

### **Message from the Program and General Chairs**

The *Pacific-Rim Symposium on Image and Video Technology* (PSIVT) is a symposium series which was started in 2006. The PSIVT in 2010 (PSIVT 2010) is the fourth issue of the series and is to be held in Singapore 14-17 November 2010. The symposium has provided a forum for presenting and discussing the latest research and development in image and video technology and explored possibilities and future directions in the field. PSIVT 2010 continues to attract researchers, artists, developers, educators, performers, and practitioners of image and video technology from the Pacific-Rim and around the world.

In PSIVT 2010, we have made a number of changes compared to the past three issues. The program committee is made up of area chairs and a technical program committee. The technical areas of PSIVT 2010 cover Image/Video Coding and Transmission; Image/Video Processing and Analysis; Image/Video Storage, Retrieval, and Authentication; Image/Video Formation, Display, and Hardware; Graphics, Visualization and Pervasive Computing; Scene Recognition and Understanding; Multiview Image/Video Coding and Processing; and Computer Vision Applications. For each technical area, at least two area chairs are assigned to coordinate the paper review process with their own team of reviewers selected from the technical program committee. The review process was double-blind in which author names and affiliations were not made known to area chairs and reviewers. Reviewers also did not know their area chairs. Each paper received at least three reviews. The reviewers were asked to submit a detailed review report and the area chairs made the final decisions on the acceptance of papers with little moderation from the program chairs. A total of 154 paper submissions underwent this review process. In the end, 71 papers were finally included in the symposium program for presentation. We also invited 13 papers addressing specific topics in PSIVT, forming 4 special sessions covering the following topics: Social Media Content Mining, Multimedia Browsing and Understanding, Graphics and Visualization and lastly a very special mix of Image/Video Technology Meets Art. This special mix aims to promote the awareness of the artists' perspective on image and video technology and the technologists' contribution to art with such technology. The format of paper presentation is also slightly different in PSIVT 2010. In addition to traditional oral presentations, there are short oral spotlight presentations with posters. The papers scheduled for traditional oral presentation are also encouraged to prepare posters for increased interaction with participants.

The proceeding of PSIVT 2010 is also for the first time published by the Conference Publishing Services of the IEEE Computer Society. All presented papers will appear in IEEE Xplore and CSDL. Among the papers included in the proceeding as well as the symposium program, 19% is from

mainland China, 14% from Singapore, 12% from Taiwan, 11% from Japan with the remaining coming from other countries in the Pacific-Rim region as well as a number from Europe and North America. We also invited three keynote speakers, two of them IEEE Fellows, to deliver keynote speeches on areas such as Visual Signal Processing and Recognition, Intelligent Video Surveillance, and 3DTV.

Singapore is a modern metropolitan city and country, and it is often known as the "garden city". We hope all the participants will enjoy the high quality program put together by the PSIVT 2010 organizing committee and also the beautiful city nation of Singapore.

**General Chairs: Kap Luk Chan, Akihiro Sugimoto, Hanqing Lu**

**Program Chairs: Kai-Kuang Ma, Han Wang, Yasuyuki Matsushita**

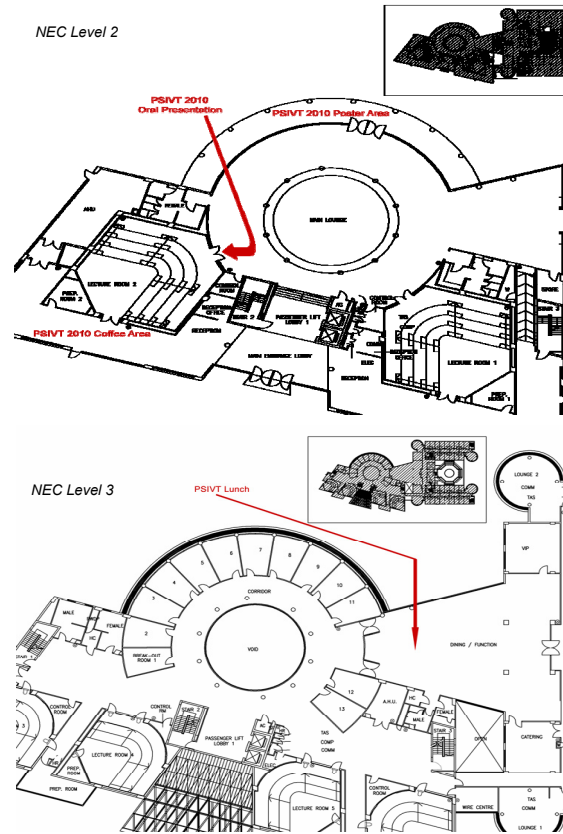
**PSIVT 2010 Program Overview**

<b>15-Nov-10</b>		Session Name
09:00-09:45	Keynote 1	Locally Adaptive Regression Kernels (LARK) for Visual Signal Processing and Recognition
09:45-10:00	Break 1	
10:00-11:00	Short Oral 1 (+Poster)	Detection, Recognition, and Image Retrieval
11:00-12:20	Oral 1	Biometrics
12:20-01:20	Lunch	
01:20-02:05	Keynote 2	Intelligent Video Surveillance
02:05-03:25	Oral 2	Computer Vision (I)
03:25-03:40	Break 2	
03:40-04:40	Short Oral 2 (+Poster)	Computer Vision (II)
04:40-06:00	SP 1	Social Media Content Mining

