HP Smartprint

has been shipped as a formal product of HP which can be downloaded from HP web print web page.

8302-10, Session 4

Kind of images in printed photo books

R. Fageth, P. Schuetz, T. Wagner, CeWe Color AG & Co. OHG (Germany)

Preserving memories combined with story telling using photo books are substituting more and more classical 4*6 prints and silver halide posters. Digital printing via electro photography and ink jet is also substituting more and more classical silver halide printing technology as the dominant production technology for these kinds of products. But there is still a remarkable number of people who want the photo quality in photo books on real photo paper (AgX) or at least order their photo book in a glossy version printed via electro photography printers.

In this paper we show the usage of these three different output possibilities as function of the event documented in the printed product as well as the ordering behavior for classical silver halide prints as function of the events.

This paper also analyses the usage of images taken by digital still cameras as well as with mobile phones in printed photo books. We evaluate the number of cameras used in photo books as well as combinations of SLRs, point and shoot and mobile phones in all of its possible combinations. We also try to analyze these combinations with the themes of the photo books, e.g. in books telling the story about a birthday one finds more images of mobile phones than in books of a wedding.

We also verify if the duration of the event (time difference of the 1st image taken and the last image taken) have an influence of the format/size of the photo book being ordered.

Additional parameters are verified, such as if the geo information (longitude and latitude) is included and used, e.g. in order to place a map of the location in the book.

8302-11, Session 4

SmartFit: automatic photo fitting for variable data printing

Z. Karni, A. Gaash, Hewlett-Packard Labs. Israel Ltd. (Israel)

We present an algorithm for automatic smart image fitting: changing the size of an image so that it may fit "naturally" within a given frame. As the frame's dimensions and aspect ratio will generally differ from that of the image, the algorithm preserves important details in their original aspect ratio, while less important details undergo more substantial deformations. This problem is useful for many commercial print and web applications. One example is an automatic retargeting of images in web pages where the page should fit smoothly into different sizes such as of cellular phones, Touch-Pads, regular monitors and printed pages. Another example is the HP SmartStream Designer, which is a tool to create variable and personalized content documents. The smart image fitting algorithm has been implemented within the SmartStream Designer, and has been presented at IPEX 2010 (Birmingham) and Graph Expo 2010 (Chicago), where it received highly positive reactions.

8302-12, Session 4

All new custom path photo book creation

W. H. Wang, R. Muzzolini, Shutterfly (United States)

In this paper, we present an all new custom path to allow consumers to have full control to their photos and the format of their books, while providing them with guidance to make their creation fast and easy. The users can choose to fully automate the initial creation, and then customize every page. The system manage many design themes along with numerous design elements, such as layouts, backgrounds, embellishments and pattern bands. The users can also utilize photos from multiple sources including their computers, Shutterfly accounts, Shutterfly Share sites and Facebook. The users can also use a photo as background, add, move and resize photos and text - putting what they want where they want instead of being confined to templates. The new path allows users to add embellishments anywhere in the book, and the high-performance platform can support up to 1,000 photos per book and up to 25 pictures per page. The path offers either Smart Autofill or Storyboard features allowing customers to populate their books with photos so they can add captions and customize the pages.

8302-13, Session 4

Investigation of the role of aesthetics in differentiating between photographs taken by amateur and professional photographers

S. Xue, Purdue Univ. (United States); Q. Lin, D. Tretter, S. Lee, Hewlett-Packard Labs. (United States); Z. Pizlo, J. P. Allebach, Purdue Univ. (United States)

In additional to previous low level features, we utilized high level rules commonly accepted by the photography community as the new features. By doing so, we can not only train the system to regard sharp, properly-lighted photos as good photos, but can also take into account the fact that photos need to be composed in an aesthetic way to be qualified as professional ones. Composition rules such as balance, rhythm, and visual paths are the main focus of the new features.

Photos from on-line sharing website serve as the source for training and testing. We used crawlers to download images, their user ratings, and other statistics. Machine learning and data mining techniques are implemented to train the system based on the features. Classification results are compared with those from methods proposed by previous papers.

In order to avoid bias from using opinions of users at a specific website, and to get viewpoints from professional photographers, we also ask professional photographers to rate some of the photos randomly chosen from the testing dataset to see if the classification results from the algorithm confirm with the viewpoints of professional photographer.

8302-14, Session 5

Learning from user data in Facebook

J. Yang, Facebook Inc. (United States)

No abstract available.

8302-15, Session 5

Measuring engagement effectiveness in social media: a time-sensitive methodology

T. Sun, L. Li, W. Peng, Xerox Corp. (United States)

The effectiveness measurements in any traditional CRM system (e.g. customer call center) are primarily based on the duration time per call and/or number of answered calls by service agents per day. In social media environment, we have a multitudes of conversational contents and underlying social graph data widely available. The effectiveness of social engagement can be measured by analyzing the content of conversations, the sentiment of customers and their social impacts in underlying social network. In this paper, we present a time-sensitive topic-relevant and sentiment-dependent metric to measure the effectiveness of service agents' engagement of customer care in social media. Specifically, we quantify the engagement effectiveness by making use of a pyramid kernel, which measures the time-sensitive distance of an employee's engagement within a multi-resolution space in terms of polarized topical changes. The proposed metric is able to capture the engagement behavior of agents by analyzing the conversations between the agents and the customers in social media environment, e.g., Twitter. In summary, the main contribution of this

paper is a new time-sensitive pyramid kernel function based on implicit topical and sentimental correspondences among daily conversations, which enables discriminative evaluation for agent engagement in social customer care.

8302-16, Session 5

Building a scalable storage for images on a social network

J. Medrano Navarro, Tuenti Technologies (Spain)

Images are one of the key components on a social network. A storage for images needs to be highly scalable, and provide redundancy, high availability and the ability to grow its size. Efficiency is also required so that disk storage and the need for processing power should be minimized.

Our storage uses a CDN (Content Delivery Network) as a web cache that allows us to meet high throughput requirements. When an image is not cached in the CDN, it is requested to the IRL (Image Routing Layer) that is in charge of finding its physical location. If IRL is not able to retrieve the image from one of the locations it can get it from the other copies preventing the CDN and the user from noticing the miss. If the requested size is not available on the storage it's resized by the IRL. Expensive operations like finding the physical location or resizing are only done when there's a cache miss on the CDN.

The physical storage is split in homogeneus buckets that are spread among the storage servers. Growth strategy is performed by adding more storage servers and rebalancing buckets towards them. Rebalancing not only provides free space on full servers but allows the upload bandwidth to increase because there will be fewer buckets, and so fewer uploads per server.

8302-17, Session 5

Color correction of smartphone photos with prior knowledge

Y. Zhao, S. Wang, Xerox Corp. (United States); J. Jiang, Rochester Institute of Technology (United States)

The color image is the result of a complex interaction between three major components: scene, illumination and camera sensor. Typically, the effect of the scene properties and the influence of the illumination are tightly coupled and very hard to separate from each other. Unlike the human visual system, which has the property of perceiving the color of the object to remain constant when the level and color of the illumination are changed, the digital camera completely lacks this property. Although there are some build-in features or functions on the sensor to set up white point and exposure time, in some cases, the photos are still not satisfactory because of unwanted color cast or poor contrast. This paper presents a method for automatically removing color cast and adjusting lightness contrast of digital photos due to the incorrect in-camera setting with prior knowledge for estimating capturing illumination. The proposed method is to use a person's head characteristics as the prior knowledge. One decent head image was taken in advance under sufficient illumination condition and saved as the standard image. Whenever the person snaps a target photo, he also takes an image of his own head as the reference image. These two head images are compared to estimate the capturing illumination and build the proper color transformation. After that, the transformation is used to correct the target photo taken under the similar illumination.

8302-18, Session 5

XML data compression in web publishing

R. Qiu, W. Hu, Z. Tang, X. Lu, Peking Univ. (China); L. Zhang, Peking Univ. Founder Group Corp. (China)

XML is widely used in various document formats on the web. But it has caused negative impacts such as expensive document distribution time over the web, and long content jumping and rendering delay,

especially on mobile devices. Hence we proposed a Schema-based efficient queryable XML compressor, called XTrim, which significantly improves compression ratio by utilizing optimized information in XML Schema while supporting efficient queries. Firstly, XTrim draws structure information from XML document and corresponding XML Schema. Then a novel technique is used to transform the XML tree-like structure into a compact indexed form to support efficient queries. At the same time, text values are obtained, and a language-based text trim method (LTT) that facilitates language-specific text compressors is adopted to reduce the size of text values in various languages. In LTT a word composition detection method is proposed to better process text in non-Latin languages. To evaluate the performance of XTrim, we have implemented a compressor and query engine prototype. Via extensive experiments, results show that XTrim outperforms XMill and existing queryable alternatives in terms of compression ratio, as well as the query efficiency. By applying XTrim to documents, the storage space can save up to 30% and the content jumping and rendering delay is reduced to less than 100ms from 4 seconds.

8302-19, Session 6

Layout hierarchies for interactive design reuse

D. S. Greig, A. A. Hunter, D. N. Slatter, Hewlett-Packard Labs. (United Kingdom)

In their 2010 paper [1] the authors describe in overview the basis of an easy to use web service for high quality authorship aimed at occasional users with limited aesthetic design skills. The system allows users to drag and drop material from existing magazines into the magazine under construction and automatically rearranges pages affected by the new content to produce a new layout. The system has a shuffle function that allows the user to choose from further additional layouts of the same material.

This paper describes an important aspect of the previous magazine design system. The problem is that a very wide space of desirable layouts exists, and individual layout engines can only span a small segment of that space. If available layout engines can be combined into robust hierarchies then much more complex designs are possible.

A hierarchical document structure is described that represents complex layouts and supports user edit operations interacting with a set of layout engines to produce aesthetically pleasing outcomes.

[1] Hunter, A., Slatter, D., Greig, D., "Web Based Magazine Design for Self Publishers", Proc SPIE 7879, 789902 (2011)

8302-20, Session 6

Automatic page composition with combined cropping and layout technologies

A. A. Hunter, D. N. Slatter, D. S. Greig, Hewlett-Packard Labs. (United Kingdom)

Automatic layout algorithms simplify the composition of image-rich documents, but they still require users to have sufficient artistry to supply well cropped and composed imagery. Combining an automatic cropping technology with a document layout system enables better results to be produced faster by less-skilled users. This paper reviews prior work in automatic image cropping and automatic page layout and presents a case for a combined crop and layout technology. The authors describe one such technology in a system for interactive publication design by amateur self-publishers. They show that providing an automatic cropping system with additional information about the layout context can enable it to generate a more appropriate set of ranked crop options for a given image. Similarly, they show that providing an automatic layout system with sets of ranked crop options for images can enable it to compose more appropriate page layouts.

8302-21, Session 6

Psychophysical evaluation of document visual similarity

A. Satkhozhina, I. Ahmadullin, Purdue Univ. (United States); S. Lee, Hewlett-Packard Co. (United States); Z. Pizlo, J. P. Allebach, Purdue Univ. (United States)

Being able to automatically compare document layouts, and classify and search documents based on their visual appearance is a very important task in many applications that manage large databases. Most of these applications are targeted for customer use. Therefore it is essential to understand how people interpret visual similarity between documents and what document features play main roles in determining visual similarity. The purpose of our research is to determine document features that are the most critical in deciding visual similarity for human perception. The text content will be ignored since we are looking for visual similarity only. In this research, psychophysical experiments will be conducted to identify what document features are the most important for human perception in deciding the visual similarity. For our study, we will use a database of documents that were collected from various Purdue organizations. Sixty document pages will be selected from the database and organized into five test sets. Also, five query document pages will be chosen from the database. The participant will be asked to place the set of printed document pages around a query document page in such way that the distance between the query document and the document page is inversely proportional to the visual similarity between them. Results will be recorded and analyzed to determine the document features that are the most important in deciding visual similarity in human perception.

Words: 228

8302-22, Session 6

Similarity pyramid: browsing a document database with respect to visual similarity

I. Ahmadullin, J. P. Allebach, Purdue Univ. (United States)

In this presentation we will talk about organization of single-page documents in a 3-D hierarchical structure called a similarity pyramid. The pyramid is constructed from a document database that is embedded on a low dimensional surface with the help of a nonlinear dimensionality reduction algorithm called Isomap. The mapping algorithm preserves similarity distances between documents by mapping documents that are close to each other in a feature space to points on low dimensional surface that are close to each other. Higher levels of the pyramid consist of document image icons that represent a large group of roughly similar documents, whereas lower levels contain document image icons representing small groups of very similar documents. A user can browse the database by moving along a certain level of a pyramid by moving between different levels.

8302-23, Session 6

Automatic design of magazine covers

A. Jahanian, Purdue Univ. (United States) and Hewlett-Packard Labs. (United States); J. P. Allebach, Purdue Univ. (United States); Q. Lin, J. Liu, N. Damera-Venkata, E. O'Brien-Strain, S. Lee, J. Fan, D. Tretter, Hewlett-Packard Labs. (United States)

In this paper we propose a system for the automatic design of magazine covers that quantifies a number of concepts from art and aesthetics. We are interested in knowing how professional designers think in the process of magazine creation and how they create a cover considering both form and functionality. If we can quantify and model this process, we can support non-designers to produce some designs close to the work of professionals. Accordingly, we have collaborated with professional designers, magazine art directors and editorial boards, and journalists to study the process of magazine cover design. We have then implemented what we have learned in the framework of a software tool. Our framework divides the task of design into three main modules: visual balance in layout, design of color for masthead and cover lines, and typography of cover lines. At this stage of our work, feedback of professional designers on our designs implies that our results are congruent with their intuition. However, our future work includes further evaluation of our automatic designs by users. We also endeavor to quantify the idea of style in design. The style of a magazine, which is heavily influenced by magazine type, has a major influence on the design of a magazine cover. We are interested in extending our model of color design to include style of the magazine. We are also interested in aligning the typeface with the style of the magazine.

8302-24, Session 7

Sentiment analysis and live customer intelligence from social channels

M. Hsu, Hewlett-Packard Labs. (United States)

The proliferation of online communities and social networks created a new channel, the Social Channel, for consumers to share their experiences and voice their opinions about everything. The dynamic social channel engages customers, prospects, partners, influencers, and employees-touching virtually every key constituent in an organization's value chain. These channels are reshaping how customers evaluate and choose products, how brands are perceived, and how business processes interact with the customers. The ability to model, identify, understand, measure, and timely react to sentiment, opinion, preference, and behavior expressed in this unstructured content is transforming the way companies interact and manage relationships with their customers. In this session we will present recent work in the area of live business intelligence at HP Labs. In particular we will discuss live customer intelligence, the use of streaming, real-time text analytics to study user sentiment and feedback collected from review forums, online retailers, blogs, and social media outlets such as Twitter and Facebook. A number of use cases will also be illustrated.