<b>16-Nov-10</b>		Session Name
09:00-09:45	Keynote 3	3DTV and Realistic Broadcasting Services
09:45-10:00	Break 3	
10:00-	Short Oral 3	Image/Video Coding
11:00-12:20	Oral 3	Video Coding
12:20-1:20	Lunch	
01:20-	Short Oral 4	Image/Video Processing
02:20-03:40	Oral 4	Object Detection and Recognition
03:40-03:55	Break 4	
03:55-05:15	SP 2	Multimedia Browsing and Understanding
07:00-09:00	Banquet	

<b>17-Nov-10</b>		Session Name
09:00-10:20	Oral 5	Super-resolution and Image Enhancement
10:20-10:35	Break 5	
10:35-11:55	Oral 6	Image Enhancement and Image Filter
11:55-12:55	Lunch	
12:55-02:15	Oral 7	Video Surveillance
02:15-03:35	SP 3	Image/Video Technology Meets Art
03:35-03:50	Break 6	
03:50-05:10	SP 4	Graphics and Visualization

### PSIVT 2010 Venue Map



### 10:00AM-11:00AM Short Oral 1 Detection, Recognition, and Image Retrieval

**Title:** One-shot Recognition Using Unsupervised Attribute-Learning

**Authors:** Zhenyu Guo, Z. Jane Wang

**Abstract:** It has been shown that incorporation of human-specified high-level description of the target objects, e.g. labeled prior-knowledge data, can increase the performance of one-shot recognition. In this paper, we introduce latent components as a high level representation of the original objects and propose a cascade model for one-shot image recognition based on latent components learned by Hierarchical Dirichlet Process (HDP). In the proposed approach, instead of solving an optimization problem in the training stage, the latent high-level components are learned efficiently in an unsupervised way from unlabeled prior-knowledge data. Motivated by the facts that HDP is an infinite mixture model proposed in the literature for document modeling that can infer the unknown mixture components and the number of components from the data, and that bag-of-feature model is a standard representation in document retrieval and computer vision areas, we adopt HDP model to infer the mixture components (like latent topics in documents) for target images from unlabeled image visual word vocabulary, and we then train a classifier to associate the components with class labels. The superior performances of the proposed one-shot recognition method are illustrated by testing the Caltech category dataset and the "Animals with Attributes" dataset.

**Title:** Optimal Training Set Selection for Video Annotation

**Authors:** Guo-Xiang Hong, Chung-Lin Huang, Shih-Chung Hsu, Chi-Hung Tsai

**Abstract:** Most learning-based video semantic analysis methods require a large training set to achieve good performances. However, annotating a large video is labor-intensive. This paper introduces how to construct the training set and reduce user involvement. There are four selection schemes proposed: clustering-based, spatial dispersiveness, temporal dispersiveness, and sample-based which can be used to construct a small size and effective training set. If the selected training data can represent the characteristic of the whole video data, the classification performance will be better even when the size of the training set is smaller than that of the whole video data. To verify the best selected samples for training a semantic model, we use SVM to classify the category of the test samples. Five different categories: person, landscape, cityscape, map and others are tested. Experimental results show that these methods are effective for training set selection in video annotation, and outperform random selection.

**Title:** Detecting the Presence of Stationary Objects from Sparse Stereo Disparity Space

**Authors:** Tiandi Duan, Wei Huang, Martin Constable

**Abstract:** This paper presents a novel approach to detecting the presence of objects in a scene from the 3D sparse disparity space obtained by stereo matching. The use of stereo imaging makes the proposed method particularly useful for detecting stationary objects without the need of learning the appearance patterns on an object or the background. Our approach is based on the fact that sparse image features on an object exhibit cluster structures in the 3D disparity space and this reveals the pres-

ence of the object. Hence, we propose to use spectral clustering for grouping matched Scale Invariant Feature Transform (SIFT) interest points in the disparity space and to automatically determine the number of groups and their positions. For grouping matched edge points in the disparity space, a Gaussian mixture model is proposed for its computational efficiency. Experimental results show that our proposed methods can accomplish the task well.

**Title: 3D Face Recognition Using Multi-level Multi-Feature Fusion**

**Authors:** Cuicui Zhang, Keiichi Uchimura, Caiming Zhang, Gou Koutaki

**Abstract:** This paper proposed a novel 3D face recognition algorithm using multi-level multi-feature fusions. A new face representation method named average edge image is proposed in addition to traditional ones such as maximal principal curvature image and range image. In the matching process stage, a new weight calculation algorithm based on the sum rule is presented for feature fusion and match score fusion in order to improve the matching precision. Depending on the complementary characteristic of feature fusion and match score fusion, a combination of them named two-level fusion is proposed. Experiments are conducted using our own 3D database consisting of nearly 400 samples. Mesh simplification is utilized for data reduction. Recognition results show that the new weight calculation method improves the recognition accuracy and the two-level fusion algorithm performs better than feature fusion and match score fusion.

**Title: Vector Quantization Based Index Cube Model for Image Retrieval**

**Authors:** B. Janet, A.V. Reddy, S. Dominic

**Abstract:** We propose a vector quantization (VQ) based index cube model for content based image retrieval. VQ captures the pixel intensity and the spatial information of the image blocks. An indexing and retrieval algorithm is implemented and different similarity measures are evaluated with the precision and recall curves. It can be used for content based image retrieval in image databases using the incremental codebook generation process. The index is scalable as new images can be easily appended to the index. The retrieval time is reduced as there is no processing of the query image before retrieval.

**Title: Locating Corresponding Regions in Urban Environments**

**Authors:** Jimmy Addison Lee, Kin-Choong Yow

**Abstract:** This paper presents a novel approach to locate corresponding regions between two views in urban environments despite the presence of repetitive structures and widely separated views. First we extract hypotheses of building facades, each defined by a rectangular region. The inputs from each pair of regions in two images derive a projective transformation model. Extracted lines and points are used to evaluate the transformation model by voting on the correctness of the model. For each region, we find its corresponding region with the highest votes. Thereafter, we have a set of most corresponding pairs of regions. We take those frequently occurred transformation models (1-3) as correct models since many of the regions in the set come from the same plane, or else they are outliers. We redo the finding most corresponding regions step with the correct transformation models for those outliers with higher votes than the best one among each correct model group. Lastly,

we choose the best pair of most corresponding regions with the highest votes among each correct model group. Experimental results show that our approach is effective and reliable in the case of viewpoint, orientation and scale changes.

**Title: Semantic Segmentation and Object Recognition Using Scene-Context Scale**

**Authors:** Yousun Kang, Hiroshi Nagahashi, Akihiro Sugimoto

**Abstract:** Scene-context plays an important role in scene analysis and object recognition. Among various sources of scene-context, we focus on scene-context scale, which means the effective region size of local context to classify an image pixel in a scene. This paper presents semantic segmentation and object recognition using scene-context scale. The scene-context scale can be estimated by the entropy of the leaf node in multi-scale text on forests. The multi-scale text on forests efficiently provide both hierarchical clustering into semantic textons and local classification depending on different scale levels. For semantic segmentation, we combine the classified category distributions of scene-context scale with the bag-of-textons model. In our experiments, we use MSRC21 segmentation dataset to assess our segmentation algorithm and show that the usage of the scene-context scale improves recognition performance.

**Title: Automated Detection of Fish Bones in Salmon Fillets Using X-ray Testing**

**Authors:** Domingo Mery, Iván Lillo, Hans Loebel, Vladimir Rizzo, Alvaro Soto, Aldo Cipriano, José Miguel Aguilera

**Abstract:** X-ray testing is playing an increasingly important role in food quality assurance. In the production of fish fillets, however, fish bone detection is performed by human operators using their sense of touch and vision which can lead to misclassification. In countries where fish is often consumed, fish bones are some of the most frequently ingested foreign bodies encountered in foods. Effective detection of fish bones in the quality control process would help avoid this problem. For this reason, we developed an X-ray machine vision approach to automatically detect fish bones in fish fillets. This paper describes our approach and the corresponding validation experiments with salmon fillets. The approach consists of six steps: 1) A digital X-ray image is taken of the fish fillet being tested. 2) The X-ray image is filtered and enhanced to facilitate the detection of fish bones. 3) Potential fish bones in the image are segmented using band pass filtering, thresholding and morphological techniques. 4) Intensity features of the enhanced X-ray image are extracted from small detection windows that are defined in those regions where potential fish bones were segmented. 5) A classifier is used to discriminate between 'bones' and 'no-bones' classes in the detection windows. 6) Finally, fish bones in the X-ray image are isolated using morphological operations applied on the corresponding segments classified as 'bones'. In the experiments we used a high resolution flat panel detector with the capacity to capture up to a 6 million pixel digital X-ray image. In the training phase, we analyzed 20 representative salmon fillets, 7700 detection windows (10x10 pixels) and 279 intensity features. Cross validation yielded a detection performance of 95% using a support vector machine classifier with only 24 selected features. We believe that the proposed approach opens new possibilities in the field of automated visual inspection of salmon and other similar fish.

**Title: View-invariant Fall Detection for Elderly in Real Home Environment**

**Authors: Muhammad Shoab, Ralf Dragon, Jörn Ostermann**

**Abstract:** We propose a novel context based human fall detection mechanism in real home environment. Fall incidents are detected using head and floor information. The centroid location of the head and feet from each frame are used to learn a context model consisting of normal head and floor blocks. Every floor block has its associated Gaussian distribution, representing a set of head blocks. This Gaussian distribution defines standard vertical distance as average height of an object at that specific floor block. The classification of blocks and average height is later used to detect a fall. Fall detection methods often detect bending situations as fall. This method is able to distinguish bending and sitting from falling. Furthermore, a fall into any direction and at any distance from camera can be detected. Evaluation results show the robustness and high accuracy of the proposed approach.

**Title: Error Analysis of Geometric Ellipse Detection Methods Due to Quantization**

**Authors: Dilip Kumar Prasad, Maylor K.H. Leung**

**Abstract:** Many geometric methods have been used extensively for detection of ellipses in images. Though the geometric methods have rigorous mathematical framework, the effect of quantization appears in various forms and introduces errors in the implementation of such models. This unexplored aspect of geometric methods is studied in this paper. We identify the various sources that can affect the accuracy of the geometric methods. Our results show that the choice of points used in geometric methods is a very crucial factor in the accuracy. If the curvature covered by the chosen points is low, then the error may be significantly high. We also show that if numerically computed tangents are used in the geometric methods, the accuracy of the methods is sensitive to the error in the computation of the tangents. Our analysis is used to propose a probability density function for the relative error of the geometric methods. Such distribution can be an important tool for determining practical parameters like the size of bins or clusters in the Hough transform. It can also be used to compare various methods and choose a more suitable method.

**Title: Ellipse Detection by Combining Division and Model Selection Based Integration of Edge Points**

**Authors: Yasuyuki Sugaya**

**Abstract:** We present a new method for detecting ellipses in images. We first detect edges, from which we select an elliptic arc by RANSAC: we randomly choose five points repeatedly until the ellipse they define closely passes through a maximum number of edge points. Those points are regarded as constituting an elliptic arc. We repeat this procedure and detect multiple elliptic arcs. Then, we integrate those arcs that define similar ellipses to one group and fit a single ellipse to them. The similarity between ellipses is measured by a model selection criterion. Thus, our method can accurately detect partially occluded ellipses. Using real images, we demonstrate the effectiveness of our method.