8302-25, Session 7

Automatic content recognition for the next-generation TV experience

X. Lin, Vobile, Inc. (United States)

In recent years internet is touching almost every corner of our life and TV is no exception. First, a new generation of Smart TVs has been introduced. Second, applications running on mobile devices (so called "second-screen apps") have significantly enriched TV watching experience. As an enabler of content-aware TVs and apps, automatic content recognition (ACR) is attracting a lot of attention recently. This paper presents an overview of ACR in this context. It attempts to answer a number of questions: Why do we need ACR for the next generation TV experience? What is the relationship between ACR and existing technologies? What are the unique requirements and challenges on ACR in those applications? What are typical architectures to implement ACR? It also describes the existing products and prototypes in this space.

8302-26, Session 7

Marketing image categorization using hybrid human-machine combinations

N. Gnanasambandam, Xerox Corp. (United States)

No abstract available.

8302-27, Session 7

Practical experiences in analog to digital content transformation

P. D. Reddy, Hewlett-Packard Labs. (United States)

The initial goal was to take out-ofcopyright texts and make them available to read online (for free) or to own (via Print-ondemand). The scanned page images of these works capture all the nuances of the original printing, however supporting rapid download, search, accessibility, and re-flowable text requires converting the analog page images into digital text in an electronic publication format. Unfortunately commercially-available Optical Character Recognition engines cannot provide adequate accuracy without expensive and time consuming manual intervention. In this paper, we describe practical techniques for achieving a high degree of accuracy, scaling to process large volumes efficiently, and web interfaces for seamless content consumption and efficient error correction.

8302-28, Session 7

Global image analysis to determine suitability for text-based image personalization

H. Ding, R. Bala, Z. Fan, Xerox Corp. (United States); C. A. Bouman, J. P. Allebach, Purdue Univ. (United States)

Lately, image personalization is becoming an interesting topic. Images with variable elements such as text usually appear much more appealing to the recipients. In this paper, we describe a method to pre-analyze the image and automatically suggest to the user the most suitable regions within an image for text-based personalization. The method is based on input gathered from experiments conducted with professional designers. It has been observed that regions that are spatially smooth and regions with existing text (e.g. signage, banners, etc.) are the best candidates for personalization. This gives rise to two sets of corresponding algorithms: one for identifying smooth areas, and one for locating text regions. Furthermore, based on the smooth and text regions found in the image, we derive an overall metric for "suitability for personalization".

8302-29, Session 7

Chrominance watermark embed using a full-color visibility model

A. M. Reed, Digimarc Corp. (United States)

No abstract available

8302-30, Session 7

Document image orientation based on both text and image

Y. Sun, C. Liu, X. Ding, Tsinghua Univ. (China); Z. Fan, F. Tse, Xerox Corp. (United States)

This paper investigated the problem of orientation detection for document images with Chinese, Japanese, or Korean characters. These images may be in four orientations: right side up, up-side down, 90° and 270° rotated counterclockwise. First, we presented the structure of text-recognition-based orientation detection algorithm. Text line verification and orientation judgment methods were mainly discussed, afterwards multiple experiments were carried. Distance-difference based text line verification and confidence based text line verification were proposed and compared with methods without text line verification. Then, a picture-based orientation detection framework was adopted for the situation where no text line was detected. This high-level classification problem was solved by relatively low-level vision features including Color Moments (CM) and Edge Direction Histogram (EDH), with distant-based classification scheme. Finally, confidence-based classifier combination strategy was employed in order to make full use of the complementarity between different features and classifiers. Experiments showed that both text line verification methods were able to improve the accuracy of orientation detection, and picture-based orientation detection had a good performance for no-text image set.

Conference 8303: Media Watermarking, Security, and Forensics 2012

Monday-Wednesday 23-25 January 2012

Part of Proceedings of SPIE Vol. 8303 Media Watermarking, Security, and Forensics IV

8303-01, Session 1

Security threat to media security applications based on scale-space feature extraction

C. Hsu, Academia Sinica (Taiwan); C. Lu, Institute of Information Science (Taiwan); S. Pei, National Taiwan Univ. (Taiwan)

Scale-space image feature extraction (SSIFE) has been widely adopted in multimedia security and retrieval due to its powerful resilience to attacks. However, the security threat to SSIFE-based media security applications is relatively unexplored. In this paper, the security threat, composed of a constrained-optimization keypoint inhibition attack (KIHA) and a keypoint insertion attack (KISA), is specifically designed for scale-space feature extraction methods such as SIFT and SURF.

The principle of KIHA is to make a fool of feature extraction protocols in that the detection rules are purposely violated so that no local maximum can be found around in a local region. On the other hand, KISA is designed to create the false positive problem. Our method is evaluated and compared with Do et al.'s method (ACM MM'10), which also figures out the weakness of our previous work (ACM MM'09). In addition, our proposed security threat is applied to an image copy detection method together with a web-scale image database for performance evaluation.

8303-02, Session 1

Robust image obfuscation for privacy protection in Web 2.0 applications

A. Poller, M. Steinebach, H. Liu, Fraunhofer-Institut für Sichere Informations-Technologie (Germany)

In this paper we present two approaches to robust image obfuscation based on permutation of image regions and channel intensity modulation. The proposed concept of robust image obfuscation is a step toward client-to-client security in Web 2.0 applications. It helps to protect the privacy of the users against threats caused by image robots and certain web applications which extract biometric and other features from images for data-linkage purposes. The approaches described in this paper consider that images uploaded to web 2.0 applications pass several transformations until the receiver downloads them. In contrast to existing approaches our focus is on

usability, therefore the primary goal is not a maximum of security but an acceptable trade-off between securityand resulting quality.

8303-03, Session 2

Improved Fourier domain template and patchwork embedding using spatial masking

H. Liu, M. Steinebach, Fraunhofer-Institut für Sichere Informations-Technologie (Germany)

Robustness against distortions caused by common image processing is one of the essential properties for image watermarking to be applicable in real-world applications. Typical distortions include lossy JPEG compression, filtering, cropping, scaling, rotation, and so on, among which geometric distortion is more challenging. Even slight geometric distortion can totally fail the watermark detection through de-synchronization. Another important property is the watermark payload. Although one-bit watermark is widely used in research work for algorithm testing and evaluation, only checking whether a specific watermark exists does not meet the requirement of many practical applications. This paper presents a practical robust image watermarking algorithm which combines template embedding and patchwork watermarking in Fourier domain. The embedded template enables the necessary robustness against geometric distortions and the patchwork approach provides a reasonable watermark payload which can meet the requirement of most applications. A spatial perceptual mask is used to reshape the embedded energy after it is inverted to the spatial domain, which significantly improves the image quality and enhances the robustness of both template and watermark. Implementation issues and solutions, e.g. fine-tuning of embedding energy of individual pixels, are also discussed. Experimental results demonstrate the effectiveness and practicability of the proposed algorithm.

8303-04, Session 2

Ranking search for probabilistic fingerprinting codes

M. Schäfer, W. Berchtold, M. Steinebach, Fraunhofer-Institut für Sichere Informations-Technologie (Germany)

Digital transaction watermarking today is a widely accepted mechanism to discourage illegal distribution of multimedia. The transaction watermark is a user-specic message that is embedded in all copies of one content and thus makes it individual. Therewith it allows to trace back copyright infringements. One major threat on transaction watermarking are collusion attacks. Here, multiple individualized copies of the work are compared and/or combined to attack the integrity or availability of the embedded watermark message. One solution to counter such attacks are mathematical codes called collusion secure fingerprinting codes. Problems arise when applying such codes to multimedia les with small payload, e.g. short audio tracks or images. Therefore the code length has to be shortened which increases the error rates and/or the effort of the tracing algorithm. In this work we propose an approach whether to use as an addition to probabilistic fingerprinting codes for a reduction of the effort and increment of security, as well as a new seperate method providing shorter codes at a very fast and high accurate tracing algorithm.

8303-05, Session 2

Stereoscopic watermarking by horizontal noise mean shifting

J. Lee, H. Kim, H. Choi, S. Choi, H. Lee, KAIST (Korea, Republic of)

Depth-image-based rendering (DIBR) is a method to represent a stereoscopic content. The DIBR consists of a monoscopic center view and an associated per-pixel depth map. Using these two components and given depth condition from an user, the DIBR renders left and right views. The advantages of DIBR is numerous. The user can choose not only the monoscopic or stereoscopic view selectively, but also the depth condition what he prefers when he watches a stereoscopic content. However, in the view of copyright protection, since not only the center view but also each left or right view can be used as a monoscopic content when they are illegally distributed, the watermark signal which is embedded in the center view must have an ability to protect the respective three views. In this study, we solve this problem by exploiting the horizontal noise mean shifting (HNMS) technique. We exploit the fact that the objects in the view are shifted only to horizontal way when the center view renders to the left and right views. Using this fact, the proposed stereoscopic watermarking scheme moves the mean of horizontal noise histogram which is invariant to

horizontal shifting, and we achieve good performance as shown in the experimental results.

8303-06, Session 2

Reversible q-ry watermarking with controllable prediction error and location map-free capability

T. Efimushkina, K. O. Egiazarian, Tampere Univ. of Technology (Finland)

In this paper a new high capacity and high quality reversible image watermarking method is presented. It is based on prediction error expansion with a controllable prediction error and a local activity indicator used to manage the embedding capacity. This method enables hiding the secret data from Galois field, which allows to embed more than one bit per pixel in a single run of the algorithm. Special conditions for location map free data embedding are derived in order to decrease the auxiliary data. However, the algorithm allows to leave out the computation of these thresholds substituting them by user-specified thresholds. In that case, only one iteration of the encoding process is performed with the utilization of losslessly compressed mask, that indicates non-embeddable pixel positions.

Algorithm has a special control mechanism that allows the data embedding process only if the prediction error is bounded. Moreover, over/underflow conditions are to be sustained.

The proposed method is compared with the state-of-the-art reversible data hiding methods and demonstrates the higher performance in terms of the image quality and capacity. The capacity can be further increased by applying multiple iterations of the proposed q-ry scheme.

8303-07, Session 3

Optimizing pixel predictors for steganalysis

V. Holub, J. Fridrich, Binghamton Univ. (United States)

A standard way to design steganalysis features for digital images is to choose a pixel predictor, use it to compute a noise residual, and then form joint statistics of neighboring residual samples (co-occurrence matrices). This paper proposes a general data-driven approach to optimizing predictors for steganalysis. First, a local pixel predictor is parametrized and then its parameters are determined by solving an optimization problem for a given sample of cover and stego images and a given cover source. Our research shows that predictors optimized to detect a specific case of steganography may be vastly different than predictors optimized for the cover source only. The results indicate that optimized predictors may improve steganalysis by a rather non-negligible margin. Furthermore, we construct the predictors sequentially - having optimized k predictors, design the k+1st one with respect to the combined feature set built from all k predictors. In other words, given a feature space (image model) extend (diversify) the model in a selected direction (functional form of the predictor) in a way that maximally boosts detection accuracy.

8303-08, Session 3

Steganalysis of JPEG images using rich models

J. Kodovsky, J. Fridrich, Binghamton Univ. (United States)

In this paper, we propose a rich model of DCT coefficients in a JPEG file for the purpose of detecting steganographic embedding changes. The model is built systematically as a union of smaller submodels constructed as joint distributions of DCT coefficients from their frequency and spatial neighborhoods covering a wide range of statistical dependencies. Due to its high dimensionality, we combine the rich model with ensemble classifiers and construct detectors for six modern JPEG domain steganographic schemes: nsF5, model-based steganography, YASS, and schemes that use side information

at the embedder in the form of the uncompressed image: MMEx, BCH, and BCHopt. The resulting performance is contrasted with previously proposed feature sets, including both low- and high-dimensional sets. We also investigate the performance of individual submodels when grouped by their type as well as the effect of cartesian calibration. The proposed rich model delivers superior performance across all tested algorithms and payloads.

8303-09, Session 3

Co-occurrence steganalysis in high dimension

T. Pevny, Czech Technical Univ. in Prague (Czech Republic)

The state of the art steganalytic features for spatial domain, and to some extent for transfer domains (DCT), are based on histogram of co-occurrences of neighboring elements. The rationale behind is that neighboring pixels in digital images are correlated, which is caused by the smoothness of our world and by the image processing. The limitation of histogram-based features is that they do not scale well with a dimension of the histogram, since the number of histogram bins (hence number of features) grows exponentially with the number of modeled neighboring elements. Despite the use of ensemble classifiers proposed to deal more efficiently with large number of features, we still believe that the small number of good features is preferable over the large number of weaker ones.

The major goal of this work is to introduce the problem of reduction of histogram bins without sacrificing the detection quality. For the problem, which was not yet tackled in the prior art, a heuristic solution is proposed. The solution, albeit not being optimal, already outperforms the ad-hoc solution used in the prior art.

The methods reducing the number of histogram bins are very important, as the results of BOSS competition showed that the features extracted from histograms of higher order and needed to detect state of the art steganographic algorithms.

8303-18, Session 4

Privacy and the social network

J. Staddon, Google Inc. (United States)

The tension between online social services and privacy is readily apparent. Online social networks are invaluable for making connections between people and facilitating sharing, but can also include the unintentional over-sharing of content and incidents of unwanted online attention from other users. I'll talk about gauging privacy concerns, expectations and feature utility as integral parts of the process of building an engaging and privacy-aware social network. Drawing examples from our experience with Google+, I'll discuss how each is key to our iterative privacy design process, which includes threat assessment, data analysis and ample user studies.

8303-11, Session 5

Source camcorder identification with cropped and scaled videos

D. Hyun, S. Ryu, M. Lee, J. Lee, H. Lee, H. Lee, KAIST (Korea, Republic of)

In this paper, we propose a new camcorder identification method based on photo-response non-uniformity (PRNU) to perform well with videos under investigation which are simultaneously cropped and scaled. One of disadvantages of PRNU-based algorithms is that it is very sensitive to de-synchronization. If an input video is slightly cropped or scaled, the the identification process without synchronization will fail. The proposed method solves this kind of out-of-sync problem, by achieving downscale-tolerance using minimum average correlation energy Mellin radial harmonic (MACE-MRH) filter. The experimental results show that the proposed method identifies source device faster and more accurately than the existing method.