**11:00AM-12:20AM Oral 1 Biometrics**

**Title: Bit Reliability-driven Template Matching in Iris Recognition**

**Authors: Christian Rathgeb, Andreas Uhl**

**Abstract:** Of all the biometric applications available today, it is generally conceded that iris recognition is one of the most accurate. In the past several years a huge amount of iris recognition algorithms have been proposed. However, the vast majority of proposed algorithms restrict to extracting distinct features out of preprocessed iris textures to generate discriminative binary iris-codes, neglecting potential improvements in matching procedures. In this work we present a new technique for matching binary iris-codes. Information of authentication procedures is leveraged by maintaining so-called reliability masks for each user, which indicate local consistency of enrollment templates. Based on user-specific reliability masks a weighted matching procedure is performed in order to improve recognition performance. We apply the proposed matching procedure to different iris recognition algorithms and compare obtained recognition rates to other matching techniques. Experimental results confirm the worthiness of our approach.

**Title: Recombination of 2D and 3D Images for Multimodal 2D + 3D Face Recognition**

**Authors: Gede Putra Kusuma, Chin-Seng Chua, Hock-Lye Toh**

**Abstract:** Exploiting the dependency between 2D and 3D information of a face at the early stage is more effective than at the later stage. Early fusion data contains richer information about the input biometric than the compressed features or matching scores. We propose an image recombination for face recognition that explores the dependency between modalities at the image level. Facial cues from 2D and 3D images are recombined into more discriminating data using PCA- and FLD-based recombination transforms. Our experimental results show that the proposed image recombination outperforms the pixel- and score-level fusions.

**Title: A Novel Facial Expression Recognition Based on the Curvelet Features**

**Authors: Juxiang Zhou, Yunqiong Wang, Tianwei Xu, Wanquan Liu**

**Abstract:** Curvelet transform has been recently proved to be a powerful tool for multi-resolution analysis on images. In this paper we propose a new approach for facial expression recognition based on features extracted via curve let transform. First curve let transform is presented and its advantages in image analysis are described. Then the coefficients of curve let in selected scales and angles are used as features for image analysis. Consequently the Principal Component Analysis (PCA) and Linear Discriminate Analysis (LDA) are used to reduce and optimize the curve let features. Finally we use the nearest neighbor classifier to recognize the facial expressions based on these features. The experimental results on JAFFE and Cohn Kanade two benchmark databases show that the proposed approach outperforms the PCA and LDA techniques on the original image pixel values as well as its counterparts with the wavelet features.

**Title: Age Regression Based on Local Image Features**

**Authors:** Azza Mokadem, Maurice Charbit, Gérard Chollet, Kevin Bailly

**Abstract:** Human age estimation using facial image is becoming more and more investigated because of its potential applications in many areas such as multimedia communication and human computer interaction. Since many factors contribute to the aging process like gender, race, health, living style, the current age estimation performance for computer vision systems is still not efficient enough for practical use. In this paper, we addressed the problem of age estimation from single facial gray-scale image since the color information appeared as not significant in considered low resolution images. Local and global Discrete Cosinus Transformation (DCT) are used for feature extraction allowing thus a first dimensionality reduction through this discriminative representation. A second reduction of dimensionality has been obtained through principal component analysis (PCA). A linear regression function has been learned and tested on different large databases extracted from MORPH. Experimental results have shown some encouraging results.

**Afternoon**

**02:05PM-03:25PM Oral 2 Computer Vision (I)**

**Title: Towards Generalized Benthic Species Recognition and Quantification Using Computer Vision**

**Authors:** Adam F. Gobi

**Abstract:** Seabed resource exploitation and conservation efforts are extending to offshore areas where the distribution of benthic epifauna (animals living on the seafloor) is unknown. There is a need to survey these areas to determine how biodiversity is distributed spatially and to evaluate and monitor ecosystem states. Seafloor imagery, collected by underwater vehicles, offer a means for large-scale characterization of benthic communities. A single submersible dive can image thousands of square metres of seabed using video and digital still cameras. As manual, human-based analysis lacks large-scale feasibility, there is a need to develop efficient and rapid techniques for automatically extracting biological information from this raw imagery. To meet this need, underwater computer vision algorithms are being developed for the automatic recognition and quantification of benthic organisms. Focusing on intelligent analysis of distinct local image features, the work has the potential to overcome the unique challenges associated with visually interpreting benthic communities. The current incarnation of the system is a significant step towards generalized benthic species mapping, and its feature-based nature offers several advantages over existing technology.

**Title: Mismatch Removal for Wide-baseline Image Matching via Coherent Region-to-Region Correspondence**

**Authors:** Chen Wang, Kai-Kuang Ma, Teck-Khim Ng, Dong Guo

**Abstract:** The conventional wide-baseline image matching solely relies on identifying the same local features of two wide-baseline images under matching to establish pixel-to-pixel correspondence based on the nearest neighbor matching criterion. However, a large number of mismatches would be incurred especially for those images containing complicated scene. In order to effectively reduce mismatches, we propose to utilize the information provided by the established coherent region-to-

region correspondence to verify whether each pixel-to-pixel match constructed by the scale invariant feature transform (SIFT) descriptors previously is indeed a correct match or a mismatch. In order to establish coherent one-to-one region correspondence, over-segmentation is first performed on the entire image, and the obtained image segments are merged into larger regions by the proposed segment-merging operations. A bipartite graph model is then applied to these merged regions, and the Hungarian method is exploited to establish one-to-one coherent region pairs. Extensive experimental results have demonstrated that our proposed mismatch removal method for wide-baseline image matching significantly reduces incorrect SIFT-based pixel-to-pixel matching pairs.

**Title: One-shot Entire Shape Acquisition Method Using Multiple Projectors and Cameras**

**Authors:** Ryo Furukawa, Ryusuke Sagawa, Hiroshi Kawasaki, Kazuhiro Sakashita, Yasushi Yagi, Naoki Asada

**Abstract:** In this paper, we propose an active scanning system using multiple projectors and cameras to acquire a dense entire shape of the object with a single scan (a.k.a. ones hot scan). One of the potential application of the system is to capture a moving object with high frame-rate. Since the pattern used for ones hot scan is usually complicated and those patterns interfere each other if they are projected on the same object, it is difficult to use multiple sets of them for entire shape acquisition. In addition, at the end of the closed loop, errors on each scan are accumulated, resulting in large gaps between shapes. To solve the problem, we propose a ones hot shape reconstruction method using a projector projecting a static pattern of parallel lines with one or two colors. Since each projector projects just parallel lines with a small number of colors, those patterns are easily decomposed and detected even if those patterns are projected multiple times on the same object. We also propose a kind of multi-view reconstruction algorithm for the proposed projector-camera system. In the experiment, we actually built a system which consists of six projectors and six cameras and dense shapes of entire objects were successfully reconstructed.

**Title: Shape Classification Using Local and Global Features**

**Authors:** Kart-Leong Lim, Hamed Kiani Galoogahi

**Abstract:** In this paper, we address the shape classification problem by proposing a new integrating approach for shape classification that gains both local and global image representation using Histogram of Oriented Gradient (HOG). In both local and global feature extraction steps, we use PCA to make this method invariant to shapes rotation. Moreover, by using a learning algorithm based on Adaboost we improve the global feature extraction by selecting a small number of more discriminative visual features through a large raw visual features set to increase the classification accuracy. Our local method is adopted from the popular bag of key points approach for shape classification. To integrate the classification results generated based on both local and global features, we use a combining classifier to perform the final classification for a new unknown image query. The experiment results show that this new method achieves the state-of-art accuracy for shape classification on the animal dataset in [8].

**03:40PM-04:40PM Short 2 Computer Vision (II)**

**Title: Modeling of Unbounded Long-Range Drift in Visual Odometry**

**Authors:** Ruyi Jiang, Reinhard Klette, Shigang Wang

**Abstract:** Visual odometry is a new navigation technology using video data. For long-range navigation, an intrinsic problem of visual odometry is the appearance of drift. The drift is caused by error accumulation, as visual odometry is based on relative measurements, and will grow unboundedly with time. The paper first reviews algorithms which adopt various methods to suppress this drift. However, as far as we know, no work has been done to statistically model and analyze the intrinsic properties of this drift. This paper uses an unbounded system model to represent the drift behavior of visual odometry. The model is composed of an unbounded deterministic part with unknown constant parameters, and a first-order Gauss-Markov process. A simple scheme is given to identify the unknown parameters as well as the statistics of the stochastic part from experimental data. Experiments and discussions are also provided.

**Title: Human Behaviour Analysis and Event Recognition at a Point of Sale**

**Authors:** R. Sicre, H. Nicolas

**Abstract:** This paper presents a new application that aims at improving communication and interactions between digital media and customers at a point of sale. Our system analyzes in real-time human behaviour while shopping. In particular, the system detects customer's interest in products and interaction such as people grabbing products. This system is based on a behaviour model. A video analysis module detects motion, tracks moving object, and describes local motion. Then specific behaviours are recognized and sentences are generated. Finally, our approach is tested on real video sequences.

**Title: A Robust Method for License Plate Detection**

**Authors:** Jian Luo, Su Yang, Ruimin Guan, Haijun Niu

**Abstract:** License plate detection plays an important role in vehicle license plate recognition for intelligent transport systems. This paper presents a robust method for license plate detection. As we observed, license plate area contains rich corner points' information. The whole procedure is as follows: (1) Extract the interesting points using Harris corner detector in enhanced car image. (2) Cluster the corner points and apply outlier detection to every cluster. (3) Merge the clusters close to each other and use the gradient information to find the proper block in which the license plate area is enclosed. (4) Train a simple cascade classifier to classify the blocks into two categories: The area possibly containing license plate and those not. (5) Detect the plate region by the rectangle window. Experimental results demonstrate the robustness and the generality of our method.

**Title: Object Tracking Based on Covariance Descriptors and On-Line Naive Bayes Nearest Neighbor Classifier**

**Authors:** Pedro Cortez Cargill, Domingo Mery Quiroz, Luis Enrique Suar

**Abstract:** Object tracking in video sequences has been extensively studied in computer vision. Although promising results have been achieved, often the proposed solutions are tailored for particular objects, structured to specific conditions or constrained by tight guidelines. In real cases it is difficult to recognize these situations automatically because a large number of parameters must be tuned. Factors such as these make it necessary to develop a method robust to various environments, situations and occlusions. This paper proposes a new simple appearance model, with only one parameter, which is robust to prolonged partial occlusions and drastic

appearance changes. The proposed strategy is based on covariance descriptors (which represent the tracked object) and an on-line nearest neighbor classifier (to track the object in the sequence). The proposed method performs exceptionally well and reduces the average error (in pixels) by 47% compared with tracking methods based on on-line boosting.