Conference 8303

8303-12, Session 5

Digital image forensics for photographic copying

Y. Fang, J. Yin, Sun Yat-Sen Univ. (China)

Image display technology has greatly developed over the past few decades, which make it possible to recapture high-quality images from the display medium, such as a liquid crystal display(LCD) screen or a printed paper. The recaptured images are not regarded as a separate image class in the current research of digital image forensics, while the content of the recaptured images may have been tempered. In order to study this problem, we firstly obtained finely recaptured images from LCD screen and printed papers, then two sets of features based on the noise and the traces of double JPEG compression are proposed to identify these recaptured images. Experimental results showed that our proposed features perform well for detecting photographic copying.

8303-13, Session 5

Forensic audio watermark detection

M. Steinebach, S. Zmudzinski, Fraunhofer-Institut für Sichere Informations-Technologie (Germany)

Digital audio watermarking detection is often computational complex and requires at least as much audio information as required to embed a complete watermark. In some applications, especially real-time monitoring, this is an important drawback. The reason for this is the usage of sync sequences at the beginning of the watermark, allowing a decision about the presence only if at least the sync has been found and retrieved. We propose an alternative method for detecting the presence of a watermark. Based on the knowledge of the secret key used for embedding, we create a mark for all potential marking stages and then use a sliding window to test a given audio file on the presence of statistical characteristics caused by embedding. In this way we can detect a watermark in less than 1 second of audio.

8303-14, Session 5

Sensor-fingerprint based identification of images corrected for lens distortion

M. Goljan, J. Fridrich, Binghamton Univ. (United States)

Computational photography is quickly making its way from professional devices to consumer cameras. Recently, camera manufacturers started using in-camera lens-distortion correction of the captured image to give users more powerful range of zoom in compact and affordable cameras. Since the distortion correction (barrel/pincushion) depends on the zoom, it desynchronizes the pixel-to-pixel correspondence between images taken at two different focal lengths. This poses a serious problem for digital forensic techniques that utilize the concept of sensor fingerprint (photo-response non-uniformity), such as "image ballistic" techniques that can match an image to a specific camera in a similar manner as a bullet can be matched to a gun barrel. Such identification techniques may in fact completely fail. This paper presents an extension of sensor-based camera identification that would work for images whose content was corrected for lens distortion. To reestablish synchronization between an image and the fingerprint, we adopt a barrel distortion model and search for its parameters to maximize the detection statistic, which is the PCE (peak to correlation energy ratio). The proposed method is tested on hundreds of images from three compact cameras to prove the viability of the concept and demonstrate its efficiency.

8303-15, Session 6

Digital audio authentication by robust feature embedding

S. Zmudzinski, B. Munir, M. Steinebach, Fraunhofer-Institut für Sichere Informations-Technologie (Germany)

In many scenarios, digital audio data contains important or sensitive information. Examples are given by recordings of telephone calls, police interrogations, air traffic communication or content preserving the cultural heritage. As the audio data can easily be semantically manipulated with modern computer hardware and software, mechanisms for verifying the integrity of the audio content and the authenticity of its origin are of special interest. We introduce and evaluate an approach for verifying the integrity of digital audio recording by means of content-based integrity watermarking. Here an audio fingerprint is extracted in the Fourier domain and embedded as a digital watermark. The design of the feature extraction allows a fine temporal resolution of the verification of the integrity.

8303-16, Session 6

High-resolution printed amino acid traces: a first-feature extraction approach for fingerprint forgery detection

M. Hildebrandt, S. Kiltz, J. Dittmann, Otto-von-Guericke-Univ. Magdeburg (Germany)

Fingerprints are used for the identification of individuals for over a century in crime scene forensics. Here, often physical or chemical preprocessing techniques are used to render a latent fingerprint visible. For quality assurance purposes of those development techniques, Schwarz (2009) introduces a technique for the reproducible generation of latent fingerprints using ink-jet printers and artificial amino acid sweat. However, this technique allows for printing latent fingerprints at crime scenes to leave wrong traces, too. Hence, Kiltz et al. (2011) introduce a first framework for the detection of printed fingerprints. However, the utilized printers have a maximum resolution of 2400x1200 dpi. In this paper, we use a printer with a much better resolution of 9600x2400 dpi, which does not produce visible dot patterns as reported in Kiltz et al. (2011) We show, that an acquisition with a resolution of 12700 to 25400 ppi is necessary to extract microstuctures which perspectively allows for an automated detection of printed fingerprint traces fabricated with high-resolution printers. Here, using our first test set with 15 printed and 5 real, natural fingerprint patterns from the human the evaluation results indicate a very positive tendency towards the detectability of such traces.

8303-17, Session 6

Image forgery detection by means of no-reference quality metrics

F. Battisti, M. Carli, A. Neri, Univ. degli Studi di Roma Tre (Italy)

In this paper a methodology for digital image forgery detection by means of an unconventional use of image quality assessment metrics is addressed. In particular, the analysis of the variation of degradations inside an image is adopted to reveal the presence of patches that have been pasted from a different source. The ratio behind this work is in the hypothesis that any image may be affected by artifacts, visible or not, caused by the processing steps: acquisition (i.e., lens distortion, imperfections in the acquisition sensors, analog to digital conversion, single sensor to color pattern interpolation), processing (i.e., quantization, storing, JPEG compression, sharpening, deblurring, enhancement), and rendering (i.e., image decoding, color/size adjustment). These impairments are usually localized and their strength is strictly dependent on the content. These characteristics can be considered as a fingerprint of each digital image. The proposed approach relies on a combination of image quality assessment systems. The adopted no reference metric do not require any information about the original image, thus allowing an efficient and stand alone blind system for image forgery detection. The experimental results show the effectiveness of the proposed scheme.

8303-10, Session 7

Advancing technology: bane and boon for banknotes

S. E. Church, Board of Governors of the Federal Reserve System (United States)

Over the last two decades, rapid advances in technology have led to significant threats to banknote security. It may seem obvious that advanced digital imaging and printing technologies provide ready accessibility to rapid and faithful image capture, easy image manipulation and customization, and high-resolution, digital desktop printing, pose threats to the security of banknotes. What may be less obvious is that the negative consequences of these advances in imaging and printing have led to positive consequences in the form of innovation and exploration of advanced technologies for use in banknote production and security.

In this way, the advances in technology have both created problems for the security of banknotes and provided potential solutions to address the problems. As the printed images of banknotes have become more vulnerable, the banknote community has turned to materials engineering, optical techniques, and other advanced approaches to provide security. As a result, modern security features may add additional dimensions to the traditional two-dimensional printed image of the notes and may require manipulation or special techniques to activate or access. Such new features are not readily captured with digital techniques alone, forcing the counterfeiter to use additional materials and skills to achieve satisfactory results. In this respect, advancing technology offers increasing opportunities to discourage counterfeiting and lessen the success of the outcome if attempted. As a result of both the threats and opportunities offered by advancing technology, modern security features and banknotes are highly engineered products based on the innovative marriage of venerably old but modernized production techniques and advanced, precisely engineered materials.

The presentation will explore some of the ways that technology has served both to threaten and to enhance banknote security.

8303-19, Session 8

Going from small to large data in steganalysis

I. Lubenko, A. D. Ker, Univ. of Oxford (United Kingdom)

No abstract available

8303-20, Session 8

Identifying a steganographer in realistic and heterogeneous data sets

A. D. Ker, Univ. of Oxford (United Kingdom); T. Pevny, Czech Technical Univ. in Prague (Czech Republic)

The traditional steganalytic scenario usually considers the case, when the steganographer tries to detect, if a single object carries payload or not. Such scenario is common in the most literature, even though it is highly theoretical. In practice, one can expect to consider rather different scenario, where the steganographer has to consider multiple users each transmitting multiple objects. This problem, known as a pooled steganalysis, has been theoretically introduced in 2006, but no practical solution has been proposed until 2011.

This work extends the experiments presented in the original publication, to a new, more realistic, domain, where we utilizes 800 000 images from 4000 users of a social networking service.

8303-21, Session 9

Asymmetric robust quantum image hashing

M. Steinebach, H. Liu, Fraunhofer-Institut für Sichere Informations-Technologie (Germany)

Forensic analysis of image sets today is most often done with the help of cryptographic hashes due to their efficiency, their integration in forensic tools and their excellent reliability in the domain of false detection alarms. A drawback of these hash methods is their fragility to any image processing operation. Even a simple re-compression with JPEG results in an image not detectable. A different approach is to apply image identification methods, allowing identifying illegal images by e.g. semantic models or facing detection algorithms. Their common drawback is a high computational complexity and significant false alarm rates. Robust hashing is a well-known approach sharing characteristics of both cryptographic hashes and image identification methods. It is fast, robust to common image processing and features low false alarm rates. To verify its usability in forensic evaluation, in this work we discuss and evaluate the behavior of an optimized block-based hash.

8303-22, Session 9

Fast detection of Tardos codes with Boneh-Shaw types

M. Desoubeaux, G. Le Guelvouit, France Telecom R&D (France); W. Puech, Lab. d'Informatique de Robotique et de Microelectronique de Montpellier (France)

Traitor tracing aims at preventing unauthorized redistribution of multimedia content by embedding individual sequences of bits within each authorized copy. These sequences have to be robust against collusion attacks. Collusion is the process used by dishonest users to attempt to forge an untraceable content with their copies. Current applications of fingerprinting deal with a large number of users and distributors need to have a fast detection process.

In this paper we present a two level approach to improve the detection of pirates with the well known Tardos fingerprinting scheme. A Boneh-Shaw random code is concatenated with the Tardos code thanks to a q-ary alphabet. A improvement at the decoding side is presented by organizing the detection of users over different subgroups of users which are selected thanks to the Boneh-Shaw random code. The efficiency of the detection is strongly dependent on the Tardos code length but experimentations confirm nevertheless that we reduce the decoding complexity.

8303-23, Session 9

Locatability of modified pixels in steganographic images

T. Quach, Sandia National Labs. (United States)

Payload location using residuals is a successful approach to identify load-carrying pixels provided a large number of stego images are available. Furthermore, each image must have the payload embedded at the same locations. The success of payload location is therefore limited if different keys are used or an adaptive embedding algorithm is used. Given these limitations, the focus of this paper is to locate modified pixels in a single stego image. Given a sufficiently large set of independent binary decision functions, each determines whether a pixel has been modified better than guessing, we show that it is possible to locate modified pixels in a single stego image with low error rate. We construct these functions using existing cover estimators and provide experimental results to support our analysis.

Conference 8303

8303-24, Session 9

Forensic characterization of camcorded movies: digital cinema versus celluloid film prints

X. Rolland-Neviere, B. Chupeau, G. Doërr, L. Blondé, Technicolor S.A. (France)

Most newly released movie piracy can be traced to individuals who used a recording device in a movie theater. Since the underlying tracing mechanism differs, the forensic analysis of a pirate copy will depend on the piracy process e.g. camcorder capture in a digital cinema vs. camcorder capture in a legacy cinema. Rapid classification of incoming pirate samples according to their origin, prior to any further in-depth analysis, is therefore desired in order to speed up the whole forensic process.

The objective of this paper is to devise an automatic oracle that discriminates in a robust manner camcorded copies originating from digital cinemas and those from legacy ones. To do so, we carefully analyse the projection process in order to pinpoint a few tell-tale visual artifacts, namely global illumination uniformity, on-screen vertical stability and temporal illumination pulse. We extract features associated to these artifacts and subsequently feed them into a state-of-the-art classifier to train it to separate the two classes of camcorded copies. Experimental results are then reported to demonstrate the discriminative power of these features.

8303-25, Session 10

Extending a context model for microphone forensics

C. Krätzer, K. Qian, J. Dittmann, Otto-von-Guericke-Univ. Magdeburg (Germany)

In the paper, we extend an existing statistical pattern recognition based microphone forensic approach by considerations on: recordings of replays, the influence of three different classes of sound sources (loudspeakers as well as harmonic and non-harmonic natural sources) as well as additional influences in sound transmission and recording (here power decay, echoes, reverberation and different pre-amplifiers). The extension of the context model is accompanied by empirical investigations on the authentication performance of our microphone forensic framework, aiming at: a) answering the question how good a replay can be differentiated from a natural, non-playback signal, and b) at the estimation of the strength of the impact of the aforementioned variations (e.g. of the source type) in the recording pipeline. The domain knowledge gained from all evaluations is used to increase the precision of our context model for describing typical microphone recording processes. Such an empirical investigation on the influences involved in the recording process allows for a more reliable estimation of the plausibility of our statistical pattern recognition based source forensic approach.

8303-26, Session 10

Simulating large-scale acoustic path benchmarking

M. Arnold, P. Baum, M. Alonso, U. Gries, Deutsche Thomson oHG (Germany); G. Doërr, Technicolor S.A. (France)

Robustness to D/A-A/D conversion is one of the key features of watermarking systems. In audio, it naturally translates to robustness against acoustic path transmission i.e. the ability to detect a watermark once the content has been played with some loudspeakers and picked up with a microphone.

Many applications inherently require robustness against acoustic path transmission. However, this problem is relatively understudied in the literature. The key issue is that in-depth evaluation of the acoustic path is time consuming, cumbersome, and error-prone. The objective of this study is to devise a methodology that would permit simulating such

large scale acoustic path benchmarking campaign without the logistic burden currently associated to it.

Relying on some simplifying assumption, we create a validated quickly adjustable automated tool to replicate the impact of acoustic path transmission in a given room, with given loudspeakers at a given location and given microphones at a given position. Experimental results demonstrate the simulations closely match results which would be achieved with real measurements.

8303-27, Session 10

Noise removing in encrypted color image by statistical analysis

N. Islam, W. Puech, Lab. d'Informatique de Robotique et de Microelectronique de Montpellier (France)

Cryptographic techniques are used to secure confidential data from unauthorized access but these techniques are very sensitive to noise. A single bit change in encrypted data can have catastrophic impact over the decrypted data. This paper addresses the problem of removing bit error in visual data which are encrypted using AES algorithm in the CBC mode. In order to remove the noise, a method is proposed which is based on the statistical analysis of each block during the decryption. The proposed method exploits local statistics of the visual data and confusion/diffusion properties of the encryption algorithm to remove the errors. Experimental results show that the proposed method can be used at the receiving end for the possible solution for noise removing in visual data in encrypted domain.