**Title: An Online Learning Method for Shadow Detection**

**Authors:** Chung-Hsien Huang, Ruei-Cheng Wu

**Abstract:** Shadow detection is a critical issue for most applications of video surveillance. In this study, we present an object-wise online learning method to detect casting shadows without providing any priori scene information or threshold parameters. Hue, saturation, and intensity- difference histograms of moving objects are collected to learn a cumulative distribution separately. The accumulating strategy strengthens the impact of shadow parts but reduces the effects of non-shadow parts. In each cumulative distribution, the most significant peak is then fit as a Gaussian function by using a robust estimation method. The fitted Gaussian is treated as a shadow likelihood function. The integration of shadow likelihoods in hue, saturation and intensity are modeled as a data term into a Graph Cut model. The Graph Cut model also incorporates edge information of the current image as a spatial smoothing term. Therefore, the shadow pixels can thus be labeled by minimizing an energy function. Compared to a supervised thresholding method, experimental results reveal that the flexibility and adaptability of the proposed learning method on real surveillance scenarios.

**Title: Cart Auditor: A Compliance and Training Tool for Cashiers at Checkout**

**Authors:** Unsang Park, Charles A. Otto, Sharath Pankanti

**Abstract:** Shopping carts have traditionally been used as a tool provided to the customers in retail stores to carry items from the shelf to checkout stations. These days shopping carts can also be used as a security checkpoint to prevent store losses. All the items collected in a shopping cart are supposed to be unloaded at the checkout station to be scanned and included in the bill. Any items left in the cart intentionally or by accident will not be charged and therefore cause a loss to the store. We propose a system that automatically detects shopping carts and verify their emptiness at the checkout station. We use motion segmentation, line detection, and template matching methods for the cart detection and emptiness verification. An inter-frame edge difference, cart's path accumulator, and a finite state model are introduced for accurate cart detection. All detected carts are compared with empty cart models and the dissimilarity scores are calculated to verify the emptiness. The proposed system was evaluated on a long video clip (~12 hours) and showed promising results both in cart detection and emptiness verification.

**Title: A Level-set Based Tracking Approach for Surveillance Video with Fusion and Occlusion**

**Authors:** Chaoqun Hong, Na Li, Mingli Song, Jiajun Bu, Chun Chen

**Abstract:** Traditional level-set-based methods of tracking contours suffered from occlusion and fusion. In this paper, the proposed method introduces dynamic inci-

dent detection to find and handle occlusion and fusion. Color histogram of the hue component in HSV color space is used to identify the objects re-entering after occlusion. On the other hand, object features including the size and the motion pattern are utilized to remove the fake regions that are fused with the object region. Besides, a comprehensive foreground extraction (CFE) method based on the combination of background subtraction and Local Binary Pattern (LBP) is proposed. It is fast and robust. Our method is Experiments show that the proposed approach outperforms previous methods on both speed and subjective quality.

**Title: Camera Calibration Using Lengths of Corresponding Line Segments**  
*Authors: Ruiyan Wang, Guang Jiang, Long Quan, Chengke Wu*

**Abstract:** Camera calibration using patterns is widely used in computer vision and industry. The accuracy of calibration depends on the accuracy of the pattern. A high accuracy pattern is usually difficult to manufacture in labs. In this paper, we propose to use two images of two identical industrial objects as natural calibration patterns. The corresponding line segments on the objects are used for a nonlinear calibration method, and only five intrinsic parameters of the camera are optimized in the procedure. The experiments show that the proposed method is both effective and practical.

**Title: A Linear-time Algorithm for the Generation of Random Digital Curves**  
*Authors: Partha Bhowmick, Oishee Pal, Reinhard Klette*

**Abstract:** We propose an algorithm to generate random digital curves of finite length, generating points of a digital path  $\rho$  on the fly. Path  $\rho$  never intersects or touches itself, and hence becomes simple and irreducible. This is ensured by detecting every possible trap formed by the previously generated part of  $\rho$ , which, if entered into, cannot be exited without touching or intersecting  $\rho$ . The algorithm is completely free of any backtracking and its time complexity is linear in the length of  $\rho$ . Implemented and tested exhaustively, it shows that it produces results as specified by the user.

**Title: Shape Matching by Region Grouping**

*Authors: Yan Gao, Wei-Yun Yau*

**Abstract:** In this paper, we propose a novel region grouping approach to shape matching. It is proposed as an alternative region based approach to the traditional edge based shape matching using distance transforms. It has the advantage of obtaining a higher detection rate and obtaining meaningful object segmentation simultaneously. Each image is first segmented into image regions, and possible matches are found among the regions based on a proposed probabilistic similarity measure to the exemplars. We evaluate this new approach against edge based matching using Chamfer and Hausdorff distances. Pros and cons of the proposed method are discussed.

04:40PM-06:00PM SP 1 Social Media Content Mining

**Title: Travel Video Scene Detection by Search**

*Authors: Wei-Ta Chu, Cheng-Jung Li, Tsung-Che Lin*

**Abstract:** We propose an approach to conduct video scene detection especially for travel videos captured by amateur photographers in journeys. The correlation between a travel video and its corresponding text-based travel schedule is discovered. Because scene boundaries are clearly defined in schedules, we segment videos into scenes by checking the discovered cross-media correlation. To make these two modalities comparable, photos related to the visited scenic spots are retrieved from image search engines, by the keywords extracted from text-based schedules. Sequences of key frames and retrieved photos are represented as visual word histograms, and the problem of correlation determination is then transformed as an approximate sequence matching problem. The experimental results verify the effectiveness of the proposed idea, and show the promising research direction of utilizing cross-media correlation in media analysis.