Conference 8304A: Multimedia on Mobile Devices 2012

Wednesday 25 January 2012

Part of Proceedings of SPIE Vol. 8304A Multimedia on Mobile Devices 2012

8304A-16, Poster Session

Low-complexity bit-plane entropy coding for 3D DWT-based video compression

E. A. Belyaev, K. O. Egiazarian, M. Gabbouj, Tampere Univ. of Technology (Finland)

An extension of the H.264/SVC standard is currently the most popular scalable video coding approach. But, because of high computational complexity of motion estimation and inter-layer prediction at the encoder side, software and hardware implementations of H.264/SVC encoder in a mobile device is a difficult task. As an alternative video encoders based on three-dimensional discrete wavelet transform (3-D DWT) can be used. These approaches do not use motion estimation for exploiting a temporal redundancy of the video source, and thus, have lower computational complexity.

For further decreasing of the 3-D DWT based video compression complexity we propose new simple bit-plane entropy coding of wavelet matrixes. All bits in bit-planes of each wavelet matrix are separated into two binary sequences. First sequence has very low probability of ones, therefore it is divided by series of zeros which are compressed by Levenshtein codes. Probability of ones in second sequence is close to 0.5, therefore bits of this sequence are placed into bit stream directly.

Proposed approach allows implementing entropy coding without using of look-up tables. We show that computation complexity of proposed entropy coder is less than bit-plane arithmetic coder in JPEG2000 and entropy encoder in 3-D SPIHT. At the same time it provides comparable compression efficiency.

8304A-18, Poster Session

Bidirectional probabilistic hyper-graph matching method using Bayes theorem

W. Cho, S. Kim, S. Park, Chonnam National Univ. (Korea, Republic of)

Establishing correspondences between two hyper-graphs is a fundamental issue in computer vision, pattern recognition, and machine learning. A hyper-graph is modeled by feature set where the complex relations are represented by hyper-edges. Hence, a match between two feature sets determines a hyper-graph matching problem. We propose a new bidirectional probabilistic hyper-graph matching method using Bayesian inference principle. First, we formulate the corresponding hyper-graph matching problem as the maximization of a matching score function over all permutations of the features. Second, we induce an algebraic relation between the hyper-edge weight matrixes and derive the desired vertex to vertex probabilistic matching algorithm using Bayes theorem. Third, we apply the well known convex relaxation procedure with probabilistic soft matching matrix to get a complete hard matching result. Finally, we have conducted the comparative experiments on synthetic data and real images. Experimental results show that the proposed method clearly outperforms existing algorithms especially in the presence of noise and outliers.

8304A-19, Poster Session

SeamCrop for image retargeting

J. Kiess, B. Guthier, S. Kopf, W. Effelsberg, Univ. Mannheim (Germany)

In this paper, we present a novel approach for the adaptation of large images to small display sizes. As a recent study suggests, most viewers prefer the loss of content over the insertion of deformations

in the retargeting process. Therefore, we combine the two image retargeting operators seam carving and cropping in order to resize an image without manipulating the important objects in an image at all. First, seams are removed carefully until a dynamic energy threshold is reached to prevent the creation of visible artifacts. Then, a cropping window is selected in the image that has the smallest possible window size without having the removed energy rise above a second dynamic threshold. As the number of removed seams and the size of the cropping window are not fix, the process is repeated iteratively until the target size is reached. Our results show that by using this method, more important content of an image can be included in the cropping window than in normal cropping. The "squeezing" of objects which might occur in approaches based on warping or scaling is also prevented.

8304A-20, Poster Session

Collecting fingerprints for recognition using mobile phone cameras

B. Yang, X. Li, C. Busch, Gjøvik Univ. College (Norway)

General-purpose cameras such as the camera embedded in a mobile phone can be expected to complement the professional fingerprint and face scanners as alternative sensors suiting some consumer biometric applications such as mobile e-payment and ticketing, for both on-line and on-site transaction, which can be used to increase the privacy and convenience on the personal side and also reduce the biometric system deployment cost on the business side. However, a big challenge of employing such consumer electronics as biometric sensor is the sample quality control which cannot be directly addressed by existing fingerprint quality estimation mechanisms due to the richer background noises, worse lightning conditions, and camera focusing inaccuracy. We propose an algorithm to check the sample quality of the fingerprint images captured by mobile phone camera in an accurate and efficient way. Ridge frequency and orientation features are fully exploited in the Discrete Fourier Transformation domain of the image blocks to give indication if the studied block is high-quality fingerprint area or not. Diversified real data collection scenarios are used to test the proposed algorithm's performance in quality estimation.

8304A-21, Poster Session

Overview of potential forensic analysis of an Android smartphone

R. Creutzburg, K. Kröger, S. Sack, Fachhochschule Brandenburg (Germany)

This paper gives an overview of potential forensic analyses of an Android Smartphone. The security concept and the underlying architecture of the operating system are introduced and forensically specific features of importance are presented. With the help of these preconsiderations and taking the example of a HTC Desire Smartphone, a forensic analysis will be carried out. The analysis explains certain features of the Android operating system such as the functionality of the lock screen, the difference between root and unroot and the use of the debug mode. The practical study shows some of the methods of how to extract information from the Smartphone. This will include the forensic analysis of the SD Card and the logical analysis of Smartphones using professional forensic software such as Oxygen Forensic Suite 2011. The creation of a bit by bit image using the Android SDK and the UNIX dd command is part of the analysis as well.

8304A-22, Poster Session

Forensics of geodata collected by Apple iOS and Google Android in mobile devices

K. Kröger, R. Creutzburg, Fachhochschule Brandenburg (Germany)

This paper gives an overview of the used technologies, the extraction and the analysis of stored geodata collected by iOS and Android in mobile devices.

The study shows how stored data can be extracted and examined.

The main aspect of the study is the analysis of the differences between iOS and Android and possibilities for the aditional use of the extracted data.

8304A-23, Poster Session

Template-based mobile platform image processing training

D. Akopian, S. C. Golagani, M. Esfahanian, The Univ. of Texas at San Antonio (United States)

No abstract available

8304A-24, Poster Session

Combining associative computing and distributed arithmetic methods for efficient implementation of multiple inner products

D. Guevorkian, Tampere Univ. of Technology (Finland); P. Liuha, T. Yli-Pietilä, Nokia Research Ctr. (Finland); K. O. Egiazarian, Tampere Univ. of Technology (Finland)

Many multimedia processing algorithms as well as communication algorithms implemented in mobile devices are based on intensive implementation of linear algebra methods, in particular, implying implementation of a large number of inner products in real time. Among most efficient approaches to perform inner products are the Associative Computing (ASC) approach and Distributed Arithmetic (DA) approach. In this work, we propose a combination of these two powerful methods that further improves the efficiency of multiple inner product computation.

8304A-25, Poster Session

Presentation of forensically interesting Microsoft XBox 360 console features

S. Luttenberger, K. Kröger, R. Creutzburg, Fachhochschule Brandenburg (Germany)

This paper deals with forensically interesting features of the Microsoft XBOX 360 game console.

The construction and the internal structure are analysed more precisely. One of the main aspects of the study is to analyse the used file system which was examined for forensic features.

Possible difficulties that might be of importance to the forensic investigator are discussed.

8304A-26, Poster Session

Presentation of forensically interesting Sony Playstation 3 console features

K. Kröger, G. Daugs, R. Creutzburg, Fachhochschule Brandenburg (Germany)

This paper deals with forensically interesting features of the Sony Playstation 3 game console.

The construction and the internal structure are analysed more precisely. Interesting forensic features of the operating system and the file system are presented.

Differences between a PS3 with and without jailbreak are introduced and possible forensic attempts when using an installed Linux are discussed.

8304A-28, Poster Session

A neural network-based approach for recognition of engraved and embossed labels on metallic parts

A. Shirkhodaie, V. K. Bandaru, Tennessee State Univ. (United States)

Conventional Optical Character Recognition (OCR) approaches have proven to be less effective for recognition of engraved and embossed (E2) characters. Their weaknesses are, particularly, paramount in inspection of engraved and embossed characters introduced on metal cast parts. Due to poor illumination, low signal-to-noise, characters font, size, spacing, color, and impression variations, and other inevitable manufacturing irregularities (e.g., poor stamping, mis-stamping, or partial stamping) in introducing these characters, most OCR are impractical to detect and recognition manufactured parts labels correctly. In this paper, we initially address the issues of characters segmentation for their noisy background and describe strength and weakness of conventional image thresholding techniques for proper segmentation of foreground and background images. Next, we discuss three competing approaches for extraction of E2 characters with low contrast relative to the background. For recognition of extracted characters, we trained a Hamming Neural Network (HNN). We demonstrate that the proposed technique neural network-based approach can effectively and efficiently recognize E2 characters with high confidence and is fault tolerant with respect to characters' font, size, color, and spacing as well as invariant of nominal manufacturing position and orientation inconsistency of E2 characters. Results from several experiments are presented. A performance comparison of proposed approach versus a conventional OCR technique is presented for demonstration of efficiency and effectiveness of the proposed approach.

8304A-30, Poster Session

A fuzzy-logic approach for metallic parts surface defects characterization and shape classification

A. Shirkhodaie, F. Vaziribozorg, Tennessee State Univ. (United States)

Recent advances in Automatic Surface Inspection System (ASIS) enable automatic and systematic inspection of quality manufactured products. Automatic visual inspection systems rely on accuracy of visual inspection algorithms and techniques to reliably detect and characterize the surface imperfections and defects. In this paper, a new approach for defect detection and shape characterization of metallic parts is proposed. The new technique is trainable and based on a fuzzy logic model which performs shape classification of defective regions. Primarily, the algorithm detects defective regions of each part based on a new thresholding technique. Image statistical properties are further used to discriminate between normal and defective regions. Once a defective region is detected, them a multi-level thresholding is applied to acquire the shape representation of the defective area. Upon the shape of defective area is determined, then, shape spatial features are computed and used as input to the fuzzy logic shape classifier. By combining qualitative and quantitative measures, then, extend of shape variation normality or defectiveness is determined. The proposed technique is tested for defect detection and shape characterization of different types of metallic parts. In comparison with other known techniques, this new approach has improved efficiency and effectiveness in terms of detecting and

Conference 8304A

IS&T / SPIE
Electronic Imaging
SCIENCE AND TECHNOLOGY

characterization variety of defective surfaces properly with high degree of confidence.

8304A-31, Poster Session

Sudoku substitution-permutation image cipher

Y. Wu, Tufts Univ. (United States); S. S. Agaian, The Univ. of Texas at San Antonio (United States); J. P. Noonan, Tufts Univ. (United States)

With the development of digital technologies, the digital image becomes a common data type in modern life. Compare to well-developed ciphers, like DES, IDEA, AES, the image ciphers are still immature in many aspects: 1) cipher security in lack of theoretical supports; 2) cipher performance is limited to several ciphertext-only measurements; and 3) cipher is not defined on the finite precision system. In this paper, we introduce a new image cipher using the Sudoku Substitution-Permutation (SSP). The cipher efficiency is guaranteed by the parameterized fast Sudoku transform, which expands a key dependent sequence of length to a square Sudoku matrix of size and then provides a reference to future substitution and permutation process. The cipher security is endorsed by the Substitution-Permutation Network (SPN) which is a well-known structure in cryptography with good confusion and diffusion properties. Moreover, the encryption quality is also enhanced by using Sudoku matrix, which is type of puzzle matrix with no repeated digits in any row, column or block. Simulation results show that the proposed cipher outperforms many existing image ciphers/encryption algorithms in both quality and efficiency.

8304A-01, Session 1

Location-aware gang graffiti acquisition and browsing on a mobile device

A. Parra Pozo, M. Boutin, E. J. Delp III, Purdue Univ. (United States)

In this paper, we propose a mobile-based system that allows first responders to identify and track gang activity, combining the use of image analysis and location-based-services. The image analysis includes obtaining the metadata (geoposition, date and time) and extracting relevant features (e.g., color, shape) from the gang graffiti image. The data obtained automatically on the device is transferred to a server and compared against a database of graffiti images. The matched results are sent back to the device where the user can then review the results and provide extra inputs to refine information.

The first responder can take advantage of the location-based-services that the mobile device provides. Along with the use of map projections or augmented reality techniques, all the graffiti data in a certain radius from the current location can be pulled from the server. The data includes not only the images, but all the information related to it, such as date and time, geoposition, gang, gang member, colors, or symbols. The user can then visuallize the data in the area in a useful manner, and take the necessary actions. Therefore, the process of identifying and tracking gang activity is made more efficient, leading to a faster intervention by the law enforcement officers.

8304A-02, Session 1

Dietary intake assessment using integrated sensors and software

J. Shang, E. Johnson, K. Sundara-Rajan, A. Teredesai, Univ. of Washington (United States); A. Kristal, Fred Hutchinson Cancer Research Ctr. (United States); A. V. Mamishev, Univ. of Washington (United States)

The area of dietary assessment is becoming increasingly important as obesity rates soar, but valid measurement of the food intakes in free-living persons is extraordinarily challenging. Traditional paper-based

dietary assessment methods have limitations due to bias, user burden and cost, so improved methods are needed to address important hypotheses related to diet and health. In this paper, we will describe the progress of our mobile Diet Data Recorder System (DDRS), where an electronic device is used for objective measurement on dietary intake in real time and at moderate cost. The DDRS consists of (1) a mobile device that integrates a smartphone and an integrated laser package, (2) software on the smartphone for data collection and laser control, (3) an algorithm to process acquired data for food volume estimation, which is the largest source of error in calculating dietary intake, and (4) database and interface for data storage and management. The estimated food volume, together with direct entries of food questionnaires and voice recordings, could provide dietitians with more complete food description and more accurate food portion sizes. In this paper, we will describe the system design of DDRS and initial results of dietary assessment.

8304A-03, Session 1

FCam for multiple cameras

A. Troccoli, NVIDIA Corp. (United States); C. Zhou, Columbia Univ. (United States); K. A. Pulli, NVIDIA Corp. (United States)

Photography was never just about recording an objective reality, but a creative task. As cameras have moved away from paper and chemicals to digital imaging, the processing became easier to control and faster. The latest trend is to move all that computation into cameras that capture, merge, and process images digitally, yielding the final image right away in the same device. This is facilitated by flexible camera control APIs such as the FCam API, which implements the FrankenCamera imaging architecture.

The first version of FCam supported only a single camera. However, many modern mobile devices have at least two cameras, one next to the display facing the user, and the other pointing away from the user. Some devices, such as the NVIDIA's Tegra 3 development board, have even a stereo camera pair. We have extended FCam API to explicitly support several cameras. The facilities include querying the number and properties of the cameras. The properties include some obvious ones, such as camera resolutions, but also indication of the geometric configuration of the cameras, from which the program can deduce the camera orientation, stereo baseline, etc. We have also created a convenience class that allows treating a stereo camera pair logically as a single sensor.

8304A-04, Session 2

Biosensing mobile display principle for healthcare

W. Mphepö, iVorex AB, Borlänge (Sweden) and Beijing Normal Univ. (China)

Convergence and the ability to harness intrinsic properties of some display technologies can open up new dimensions to the ubiquitous mobile devices. Biosensing mobile displays being one such dimension we hereby present based on our Digital Micro Hinge [DMH] Display technology research. Interest in such biosensing devices is obvious from the recently announced $10 million Qualcomm Tricorder X Prize competition.

8304A-05, Session 2

Continuously adjustable Pulfrich spectacles for mobile devices

K. M. Jacobs, Binghamton Univ. (United States); R. S. Karpf, Consultant (United States)

Mobile devices present a challenging platform for 3-D video because of inherent device limitations. Continuously Adjustable Pulfrich Spectacles (CAPS) is a new implementation of the Pulfrich 3-D stereoscopic effect. For every scene that contains lateral motion in a 2-D movie, CAPS

144   electronicimaging.org   •   TEL: +1 703 642 9090   •   ei@imaging.org

Conference 8304A

provides realistic 3-D. Since it requires minimal additional processing, it is appropriate for mobile devices.

3-D movies utilizing the Pulfrich stereoscopic effect have been made for 80 years using passive viewing spectacles. CAPS use active viewing spectacles to overcome passive glasses' limitations. 3-D movies normally employ the asymmetry of dual images to produce stereopsis. CAPS works on the principle of illumination asymmetry, and only needs to control the differential lens optical densities.

CAPS are fabricated from optoelectronic materials that electronically control the lens optical densities. The eye's retinal triggering is used by CAPS to determine the differential lens optical densities. Motion estimation calculations from the digital image processing used to display 2-D video on mobile devices are reused to calculate real-time lens adjustments so CAPS always conform to the optical density that optimizes the Pulfrich stereoscopic effect.

Only negligible additional processing is necessary for CAPS to show 3-D for every scene that contains lateral motion in any 2-D movie.

8304A-06, Session 2

Parameters of the human 3D gaze while observing portable autostereoscopic display: a model and measurement results

A. R. Boev, M. Hanhela, A. P. Gotchev, T. Utirainen, S. Jumisko-Pyykkö, Tampere Univ. of Technology (Finland); M. Hannuksela, Nokia Research Ctr. (Finland)

We present an approach to measure and model the parameters of human point-of-gaze (PoG) in 3D space. We have developed a methodology which allows us to suppress most of the measurement noise. We identify three temporal properties of the binocular PoG. The first is reaction time, which is the minimum time that the vision reacts to a stimulus position change, and is measured as the time between the event, and the time the PoG leaves the proximity of the old stimulus position. The second is travel time of the PoG between the old and new stimulus position. The third is the time-to-arrive, which is the time combining the reaction time, travel time, and the time required for the PoG to settle in the new position.

We present the method for filtering the PoG outliers, for deriving the PoG center from binocular eye-tracking data and for calculating the gaze volume as a function of the distance between PoG and the observer. We show the mean values for all temporal properties separately for x, y and z direction averaged over all observers.

8304A-07, Session 2

Deblocking of mobile stereo video

A. P. Gotchev, L. Azzari, K. O. Egiazarian, Tampere Univ. of Technology (Finland)

Standard block transform-based compression methods often cause blocking artefacts, which have been found particularly annoying and also degrading the overall quality and the perception of depth. Suitable deblocking is required at the receiver side to tackle such artefacts. Current trend in restoration algorithms suggests applying non-local collaborative filtering methods. In this contribution, such a technique is proposed for deblocking of stereo video compressed by Simulcast or multi-view coding (MVC). The algorithm searches for similar patches exhibiting high spatial correlation along temporal dimension and between the two views. The selected patches are grouped into comprehensive 4D structure, which is decorrelated by efficient transforms, namely DCT along spatio-temporal directions and Haar wavelet transform between stereo frames. Artefact suppression is performed through transform-domain thresholding to result in first empirical estimate of the deblocked signal which is then used for a second-stage transform-domain Wiener filtering. Furthermore, an elegant stereo sharpening can be accomplished in the transform domain by alpha-rooting. We demonstrate that the procedure leads to visually pleasant results and is compares favourably against its simplified versions, i.e. sliding DCT filtering and bilateral filtering.

8304A-08, Session 3

SUPL support for mobile devices

J. M. Narisetty, A. Soghoyan, M. C. Sundaramurthy, D. Akopian, The Univ. of Texas at San Antonio (United States)

The conventional Global Positioning System (GPS) receivers are well known positioning tools in open-sky environments. But their performance degrades in urban canyons, indoors and underground due to multipath, foliage, dissipation, etc. To overcome such situations, extensive research has been done by engineers in recent years resulting in an approach known as Assisted GPS (A-GPS). Using this approach, orbital parameters such as ephemeris, almanac, reference time and coarse location information is provided to the GPS receivers which significantly enhances the acquisition of weak signals. To test A-GPS enabled receivers high-end simulators are used, as they involve additional assistance data generation, GPS signal, positioning measurements, acquisition of coarse location and delivery to target devices. Typically such simulators are not affordable by many academic institutions. National Instruments (NI) GPS Simulation Toolkit with NI RF hardware support provides an efficient testing environment for A-GPS enabled receivers. This paper describes an economic potential AGPS receiver simulated on NI LabVIEW platform. The A-GPS simulator support is provided to the mobile device through Wi-Fi technology using Packed Encoding Rules (PER). The communication between the simulator and the receiver is in accordance with the Secure User Plane Location (SUPL) protocol encapsulated with Radio resource location services (LCS) protocol (RRLP) applies to GSM and UMTS Cellular Networks.

8304A-09, Session 3

Measuring ionizing radiation with a mobile device

M. Michelsburg, T. Fehrenbach, F. Puente León, Karlsruher Institut für Technologie (Germany)

In cases of nuclear disasters it is desirable to know one's personal exposure to radioactivity and the related health risk. Usually, Geiger-Mueller tubes are used to assess the situation. Fitting everyone with such a device in a short period is very expensive. We propose a method to detect ionizing radiation with a mobile consumer device, e.g., a cell phone, using its integrated camera. In emergency cases, millions of existing mobile devices could then be used to monitor the exposure of its owners. In combination with internet access and GPS, measured data can be collected by a central server to build an overview of the situation.

During a measurement, the CCD or CMOS sensor of a mobile device is shielded from surrounding light through an attachment in front of the lens or an internal shutter. The high-energy radiation produces free electrons on the sensor chip which then are analyzed and separated from the sensor noise by using an adaptive threshold on the brightness of the pixels. With radioactive sources present significant increases in detected pixels can be seen. By adapting exposure and integration times a wide variety of radioactive intensities can be measured. Furthermore, the software can make a preliminary estimate on the collected dose of an individual and associated health risks.

8304A-10, Session 3

Design and evaluation of security multimedia warnings for children's smart phones

J. Fruth, S. Tuchscheerer, Otto-von-Guericke-Univ. Magdeburg (Germany)

In this article we introduce a children specific security-warning design approach for smart phones. We focus on those daily dual-task scenarios, where a user is warned by its antivirus application on its smart phone, while he is busy with another task.

Conference 8304A

Smart phones are more and more used by children. Security warnings of anti-virus applications on smart phones are designed to inform user about detected malicious codes. In our opinion common security warnings of those applications are designed for standard users. They are not adequately adapted to children. In respect to this, we developed a first design approach for security warnings for primary school children's smart phones. The warnings are designed in a comic style, including different threat levels, multi-media and textual information.

The developed security warnings were implemented on an iPhone 4 and evaluated in an empiric user study, using basic concepts known from usability testing. The evaluation analysis shows, e.g. that nearly all children liked the warning's design, but some children had problems by interpreting the meaning of the warning. In the future we plan a refinement of our design and the preparation of further studies.

8304A-11, Session 3

Using Wi-Fi hotspots as an intrusion vector into corporate networks

M. Scharsich, F. Holl, Fachhochschule Brandenburg (Germany)

Sending your data over a public Wi-Fi-Hotspot makes it readable to everyone. This is not just known since the FireSheep-Plugin for Firefox. To protect your data you may use secure communication technics like VPN. But what if not the unencrypted data is a problem, but the Hotspot you use, what if the Hotspot is already under control of an attacker? This can be done with minimal financial effort and offers the attacker the possibilities to gather user credentials and to deliver malware on the victim's devices. With these malware the attacker is able to take completely control over the device. The victims usage of VPN enables him to use the secure connection to get into an otherwise protected network just by using the device as a proxy. That way the attacker is able to circumvent the established perimeter protection like Firewalls and IDS and can get deeply into a corporate network.

8304A-12, Session 4

Frame rate up-conversion assisted with camera auto exposure information

L. Liang, B. Hung, G. Dane, QUALCOMM MEMS Technologies, Inc. (United States)

Many algorithms have been proposed for decoder side frame rate up conversion. However, most of them are from video encoding/decoding's point of view. We proposed to utilize the video front end 3A (auto-exposure, auto-white-balance, auto-focus) to assist frame rate up conversion. In this paper, we focus on using auto-exposure information to assist frame rate up conversion process.

In the proposed strategy the video front end 3A information is packetized as the meta data which is attached to the corresponding frame and transmitted together with the main video bit stream to the decoder side for frame rate up conversion assistance.

Video front end auto-exposure control (AEC) is required to set up camera sensor to achieve a brightness within a desired range. Once the brightness in the recorded scene changes, the AEC convergence function is proceeded to readjust the brightness level back to the tolerance range of the luma target. The abrupt brightness change could cause the difficulty of the motion vector computation of the insert frames during frame rate up conversion process.

In the proposed strategy, the intermediate reference frames are used to assist the interpolation of the inserted frames. By using the auto-exposure information, the motion vector interpolation of the inserted frames can be based on the two or more reference frames that has the same brightness level which results in more accuracy and less compuation.

8304A-13, Session 4

Fused Fibonacci-like (p,q) sequences with compression and barcoding applications

S. Agaian, J. Garcia, S. S. Abdul-Kafi, J. T. Gill III, Stanford Univ. (United States)

A Double-base Number System (DBNS) has recently been introduced and investigated. This system has been shown to have some interesting and potentially far-reaching applications in digital filtering, encryption, digital electronics, and image enhancement. In this paper we introduce a new concept of generating parametric number representations by fusing systems using multiplication and addition operations. More specifically, we introduce Fibonacci like (p,q)-sequences and determine their efficiency in representing data. We develop an efficient algorithm to test the sparsity of several fused number representation systems and explore the dual relationship between sparsity and memory. We also look at the initial applications of these representations in data compression and barcoding. Simulation results are presented to demonstrate the performance of the new class of systems. A comparison with commonly used double-base number systems is also presented.

8304A-14, Session 4

White synthesis with user input for color balancing on mobile camera systems

S. Srivastava, C. Xu, E. J. Delp III, Purdue Univ. (United States)

In this paper we extend the custom white balancing technique available on many imaging devices by allowing a user to specify any number of arbitrary colors in the scene. We derive an interpolation technique to assign weights to the arbitrary colors which are then summed to estimate the RGB corresponding to a white target. We obtain the user input by displaying a captured image alongside a color grid of commonly occurring colors. The user specifies color pairs - patches in the scene and corresponding veridical colors on the grid. We then use these pairs to estimate the white with our interpolation method. The white point thus estimated is used to construct a diagonal transform to estimate the camera output under a desired illuminant.

We obtain encouraging results from testing our methods on images acquired under several illumination conditions. Our approach is very suitable for mobile devices because most such devices are equipped with only moderately sophisticated imaging systems and our method allows better color capture with small common-sense user input. Further, we can realize our method on these devices since many such devices have built-in tools for graphical user input. Our method can be useful in photography and image analysis applications.

8304A-15, Session 4

Detection of Symmetric shapes on a mobile device with application to automatic sign interpretation

A. W. Haddad, S. Huang, M. Boutin, E. J. Delp III, Purdue Univ. (United States)

In this paper we present a light-weight method for automatically detecting a segmenting shapes that have rotational symmetry, such as a square or equilateral triangle. The challenges of limited memory, processing power and battery life make many common techniques unsuitable for mobile computation. It is for this reason that we have developed our light-weight approach. We leverage the n-fold symmetry of particular signs - in this case Hazardous Material Placards. The coefficients of the n-th central moment, as defined by Huang, of a given component can be used to accurately determine the symmetries of even a noisy component. This flexibility allows us to choose a method of segmentation which requires no post processing such as dilation or erosion, thus minimizing the complexity of existing segmentation techniques.

Our approach has been tested at distances of up to 250 feet, using a number of camera resolutions, with and without analog zoom. Our approach has been shown to be scale invariant, translation invariant and rotation invariant. The resilience of our method towards these types of transformations make it a good choice for mobile signal processing. The ability to work with a scaled image is particularly important when dealing with the challenges of limited RAM for processes often imposed by mobile systems. Translation and rotation resilience is important because the position and orientation of the camera in space is dependent on the user.

8304A-17, Session 4

Raster image adaptation for mobile devices using profiles

R. Rosenbaum, B. Hamann, Univ. of California, Davis (United States)

With the focus being on digital imagery, this contribution proposes strategies to cope with the heterogeneous viewing hardware in mobile environments. Constrained system resources of most mobile viewing devices require contents that are tailored to the requirements of the user and the capabilities of the device. Appropriate adaptation of the contents is still an unsolved research question. Due to the complexity of the problem, available solutions are either too resource-intensive or inflexible to be applied broadly.

The proposed approach is based on scalable image compression and progressive refinement as well as data and user profiles. A scalable image is created once and multiply used for different kinds of devices and user requirements. Profiles available on server side allow for an image representation that is adapted to the most important resources in mobile computing: screen space, computing power, and the volume of the transmitted data. Options for progressively refining contents thereby allow for a fluent viewing experience during adaptation. Due to its flexibility and low-complexity, the proposed adaptation approach is a much more general solution compared to related approaches. This is underpinned by empirical results obtained by the stated experiments and a discussion of its properties.