**Title: Multi-Modality Transfer Based on Multi-Graph Optimization for Domain Adaptive Video Concept Annotation**

*Authors: Shaoxi Xu, Sheng Tang, Yongdong Zhang, Jintao Li*

**Abstract:** Multi-modality, the unique and important property of video data, is typically ignored in existing video adaptation processes. To solve this problem, we propose a novel approach, named multi-modality transfer based on multi-graph optimization (MMT-MGO) in this paper, which leverages multi-modality knowledge generalized by auxiliary classifiers in the source domain to assist multi-graph optimization (a graph-based semi-supervised learning method) in the target domain for video concept annotation. To our best knowledge, it is the first time to introduce multi-modality transfer into domain adaptive video concept detection and annotation. Moreover, we propose an efficient incremental extension scheme to sequentially estimate a small batch of new emerging data without modifying the structure of multi-graph scheme. The proposed scheme can achieve a comparable accuracy with that of the brand-new round optimization which combines these data with the data corpus for the nearest round optimization, while the time for estimation has been greatly reduced. Extensive experiments over TRECVID2005 and 2007 data sets demonstrate the effectiveness of both the multi-modality transfer scheme and the incremental extension scheme.

**Title: A Collaborative Approach for Image Annotation**

*Authors: Fuming Sun, Yong Ge, Dongxia Wang, Xueming Wang*

**Abstract:** Automatic image annotation is a promising solution to enable more effective image retrieval by keywords. Different statistical models and machine learning methods have been introduced for image auto-annotation. In this paper, we propose a collaborative approach, in which multiple different statistical models are combined effectively to predict the annotation for each image. Moreover, we combine both low-level feature of image and semantic information naturally. In addition, we also combine the correspondence between keywords and image visual tokens/regions, and the word-to-word correlation to enhance the annotation. We employ the conditional probability to express two kinds of correlation uniformly and obtain the correspondence between keyword and visual feature with two typical statistical models. Ex-

periments conducted on standard Corel dataset demonstrate the effectiveness of the proposed method for image automatic annotation.

**Title: Mash-Up Approach for Web Video Category Recommendation**

**Authors:** Yi-Cheng Song, Haojie Li

**Abstract:** With the advent of web 2.0, billions of videos are now freely available online. Meanwhile, rich user generated information for these videos such as tags and online encyclopedia offer us a chance to enhance the existing video analysis technologies. In this paper, we propose a mash-up framework to realize video category recommendation by leveraging web information from different sources. Under this framework, we build a web video dataset from the YouTube API, and construct a concept collection for web video category recommendation (CCVV-CR) from this dataset, which consists of the web video concepts with small semantic gap and high categorization distinguish ability. Besides, Wikipedia Propagation is proposed to optimize the video similarity measurement. The experiments on the large-scale dataset with 80,031 web videos demonstrate that: (1) the mash-up category recommendation framework has a great improvement than the existing state-of-art methods. (2) CCVV-CR is an efficient feature space for video category recommendation. (3) Wikipedia Propagation could boost the performance of video category recommendation.

## **16-Nov-2010**

### **Morning**

#### **10:00AM-11:00AM Short Oral 3 Image/Video Coding**

**Title: Region Adaptive Inter-Color Prediction Approach to RGB 4:4:4 Intra Coding**

**Authors:** Lili Zhao, Mingjing Ai

**Abstract:** RGB color signals encoding is more desirable for high resolution and accurate definition for saturations and hue. This paper presents a region adaptive inter-color prediction scheme for 4:4:4 RGB signals, which treats G component as a primary component and predicts B/R component from G component with different weighting factors based on different texture regions. The proposed algorithm demonstrates more efficient coding performance both than the H.264/AVC intra coding algorithm and than the block adaptive prediction algorithm. Experimental results show that, compared with the H.264/AVC intra coding algorithm, the proposed algorithm can produce a bit-rate saving of up to 26-30% with PSNR gain of 0.3-0.5dB for some high resolution video sequences and low resolution static images. Meanwhile, the complexity of the proposed algorithm is much less than that of H.264/AVC intra prediction technique in independent mode.

**Title: An Extension of CCSDS Algorithm by ROI Capability**

**Authors:** Shen-Chuan Tai, Tse-Ming Kuo, Ng Keng San

**Abstract:** In the modern compression system, besides high compression qualities and low bit rates, others extra functional capabilities become more and more important. For the use of satellite images, the data size is always very large and the users might be just interested in some regions of a satellite image. Thus it is an ideal way to keep the better quality in the regions of interest than in others in a satellite image. In JPEG2000 standard part 1, the region of interest (ROI) is implemented by MAX-SHIFT, which needs to generate an ROI mask with extra memory space. For the computers in a satellite, lower computing complexity and fewer memory are required. Thus the Consultative Committee for Space Data Systems Recommendation (CCSDS) proposes a specification of image data compression for satellite images. An ROI extension based on CCSDS is proposed in our study. Experimental results show that it saves more compression time and moreover it is competitive with the JPEG2000 standard without the memory used for masks.

**Title: An Object-based Full Frame Concealment Strategy for H.264/AVC Using True Motion Estimation**

**Authors:** Shen-Chuan Tai, Chien-Shiang Hong, Cheng-An Fu

**Abstract:** In packet-based video transmissions, packets loss may result in the loss of the whole video frame due to channel errors. A lot of error concealment algorithms have been proposed to overcome the channel errors, but most of them only can deal with the loss of macro blocks but whole frame. In this thesis, we propose an efficient full frame recovery algorithm which is based on object-based module for a new coding standard H.264/AVC. First, our proposed method estimates the true motion of each block in the reference frame. Besides, the reference frame will be divided into several objects according to the estimated motion vectors. Furthermore, each object will be extrapolated to its corresponding place in the missing frame. Finally, the experimental results show that our algorithm could provide a more efficient solution to recover a missing whole frame than previous methods. In addition, based on the object-based module, our proposed method provides the better visual quality than other comparative algorithms.