Conference 8304B: Multimedia Content Access: Algorithms and Systems VI

Monday 23 January 2012

Part of Proceedings of SPIE Vol. 8304B Multimedia Content Access: Algorithms and Systems VI

8304B-32, Session 5

Searching through photographic databases with QuickLook

C. Cusano, G. Ciocca, R. Schettini, Univ. degli Studi di Milano-Bicocca (Italy); S. Santini, Univ. Autónoma de Madrid (Spain); A. De Polo, F. Tavanti, Fratelli Alinari (Italy)

We present here the results obtained by including within the framework of QuickLook image retrieval system, that exploits a relevance feedback mechanism, a new image descriptor that we called prosemantic feature vector. By coupling the prosemantic features and the relevance feedback mechanism, the user can move in a more rapid and precise way through the feature space toward the intended goal. The prosemantic features are obtained by a two-step feature extraction process. At the first step, low level features related to image structure and color distribution are extracted from the images. At the second step, these features are used as input to a bank of classifiers, each one trained to recognize a given semantic/concept category, to produce score vectors. We evaluated the efficacy of the prosemantic features under a target search task on two datasets: the Benchathlon dataset and a dataset provided by Fratelli Alinari Photo Archive.

8304B-34, Session 5

Large-scale classification of traffic signs under real-world conditions

L. Hazelhoff, I. M. Creusen, CycloMedia Technology B.V. (Netherlands); D. Van de Wouw, P. H. N. de With, Technische Univ. Eindhoven (Netherlands)

Traffic-sign inventories are important to governmental agencies as they facilitate evaluation of traffic-sign locations and are beneficial for road and sign maintenance.

These inventories can be created (semi-)automatically based on street-level panoramic images.

In these images, object detection is employed to detect the signs in each image, followed by a classification stage, to retrieve the specific sign type.

Classification of traffic signs is a complicated matter, since sign types are very similar with only minor differences within the sign, a high number of different signs is involved and multiple distortions occur, including variations in capturing conditions, occlusions, viewpoints and sign deformations.

Therefore, we propose a method for robust classification of traffic signs, based on the Bag of Words approach for generic object classification.

We extend the approach with a flexible, modular codebook to model the specific features of each sign type independently, in order to emphasize at the inter-sign differences instead of the parts common for all sign types.

Additionally, this allows us to model the false detections. Furthermore, analysis of the classification output provides the unreliable results.

This classification system is extensively tested on two large datasets, containing the output of our red triangular and blue circular traffic-sign detectors on street-level panoramic images, obtained from a county-wide database. The introduction of the modular codebook shows a significant improvement for all three sets, where the system is able to classify 99% of the reliable results correctly.

8304B-35, Session 5

Human action recognition using a Markovian conditional exponential model

A. Velivelli, A. G. Hauptmann, Carnegie Mellon Univ. (United States)

We model the sequence of human actions operating an infusion pump using a Markovian conditional exponential model. We divide each video recorded by a camera into video action units.

A video action unit corresponds to the start of a unique human action operation of the infusion pump to the end of that human action operation of infusion pump.

We calculate the MOSIFT features of video action units which combines the spatial and temporal dimensions from videos. We vector quantize the MOSIFT features of video action unit using K means clustering as video codebook elements. We estimate the conditional exponential model parameters from a training set using maximum entropy constraint and use the video codebook elements as maximum entropy constraint features.

We estimate the parameters of the Markovian conditional exponential model from a training set.

This Markovian conditional exponential model has 6 states which correspond to the 6 classes of infusion pump operation.

To find the optimal state sequence of the Markovian conditional exponential model we use the Viterbi algorithm.

This optimal state sequence corresponds to the class label sequence.

The infusion pump operation is recorded from 4 video cameras. We calculate the results of classification of 6 classes of infusion pump operation using the conditional exponential model for the 4 video cameras and also we calculate the results of of classification of 6 classes of infusion pump operation using the Markovian conditional exponential model for the 4 video cameras.

The classification performance of the Markovian conditional exponential model is better than the classification performance of conditional exponential model.

8304B-36, Session 5

Human activity discovery and recognition based on state transitions modeling in persistent surveillance systems

A. Shirkhodaie, V. Elangovan, Tennessee State Univ. (United States)

In this paper, we present a State Transitions-based tracking model for spatiotemporal HVI patterns exploitation with a link to known HVI ontologies. More specifically, this paper discusses development of Adaptive State Transition models in discovering HVI via a Hidden Markov Modeling (HMM) technique. Five categories of state transitions are considered including: Human state transitions of Object handling, Visibility, Entity-entity relation, Entity Postures and Kinematics. A new approach for recognizing static and dynamic HVIs taking plan in dark environments are represented. The proposed approach uses low-cost Kinect depth map cameras for detection, tracking, and characterization of HVI activities. Experimental results with Kinect cameras from both indoor and outdoor are presented. The proposed model, furthermore, generates semantic messages describing the HVI ontology using Zoning of Vehicle (ZoV) technique and also demonstrates the efficiency in Human Activity Discovery & Recognition (HADR) for improved situational awareness.

Conference 8304B

8304B-37, Session 6

Swimmer detection and pose estimation for continuous stroke-rate determination

D. Zecha, T. Greif, R. W. Lienhart, Univ. Augsburg (Germany)

In this work we propose a novel approach to automatically detect a swimmer and estimate his/her pose continuously in order to derive an estimate of his/her stroke rate given that we observe the swimmer from the side. We divide a swimming cycle of each stroke into several intervals. Each interval represents a pose of the stroke. We use specifically trained object detectors to detect each pose of a stroke within a video and count the number of occurrences per time unit of the most distinctive poses (so-called key poses) of a stroke to continuously infer the stroke rate. We extensively evaluate the overall performance and the influence of the selected poses for all swimming styles on a data set consisting of a variety of swimmers.

8304B-38, Session 6

Multi-view face detection based on position estimation on multi-camera surveillance system

C. Huang, National Kaohsiung Univ. of Applied Sciences (Taiwan); J. Chou, J. Syu, S. Wang, National Chiao Tung Univ. (Taiwan)

In this paper, we propose a multi-view face detection system that locates head positions and indicates the direction of each face in 3-D space over a multi-camera surveillance system. To locate 3-D head positions, conventional methods relied on face detection in 2-D images and projected the face regions back to 3-D space for correspondence. However, the inevitable false face detection and rejection usually degrades the system performance. Instead, our system searches for the heads and face directions over the 3-D space using a sliding cube. Each searched 3-D cube is projected onto the 2-D camera views to determine the existence and direction of human faces. Moreover, a pre-process to estimate the locations of candidate targets is illustrated to speed-up the searching process over the 3-D space. In summary, our proposed method can efficiently fuse multi-camera information and suppress the ambiguity caused by detection errors. Our evaluation shows that the proposed approach can efficiently indicate the head position and face direction on real video sequences even under serious occlusion.

8304B-39, Session 7

Mobile visual search

R. Grzeszczuk, Nokia Research Ctr. (United States)

No abstract available

8304B-40, Session 7

Discriminative tag learning at YouTube

G. Toderici, Google Inc. (United States)

No abstract available

8304B-41, Session 7

Revisiting K-means quantization and image object retrieval in an industrial context

R. van Zwol, Yahoo! Inc. (United States)

In this presentation we revisit the k-means algorithm as a means to efficiently quantize local features for large scale image similarity search. We propose a balanced k-means clustering algorithm that allows us to maintain near optimal minimization of the within-cluster distances, while at the same time to balance the number of features assigned to each cluster. We believe that balanced k-means clustering is beneficial in various computer vision applications, and demonstrate one such example: to quantize local feature descriptors for large scale image similarity search. We first formally define the balanced k-means clustering algorithm and perform extensive experimentation on large image data-sets to analyze the impact in terms of mean squared error, balance, and quantization quality. Moreover, we demonstrate that a balanced k-means clustering leads to a significant improvement in retrieval and system performance when applied to large-scale image similarity search based on local feature descriptors.

8304B-42, Session 7

Multimedia technologies for content creation and consumption

Q. Lin, Hewlett-Packard Labs. (United States)

With the growth of images captured by cameras and smart phones, people are increasingly using image-rich content in their communications. While the traditional media is largely created by professionals, and consumed by a large population of people, new multimedia technologies are making it easy for ordinary people to create image-rich content. In addition, new technologies are also making it possible for published content to be personalized for individuals based on their interests. In this presentation, we will discuss how multimedia analysis and composition technologies are used in content creation and consumption applications.

8304B-43, Poster Session

Keyframe generation from cartoon animation using rule-based optical flow

P. Tanapichet, N. Cooharojananone, R. Lipikorn, Chulalongkorn Univ. (Thailand)

This paper proposes a novel method to generate keyframes from cartoon animation with the aim to improve the details and accuracy of contents represented by keyframes. Consider that general techniques on video summarization usually drop some important contents due to its restriction on aspect ratio; this paper thus proposes a new method using panorama technology to add more details to be included in each keyframe. The concept is to mark the time code based on shot boundary and optical flow direction. The period of time between every two consecutive marked time codes is used to form a shot sequence which is actually a sequence of frames. The global and local optical flows are also used to determine how to select the frames and when to stitch the frames together according to the rules. The results of this proposed method are keyframes generated from various types of cartoon animation which are outstanding compared to their comic adaptations.

8304B-44, Poster Session

Adaptive characterization, tracking, and semantic labeling of human-vehicle interactions via multimodality data fusion techniques

A. Shirkhodaie, V. Elangovan, Tennessee State Univ. (United States)

Exploitations of Human-Vehicle Interactions (HVI) can be beneficial for Persistent Surveillance Systems (PSS). In particular, by proper characterization of HVI, certain pertinent threats can be prevented. In this paper, we present a new model for semantic message generation based on fusion of different sensing modalities. An ontology-based approach is proposed for localization and characterization of traceable HVI activities. Generation of HVI Semantic messages undergoes series of refinement including: Human Attributes refinement, Object

Attributes Refinement, Action Attributes refinement, Spatial refinement and Temporal refinement. Fusion of semantic messages takes place at each refinement stages and the protocol used for semantic message generation had also been addressed in this paper. State transitions of Human-Vehicle Interactions and Human Object Interactions had also been discussed here. Furthermore, we have demonstrated that by fusing imaging and acoustic data at decision level, a higher degree of confidence can be realized for describing the HVI activities with appropriate traceability. This paper also presents the results of our experimental work, and demonstrates efficiency and effectiveness of the proposed adaptive fusion technique towards achievement of semantic messages describing HVI activities with different level of abstract and details.

Conference 8305: Visual Information Processing and Communication III

Tuesday 24 January 2012

8305-01, Session 1

A novel distortion model for quadtree coding in high-efficiency video coding

B. Lee, S. Ahn, M. Kim, KAIST (Korea, Republic of)

In this paper, a novel distortion model based on a mixture of multiple Laplacian distributions are presented for the transform coefficients of predicted residues in quadtree coding. The proposed mixture model of multiple Laplacian distributions is tested for the High Efficiency Video Coding (HEVC) Test Model (HM) with quadtree-structured Coding Unit and Transform Unit. The experimental results show that the proposed model achieves more accurate results of distortion estimation than the single probability models.

8305-02, Session 1

Weighted prediction for HEVC

P. Bordes, Technicolor S.A. (France)

HEVC is the new video coding standard developed by ISO MPEG and ITU-T VCEG. As other state-of-the-art block-based inter-prediction codec, it is very sensitive to illumination variations in-between frames. To cope with this limitation, the weighted prediction (WP) tool has been proposed. A comparison of the performance of WP in HEVC and in MPEG-4 AVC/H.264 is carried out. The efficiency of WP is very dependent on the quality of the estimated WP parameters. The different stages of state-of-art WP parameters estimators are discussed and a new algorithm is proposed. It is based on histogram matching with global motion compensation. Several options are evaluated and comparison is made with other existing methods.

8305-03, Session 1

Impact of video parameters on the DCT coefficient distribution for H.264-like video coders

N. Kamaci, G. Al-Regib, Georgia Institute of Technology (United States)

With the existing and emerging advanced video coding standards, video source modeling becomes more important. Particularly, the statistical properties of the transform coded video is of great importance. In this work, we examine the impact of various encoding parameters on the distribution of the discrete cosine transform (DCT) coefficients for H.264-like video coders. We model the distribution of the frame DCT coefficients using the most common Laplacian and Cauchy distributions. We show that the resolution, the quantization levels and the coding type have significant impact on the accuracy of the Laplacian and Cauchy distribution based models. We also show that the transform kernel (4x4 vs 8x8) has little impact. Moreover, we show that for the video sources that have little temporal or spatial detail, such as flat regions, the distribution of the frame DCT coefficients resembles a Laplacian distribution. When the video source exhibits more detail, such as texture and edges, the

distribution of the frame DCT coefficients resembles a Cauchy distribution. The correlation between the details of the video source to the two probability distributions can be used to further improve the estimation of the distribution of the frame DCT coefficients, by using a classification based approach.

8305-04, Session 1

Adaptive loop filter with directional similarity mapping for video coding

P. Lai, F. C. A. Fernandes, SAMSUNG Telecommunications America Inc. (United States)

To improve coding efficiency and visual quality in a state-of-the-art video codec, we present an adaptive loop filtering design, that combines linear spatial filtering and directional similarity filtering with a mapping function. It exploits directional features in video frames by classifying blocks in a frame based on the direction of local gradients. To reduce side-information overhead, symmetric constraints are imposed on the directional filter coefficients according to the gradient direction. To emphasis pixel similarity for explicit adaptation to edges, we use a simple hard-threshold mapping function to avoid artifacts arising from across-edge filtering. Compared to the state-of-the-art approach, our design has fewer filters (4 instead of 16) and smaller worst-case filter size (7×7 as compared to 7 vertical $\times$ 9 horizontal), while achieving better objective coding performance with 0.2% BD-rate gains and improved subjective visual quality, especially along edges.

8305-05, Session 2

Distributed video coding with progressive significance map

W. A. Pearlman, Y. Hu, Rensselaer Polytechnic Institute (United States)

A distributed video coding (DVC) system based on wavelet transform and set partition coding (SPC) is presented in this paper. Conventionally the significance map (sig-map) of SPC is not conducive to Slepian-Wolf (SW) coding, because of the difficulty of generating a side information sig-map and the sensitivity to decoding errors. The proposed DVC system utilizes a higher structured significance map, named progressive significance map (prog-sig-map), which structures the significance information into two parts: a high-level summation significance map (sum-sig-map) and a low-level complementary significance map (comp-sig-map). This prog-sig-map solves the above difficulties and thus makes part of the prog-sig-map (specifically, the fixed-length comp-sig-map) suitable for SW coding. Simulation results are provided showing the improved rate-distortion performance of the DVC system even with very simple system configurations.

8305-06, Session 2

Improving side information generation using dynamic motion estimation for distributed video coding

I. Park, D. W. Capson, McMaster Univ. (Canada)

A new side information generation algorithm using dynamic motion estimation and post processing is proposed for improved distributed video coding. Multiple reference frames are employed for motion estimation at the side information frame generation block of the decoder. After motion estimation and compensation, post processing is applied to improve the hole and overlapped areas on the reconstructed side information frame. The proposed side information method contributes to improve the quality of reconstructed frames at the distributed video decoder. The average encoding time of the distributed video coding is around 15% of H.264 inter coding and 40% of H.264 intra coding. The proposed side information based distributed video coding demonstrates improved performance compared with that of H.264 intra coding.

8305-07, Session 2

Directional frame interpolation for MPEG compressed video

C. Zhao, X. Gao, X. Fan, D. Zhao, Harbin Institute of Technology (China)

Image interpolation is one of the most elementary imaging research topics. A number of image interpolation methods have been developed for uncompressed images in the literature. However, a lot of videos have already been stored in MPEG-2 format or have to be transmitted in MPEG-2 format due to bandwidth limitation. The image interpolation methods developed for uncompressed images may not be effective when directly applied to compressed videos, because on the one hand, they do not utilize the information existed in the coded bitstreams; on the other hand, they do not consider quantization error, which may be dominant in some cases. Inspired by the success of the intra prediction in H.264/AVC and the edge-directed image interpolation methods (such as LAZA and NEDI), we propose a directional frame interpolation for MPEG compressed video. In the proposed method, 8×8 intra blocks in I frames are first classified to the nine block directions in transform domain. Then the interpolation on each block is performed along its block direction. For each block direction, an optimal Wiener filter is trained based on the representative video sequences and then used for its interpolation. In the similar way, for each pixel in an inter block in P or B frames, the interpolation is performed along the direction of its corresponding reference block. The experimental results demonstrate that the proposed method achieves better performance than the traditional linear methods such as Bicubic and Bilinear and the edge-directed methods such as LAZA and NEDI, while keeping low computational complexity which meets the requirement of practical applications.

8305-08, Session 2

A fast intra-prediction method for high-efficiency video coding using Hadamard transform

Y. Kim, George Mason Univ. (United States) and Electronics and Telecommunications Research Institute (Korea, Republic of)

For the high efficiency video compression, intra prediction mode has been increased and it is computationally expensive. In this paper, we propose a fast intra prediction mode based on the estimation of Rate distortion cost using Hadamard transform and early coding unit determination. The proposed method reduce the computational complexity with comparable image quality and bit-rate..

8305-09, Session 3

Lossless description of 3D range models

N. Bayramoglu, A. A. Alatan, Middle East Technical Univ. (Turkey)

The improvements in 3D scanning technologies have led the necessity for managing range image databases. Hence, the requirement of describing and indexing this type of data arises. Since a range model has different properties compared to complete 3D models, we propose a method that relies on Spherical Harmonics Transform (SHT) for retrieving similar models where the query and the database both consist of only range models. Although SHT, is not a novel concept in shape retrieval research for 3D complete models, we utilize it for 2.5D range images by representing the models in a reciprocal world observed from the camera. The difference, as well as the advantage of our algorithm, is being information lossless. In other words, the available shape information is completely exploited for obtaining the descriptor, whereas other mesh retrieval applications utilizing SHT "approximates" the shape that yields information loss. The descriptor is invariant to scale and rotations about z-axis. The proposed technique is tested on a large database having high diversity and its performance of the proposed method is superior to the performance of popular D2 distribution.

8305-10, Session 3

Reference frame selection for loss-resilient depth map coding in multiview video conferencing

B. Macchiavello, C. Dorea, M. Hung, Univ. de Brasília (Brazil) and Hewlett Packard Labs. (United States); G. Cheung, National Institute of Informatics (Japan); W. Tan, Hewlett-Packard Labs. (United States)

Multiview video in "texture-plus-depth" format enables decoder to synthesize freely chosen intermediate views for enhanced visual experience. Nevertheless, transmission of multiple texture and depth maps over bandwidth-constrained and loss-prone networks is challenging, especially for conferencing applications with stringent deadlines. In this paper, we examine the problem of loss-resilient coding of depth maps. Our strategy is to first establish that different depth macroblocks have very different error sensitivity to the reconstructed images. We then selectively employ reference picture selection to provide offer stronger protection for more important depth macroblocks. Methods for assigning weights to depth macroblocks and for modeling distortions in a candidate reference block are provided to allow optimized selection of reference.

Preliminary results show that (1) errors in depth maps with higher border content yields significantly higher distortion in reconstructed images, and (2) reference picture selection applied to depth maps requires significantly less percentage overhead than to images. Both of these suggest that our scheme is promising. In our final paper, we plan to implement the presented optimization into H.264, and present simulation results under various network loss conditions.

8305-11, Session 3

Low-complexity automated depth-order estimation for 2D-to-3D video conversion

R. Klepko, Communications Research Ctr. Canada (Canada)

The increasing popularity of 3D TV creates the desire for more 3D video content. Waiting for there to be an abundance of 3D video content derived from stereoscopic cameras will take too much time. However, there currently exists a vast quantity of 2D video material that can potentially be converted to 3D. This 2D-to-3D conversion is a complex process and so can be costly. Thus, an automated solution that can be achieved with low-complexity would be desirable. Our past research work has already resulted in a real-time 2D-to-3D conversion technique, but this creates pseudo-3D and not accurate 3D. Thus, current research, to be described in this paper, focuses on improving the accuracy of the 3D by implementing and incorporating a technique composed of a multi-step process to determine the depth-order of objects, with respect to the camera, in each video sequence frame. The multi-step process can be summarized as follows: edge detection; determine if an edge point belongs to a moving edge to identify the occlusion boundary; determine which of the left or right side block region adjacent to the edge moves with the edge, and by deduction determines the order; segment a frame and assign block regions to segments; cluster segments into objects; globally assign depth-order; and modify surrogate depth map to create more accurate depth map. Test results show that this is a very effective and fast technique for deriving the depth-order of objects.

8305-12, Session 3

Block-layer, optimal bit allocation based on constant perceptual quality

C. Wang, X. Mou, Xi'an Jiaotong Univ. (China); L. Zhang, The Hong Kong Polytechnic Univ. (China)

Bit allocation is a key issue in image/video coding. We suggest that the bit allocation among the macro blocks can be optimized by aiming at the constant perceptual quality (CPQ) inside an image/a frame. Based

on the MINMAX criterion, we propose a CPQ based block-layer bit allocation method, which is a multi-pass scheme. The results show that the CPQ can improve the encoding performance obviously.

8305-13, Session 4

Patch-wise ideal stopping time for anisotropic diffusion

H. Talebi, P. Milanfar, Univ. of California, Santa Cruz (United States)

Data-dependent filtering methods are powerful techniques for image denoising. These algorithms are mostly patch-wise, and also employ local and non-local similarities in the signals. All of these restoration methods work based on the same framework in which some data-adaptive weights are assigned to each pixel contributing to the filtering. Beginning with any base procedure (nonlinear filter), repeated applications of the same process can be interpreted as a discrete version of anisotropic diffusion. Diffusion filtering gradually removes noise in each iteration, but also takes away latent details from the underlying signal. Choosing a small iteration number preserves the underlying structure, but also does little denoising. On the other hand, a large iteration number tends to over-smooth and remove noise and high frequency details at the same time. As such, a nature question is "What is the best stopping time in iterative data-dependent filtering?". This is the general question we address in this paper. If we clairvoyantly had the mean-square error (MSE) function in each iteration, finding the optimal stopping time would be possible, as this would correspond to the minimum of the MSE. To develop our new method, we estimate the MSE in each image patch. This estimate is used to characterize the effectiveness of the iterative filtering process, and its minimization yields the ideal stopping time for the diffusion process.

8305-14, Session 4

Video attention deviation estimation using inter-frame visual saliency map analysis

Y. Feng, G. Cheung, National Institute of Informatics (Japan); P. Le Callet, Polytech' Nantes (France); Y. Ji, National Institute of Informatics (Japan)

A viewer's visual attention during video playback is the matching of his eye gaze movement to the changing video content over time. If the gaze movement matches the video content (e.g., follow a rolling soccer ball), then the viewer keeps his visual attention. If the gaze location moves from one video object to another, then the viewer shifts his visual attention. A video that causes a viewer to shift his attention often is a "busy" video. Determination of which video content is busy is an important practical problem; a busy video is difficult for encoder to deploy region of interest (ROI)-based bit allocation, and hard for content provider to insert additional overlays like advertisements, making the video even busier. One way to determine the busyness of video content is to conduct eye gaze experiments with a sizable group of test subjects, but this is time-consuming and cost-ineffective. In this paper, we propose an alternative method to determine the busyness of video---formally called video attention deviation (VAD)---by analyzing the spatial visual saliency maps of the video frames across time. We first derive transition probabilities of a Markov model for eye gaze using saliency maps of a number of consecutive frames. We then compute steady state probability of the saccade state in the model---our estimate of VAD. We demonstrate that the computed steady state probability for saccade using saliency map analysis matches that computed using actual gaze traces. Further, our analysis can also be used to segment video into shorter clips of different degrees of busyness by computing the Kullback-Leibler divergence using consecutive computed gaze model statistics.

8305-15, Session 4

Robust grid registration for non-blind PSF estimation

J. Simpkins, R. L. Stevenson, Univ. of Notre Dame (United States)

Given a blurred image of a known test grid and an accurate estimate of the unblurred image, it has been demonstrated that the underlying blur kernel (or point-spread function, PSF) can be reliably estimated. Unfortunately, the estimate of the sharp image can be sensitive to common imperfections in the setup used to obtain the blurred image, and errors in the image estimate result in an unreliable PSF estimate.

We propose a robust ad-hoc method to estimate a sharp prior image, given a blurry, noisy image of a test grid taken in imperfect lab and lighting conditions. The proposed algorithm is able to reliably reject superfluous image content, can deal with spatially-varying lighting, and is insensitive to errors in alignment of the grid with the image plane.

We demonstrate the algorithm's performance through simulation, and with a set of test images. We also show that our grid registration algorithm leads to improved PSF estimation and deblurring, compared to an affine registration using spatially invariant lighting correction.

8305-16, Session 4

Fast pseudo-semantic segmentation for joint region-based hierarchical and multiresolution representation

R. Sekkal, C. Strauss, F. Pasteau, M. Babel, O. Déforges, Institut National des Sciences Appliquées de Rennes (France)

In this paper, we present a new scalable segmentation algorithm characterized by region-based hierarchy and resolution scalability. Most of the proposed algorithms applied either multiresolution segmentation following a coarse to fine segmentation approach or a hierarchical segmentation with fine to coarse segmentation. Our approach combines both multiresolution and hierarchical segmentation. We consider the image as a set of different images at different levels of resolution, where at each level a hierarchical segmentation is performed. Multiresolution implies that a segmentation of one given level is reused in further segmentation process operated at the next level. Each level of resolution provides a Region Adjacency Graph (RAG) that describes the neighborhood relationships between regions within one given level of the multiresolution representation. Region labels consistency is preserved thanks to a dedicated projection algorithm based on inter-level relationships. The main advantage of our approach is the low computational complexity, as the segmentation process only requires low resolutions of the image. Experiments show that we can obtain very effective results compared with the state of the art together with less computation time.

8305-17, Session 5

Optimal local dimming for LED-backlit LCD displays via linear programming

X. Shu, X. Wu, McMaster Univ. (Canada); S. Forchhammer, Technical Univ. of Denmark (Denmark)

LED-backlit LCD displays hold the promise of improving the image quality while reducing the energy consumption with signal-dependent local dimming. To fully realize such potentials we propose a novel local dimming technique that jointly optimizes the intensities of LED back lights and the attenuations of LCD pixels. The objective is to minimize the distortion in luminance reproduction due to the leakage of LCD and the coarse granularity of the LED lights. The optimization problem is formulated as one of linear programming, and both exact and approximate algorithms are proposed. Simulation results demonstrate superior performances of the proposed algorithms over the existing local dimming algorithms.

Conference 8305

8305-18, Session 5

Gestures for natural interaction with video

N. Fourati, E. Marilly, Alcatel-Lucent Bell Labs. Villarceaux (France)

In the context of immersive communications that aim to enable natural experiences and interactions among people, objects, and environment, we propose a method enabling natural video interactions through hand gesture recognition between users and a video meeting system. An end-to-end study is done starting from the gesture recognition algorithms to the user evaluation for the validation.

Identification of gestures to recognize and their associated functionalities has been done from a user survey. In this survey, two concepts, that are often confused, have to be distinguished: hand posture and hand gesture (i.e. static versus dynamic).

Therefore, our hand gesture recognition process is composed of 2 main tasks: the hand posture recognition (i.e. skin segmentation, background subtraction, regions combination, features extraction and classification) and the hand gesture recognition (tracking and recognition). In this approach, we highlight the interest to combine a signal similarity study with a data mining tool for dynamic gesture recognition. We have also combined color and motions cues to avoid the detection of static objects.

Our perspectives are concentrated on the experimentation and user evaluation in order to improve our approach, taking into account user's feedback and performance analysis in different environments and for different users.

8305-19, Session 5

Improving underwater visibility using vignetting correction

K. Sooknanan, A. Kokaram, Trinity College Dublin (Ireland)

Underwater survey videos of the seafloor are usually plagued with heavy vignetting (radial falloff) outside of the light source's beam footprint on the seabed. These surveys are important for assessing the biological environment. In this paper we propose a novel multi-frame approach for removing this piecewise vignetting phenomenon which involves estimating the light source's footprint on the seafloor, and the parameters for our proposed vignetting model. Within the footprint, we leave the image contents as is, whereas outside this region, we perform vignetting correction. We verify our algorithm with both synthetic and real data, and then compare it with an existing technique. Results obtained show significant improvement in the brightness levels.

8305-20, Session 5

Defect pixel interpolation for lossy compression of camera raw data

M. Schöberl, Friedrich-Alexander-Univ. Erlangen-Nürnberg (Germany); J. Keinert, Fraunhofer-Institut für Integrierte Schaltungen (Germany); J. Seiler, Friedrich-Alexander-Univ. Erlangen-Nürnberg (Germany); S. Foessel, Fraunhofer-Institut für Integrierte Schaltungen (Germany); A. Kaup, Friedrich-Alexander-Univ. Erlangen-Nürnberg (Germany)

The image processing pipeline of a traditional digital camera is often limited by processing power. A better image quality could be generated if more complexity were allowed. In a raw data workflow most algorithms are executed off-camera. This allows the use of more sophisticated algorithms for increasing image quality while reducing camera complexity. However, this requires a major change in the processing pipeline: a lossy compression of raw camera images might be used early in the pipeline. Subsequent off-camera algorithms then need to work on modified data. We analyzed this problem for the interpolation of defect pixels. We found that a lossy raw compression spreads the error from uncompensated defects over many pixels. This leads to a problem as this larger error cannot be compensated for after compression. The use of high quality, high complexity algorithms in the

camera is also not an option. We propose a solution to this problem: Inside the camera only a simple and low complexity defect pixel interpolation is used. This significantly reduces the compression error for neighbors of defects. We then perform a lossy raw compression and compensate for defects afterwards. The high complexity defect pixel interpolation can be used off-camera.

This leads to a high image quality while keeping the camera complexity low.

8305-21, Session 6

Cubic-panorama image dataset compression

S. Salehi, E. Dubois, Univ. of Ottawa (Canada)

This work is part of the NAVIRE project at the University of Ottawa which aims at developing the necessary technology to allow a user to virtually walk through in an image-based representation of a remote environment. Our goal is to propose an efficient method to compress such high volume of information, keeping the desired rate, distortion, and random access requirements in mind. After the acquisition and preprocessing stage, a group of six raw full RGB images can be used to create a basis panorama which consists of six side images in cubic format. Following the image dataset analysis stage, visual data will be ready for compression. We apply and compare the two major existing approaches from the literature, i.e., a standardized method based on H.264/MPEG4 AVC and an existing wavelet based method called Dirac. The problem of indexing is addressed considering the compression efficiency, random access, and other requirements of our application. Advantages of using B frames are shown. Based on the abovementioned considerations an appropriate bitstream syntax is introduced. Disparity estimation which plays an important role in any compression scheme is given specific attention and an efficient method based on our previous work on Epipolar geometry is presented.

8305-22, Session 6

Lossless halftone image compression using adaptive context template update

S. Park, D. Choi, J. Yoon, Samsung Electronics Co., Ltd. (Korea, Republic of)

In electronic printing system, many research efforts have been made to compactly express halftone images, based on a context-based binary arithmetic coding. For example, in the standard JBIG, 10 bit-sized context template was used to reduce conditional entropy. Also, a template modification scheme was presented, which uses the autocorrelation of input image. Combined with 12 bit-sized context templates, Reavy and Boncelet presented a block arithmetic coding (BAC), which uses fixed length arithmetic codes. Moreover, a context weighting scheme was introduced to find the best context model among several context templates.

In this work, an adaptive context template design scheme is presented for efficient halftone image compression. As large-sized templates are prohibitive for hardware implementation, the proposed scheme keeps the template size as same as the standard JBIG. Also, the proposed scheme modifies the template shape in each pixel traversal. Specifically, the template is changed by selecting binary pixels having high correlations to the current pixel. The proposed scheme outperforms the conventional JBIG with 29 % bit saving, while using only 1,024 context conditions as same as the JBIG.

8305-33, Session 7

Recognition of sport players' numbers using fast-color segmentation

C. Verleysen, C. De Vleeschouwer, Univ. Catholique de Louvain (Belgium)

This paper builds on a prior work for player detection, and proposes an efficient and effective method to distinguish among players based on the numbers printed on their jerseys. To extract the numbers, the dominant colors of the jersey are learnt during an initial training phase and used to speed up the segmentation of the candidate digit regions. An additional set of criteria considering the relative position and size (compared to the player bounding box) and the density (compared to the digit rectangular support) of the digit are used to filter out the regions that obviously do not correspond to a digit. Once the plausible digit regions have been extracted, their recognition is based on feature-based classication. A number of original features are proposed to increase the robustness against digit appearance changes, resulting from the font thickness variability and from the deformations of the jersey during the game. Finally, the efficiency and the effectiveness of the proposed

method are demonstrated on a real-life basketball dataset. More than 50% of the jersey samples that can be visually recognized are detected as numbers with 93% of correct classication.

8305-34, Session 7

On the use of clustering for resource allocation in wireless visual sensor networks

A. V. Katsenou, L. P. Kondi, K. E. Parsopoulos, Univ. of Ioannina (Greece)

In the present paper, we focus on the problem of quality-driven cross-layer optimization of Direct Sequence Code Division Multiple Access (DS-CDMA) wireless visual sensor networks (WVSNs). We consider a centralized topology, where each sensor transmits directly to the centralized control unit (CCU), who manages the network resources. In real environments, the visual sensors view and transmit scenes with varying motion amounts, thus each recorded video has its individual motion characteristics. Our aim is to enable the CCU to allocate jointly the transmission power and source-channel coding rates for each WVSN node, under certain quality-driven criteria and the constraint of a constant chip rate. We consider two approaches for the cross-layer optimization scheme. In the first approach, the optimal set of network resources is assigned to each node according to its individual motion characteristics. In the second approach, the nodes are partitioned into clusters according to the amount of motion in the recorded scenes, and the same network resources are assigned to all nodes within a cluster. The resulting mixed-integer optimization problem is solved using the Particle Swarm Optimization (PSO) algorithm. The experimental results demonstrate the quality/complexity tradeoff for the two approaches.

8305-35, Session 7

Kalai-Smorodinsky bargaining solution for optimal resource allocation over wireless DS-CDMA visual sensor networks

K. Pandremmenou, L. P. Kondi, K. E. Parsopoulos, Univ. of Ioannina (Greece)

Surveillance applications usually require high levels of video quality, resulting in high power consumption. The existence of a well-behaved scheme to tradeoff video quality and power consumption is crucial for the system performance. In this work, we adopt the game theoretic approach of Kalai-Smorodinsky Bargaining Solution (KSBS) to deal with the problem of optimal resource allocation in a multinode wireless visual sensor network (VSN). The Direct Sequence Code Division Multiple Access (DS-CDMA) method is used for channel access. Moreover, a cross-layer optimization design, which employs a central processing server, accounts for the overall system efficacy through all network layers. The task assigned to the central server is the communication with the nodes and the joint determination of their transmission parameters. The KSBS is applied to non-convex utility spaces, distributing the source coding rate, channel coding rate and transmission powers fairly and efficiently among the nodes. The transmission powers can be assigned continuous values, whereas the source and channel coding rates can only take discrete values.

Investigating the provided experimental results, we demonstrate the merits of KSBS over competing fairness policies.

8305-36, Session 7

State-of-the-art lossy compression of Martian images via the CMA-ES evolution strategy

F. W. Moore, B. Babb, Univ. of Alaska Anchorage (United States); S. Aldridge, The Univ. of Southern California (United States); M. R. Peterson, Univ. of Hawai'i at Hilo (United States)

The research described in this paper uses the CMA-ES evolution strategy to optimize matched forward and inverse transform pairs for the compression and reconstruction of images transmitted from Mars rovers under conditions subject to quantization error. Our best transforms outperform both the integer and floating-point implementations of the 2/6 wavelet, substantially reducing error in reconstructed images without allowing increases in compressed file size. This result establishes a new state-of-the-art for the lossy compression of images transmitted over the deep-space channel.

8305-37, Session 7

Spatially adaptive superresolution using the optimal recovery framework

A. J. Shaik, S. D. Cabrera, The Univ. of Texas at El Paso (United States)

Multi-Frame image restoration or Super-Resolution (SR) is the process of combining multiple Low-Resolution (LR) images in order to reconstruct a single High-Resolution (HR) image. Our previous approach to SR based on the framework of optimal recovery is extended to a spatially-adaptive scheme whereby the block-by-block processing is modified based on the properties of the corresponding local LR image data. The bandwidth parameter is adapted based on the local variance of each corresponding block from all LR images that contribute to the HR image block to be reconstructed. The optimal regularization parameter for the reconstruction of each HR output block is calculated using the Generalized Cross Validation (GCV) approach chosen based on the bandwidth as well as the contributing local LR data pixels. Simulation results show the superiority of the adaptive scheme over the fixed optimal recovery based super-resolution with and without the use of overlapping blocks. A brief comparison with another state-of-the-art SR scheme is also included and we obtain comparable results.

8305-25, Session 8

A semi-automatic traffic sign detection, classification, and positioning system

I. M. Creusen, L. Hazelhoff, P. H. N. de With, CycloMedia Technology B.V. (Netherlands) and Technische Univ. Eindhoven (Netherlands)

The availability of large-scale databases containing street-level panoramic images offers the possibility to perform semi-automatic surveying of real-world objects such as traffic signs. These inventories can be performed significantly more efficiently than using conventional methods. Governmental agencies are interested in these inventories for maintenance and safety reasons. This paper introduces a complete semi-automatic traffic sign inventory system. The system consists of several components. First, a detection algorithm locates the 2D position of the traffic signs in the panoramic images. Second, a classification algorithm is used to identify the traffic sign. Third, the 3D position of the traffic sign is calculated using the GPS position of the photographs. Finally, the results are listed in a table for quick inspection and are also visualized in a web browser.

8305-27, Session 8

Compression of 2D navigation views with rotational and translational motion

D. Springer, Friedrich-Alexander-Univ. Erlangen-Nürnberg (Germany); F. Simmet, D. Niederkorn, Audi AG (Germany); A. Kaup, Friedrich-Alexander-Univ. Erlangen-Nürnberg (Germany)

In-car navigation systems have grown in complexity over the recent years, most notably in terms of route calculation, usability and graphical rendering. In order to guarantee correct system behavior, navigation systems need to be tested under real operating conditions, i.e. with field-tests on the road. In this paper, we will focus on a fast and embedded compression solution for 2D navigation renderings, so that field-tests can be archived and handed over to software engineers for subsequent evaluation. No parameters from the rendering procedure are available since access to the system is limited to the LVDS display signal. We show how to reconstruct all relevant motion parameters of the scenes with low computational complexity and develop a Global Motion Estimation (GME) method as support for a subsequent video encoding process. By integrating rate-distortion optimization concepts into our scheme, we can efficiently omit the segmentation of static and non-static areas. The proposed GME technique is inherently suited for parallel implementation, which is demonstrated on an embedded OMAP4430 SoC.

8305-31, Session 8

Survey of imaging applications used in roadway transportation

N. Manikoth, ACS, A Xerox Co. (United States); R. P. Loce, W. Wu, E. Bernal, Xerox Corp. (United States)

No abstract available.

8305-26, Session 9

Traffic camera markup language (TCML)

Y. Cai, Carnegie Mellon Univ. (United States)

Rapidly growing traffic cameras revolutionize transportation management systems. They also enable peripheral applications such as driver's information, emergency response, and pollution control. Current traffic imaging techniques are lacking in standardization and flexibility, presenting difficulties in interfacing with other systems such as optimization models and mobile phones. Furthermore, most camera signals are not registered with real-time spatiotemporal databases including conventional traffic sensors (e.g. Radar) and contemporary sensors (e.g. Bluetooth).

Here I will present a novel video markup language for articulating semantic traffic data from surveillance cameras. The markup language includes three layers: traffic flow descriptions, geographical descriptions, and application interface descriptions. I will also present case studies, including the HOV lane safety system, highway flow management and traffic situation navigation system, using Traffic Camera Markup Language.

8305-28, Session 9

On-board side pedestrian detection for automotive active safety system

R. Cheng, Y. Zhao, X. Wang, J. Xu, S. Lv, Peking Univ. (China)

Automotive Active Safety(AAS) is the main branch of intelligence automotive study and pedestrian detection is the key problem of AAS, because it is related with the casualties of most vehicle accidents. For on-board pedestrian detection algorithms, the main problem is to balance efficiency and accuracy to make the on-board system available in real scenes, so a new side pedestrian detection method considered common urban scene that people crossing the street is proposed.

The approach includes two steps, detection and verification. In detection step, Haar feature and a cascade of stage classifiers trained by Adaboost are applied. In verification step, HOG feature and SVM classifier are used to refine false positives. After these two steps, get a result with both high detection rate and low false alarm rate. To speed up the processing time, we use divide-window method together with OCS algorithm and scaling method in detection step. With a new dataset using an onboard camera driving through urban environment to shoot side pedestrians on zebra. This approach performs an on-board available result on side pedestrian detection. A novel method focus on pedestrian's leg feature applied with Haar cascade classifier and HOG descriptors is also explored, it also shows good performance.

8305-29, Session 9

Passive detection of heavily laden vehicles

T. McKay, C. Salvaggio, P. S. Salvaggio, J. Faulring, D. M. McKeown, Rochester Institute of Technology (United States); A. J. Garrett, D. Coleman, L. Koffman, Savannah River National Lab. (United States)

The Digital Imaging and Remote Sensing Laboratory (DIRS) at the Rochester Institute of Technology, along with the Department of Energy is investigating passive methods to quantify the weight of heavily laden vehicles. This paper investigates multiple vehicle signatures including brake temperature, tire temperature, engine temperature, acceleration and deceleration rates, engine acoustics, vehicle stability, suspension response, tire deformation and vibrational response as possible indicators of vehicle weight. Our investigation into these variables includes building and implementing a sensing system for data collection as well as multiple full-scale vehicle tests. This sensing system includes; infrared video cameras, triaxial accelerometers, microphones, video cameras and thermocouples. The full scale testing includes both a medium size dump truck and tractor-trailer truck on closed courses with loads spanning the full range of the vehicle's capacity. Statistical analysis of the collected data is used to determine the effectiveness of each of the signatures for characterizing the weight of a vehicle. The final sensing system will monitor multiple heavy vehicle signatures and combine the results to achieve a more accurate measurement than any of the signatures could provide alone.

8305-30, Session 9

Application of the SNoW machine learning paradigm to a set of transportation imaging problems

P. Paul, A. Burry, Xerox Corp. (United States); Y. Wang, Rochester Institute of Technology (United States); V. Kozitsky, Xerox Corp. (United States)

No abstract available.

8305-32, Session 9

Image simulation for automatic license plate recognition

R. Bala, Y. Zhao, A. Burry, V. Kozitsky, Xerox Corp. (United States); C. Saunders, Xerox Research Ctr. Europe Grenoble (France)

No abstract available.