**Title: A Perceptually Lossless Image Compression Scheme Based on JND Refinement by Neural Network**

**Authors:** Wen-Nung Lie, Wen-Ching Liu

**Abstract:** In this paper, we propose a JND (Just Notice Distortion)-loss less image compression scheme that can improve the compression performance for JPEG-LS, while maintaining the image perceptual quality simultaneously. JND-loss less can be easily achieved by setting the quantization step size (QSS) to be double the JND value. However, dynamic JNDs make the coding of varying QSSs difficult and the JND estimated at decoder is often inaccurate due to incomplete background information therein. Here, we propose to use a neural classifier at both the encoder and decoder to correct the JND mismatch and make the coding of varying QSSs unnecessary. Experiments show that our proposed JND-refinement scheme is capable of increasing the compression performance by up to 15% (w.r.t. JPEG-LS) and the proposed neural classifier is capable of correcting 94% of JND mismatch.

**Title: High-Level Synthesis of a Unified 2-D DWT System Architecture for**

#### **JPEG2000 Using FPGAs**

*Authors: Ishmael Sameen, Yoong Choon Chang, Mow Song Ng, Bok-Min Goi, Chee Siong Lee*

*Abstract:* In this paper, a programmable 2-D DWT system architecture designed for the JPEG2000 standard is proposed. The proposed system architecture, developed through an iterative design space exploration methodology using Altera's C2H compiler, provides a significant 2-D DWT performance improvement when compared to an optimized 2-D DWT software implementation and is capable of real-time video processing performance for high resolution grayscale images up to 1280 x 720 (720p) when synthesized and benchmarked in an Altera DE3 Stratix III FPGA board.

#### **Title: A Perceptually Scalable and JPEG Compression Tolerant Image Encryption Scheme**

*Authors: Fawad Ahmed, M Y Siyal, Vali Uddin Abbas*

*Abstract:* In this paper, a perceptually scalable and JPEG compression tolerant DCT-based image encryption scheme is proposed. There are two main features of the proposed scheme, first the decryption algorithm is able to reconstruct the plaintext-image even if the encrypted image is JPEG compressed. Secondly, the encryption algorithm can be adjusted to produce cipher-image with varying perceptual distortion. The proposed encryption scheme relies on some very interesting properties of orthogonal matrices containing columns that form a set of orthonormal basis vectors. It is envisaged that the proposed cryptosystem would enable uncompressed encrypted image files to be transmitted or stored in the JPEG compressed domain thus saving bandwidth and or storage space. Further, the proposed scheme can be used either for low-level encryption in which the cipher-image has a degraded visual quality but reflects the contents of the plaintext-image, or high-level encryption in which the cipher-image does not reveal any information about the plaintext-image. Experimental results and security analysis is presented to show the viability of the proposed scheme.

#### **Title: N-Way Video Communication over Enterprise Networks Based on Adaptive Bit Stream Extraction in Scalable Video Coding**

*Authors: Ramanathan Palaniappan, Nikil Jayant*

*Abstract:* In this paper, we explore real time N-way video communication over enterprise networks based on scalable video coding, SVC (Scalable extension of H.264/AVC). We present a bit stream extraction strategy based on GOP size prediction and perceptual importance of temporal and quality enhancements, and validate the strategy through initial subjective testing. We also compare it with extraction based on average bit rate of each constituent layer of the bit stream. Our extraction is adaptive with respect to variations in available bandwidth and is proxy driven, in that the decision process and the adaptation are preformed at proxy servers located at the edges of the backbone network. The main goals are to 1) maximize the user perceived quality even during deteriorating channel conditions, 2) maximize the reaction speed to changes in available bandwidth and 3) minimize the extraction delay. We report objective and subjective results for the N=2 case, with HD video clips encoded at 600-900 kbps. With a channel utilization of 92.7%, our extraction algorithm based on GOP size prediction shows an average increase in PSNR of about 2.2 dB over the extraction based on average bit rate. Initial subjective tests also prove that our layer

extraction strategy is perceptually more efficient than other extraction schemes.

#### **11:00AM-12:20AM Oral 3 Video Coding**

#### **Title: Zero-block Mode Decision Algorithm for High Bit-Rate Coding in H.264/AVC**

*Authors: Yu-Ming Lee, Shen-Hsien Liu, Wei-Yao Chiu, Yinyi Lin*

*Abstract:* This paper proposes another zero-block inter mode decision algorithm for high bit-rate coding in H.264/AVC, to improve computation efficiency of zero-block decision [2]. In the proposed algorithm we suggest using zero-blocks of both 8X8 and 4X4 zero-blocks to describe video sequences in detail, including global and local stationary characteristics. The experimental results reveal that significant computation improvement (up to 20% reduction) for high bit-rate coding can be achieved over the original zero-block decision algorithm.

#### **Title: Low Complexity Hierarchical Prediction Algorithm for H.264/SVC**

*Authors: Abdelrahman Abdelazim, Stephen Mein, Martin Varley, Djamel Ait-Boudaoud*

*Abstract:* In the scalable video coding extension of the H.264/AVC standard, an exhaustive search technique is used to select the best coding mode for each macroblock. This technique achieves the highest possible coding efficiency, but it demands a higher video encoding computational complexity which constrains its use in many practical applications. This paper proposes combined fast sub-pixel motion estimation and a fast mode decision algorithm for inter-frame coding for temporal, spatial, and coarse grain signal-to-noise ratio scalability. It makes use of correlation between the macroblock and its enclosed partitions at different layers. Experimental results show that the scheme reduces the computational complexity significantly with negligible coding loss and bit-rate increases when compared to JSVM 9.15 and recently reported fast mode decision algorithms.

#### **Title: An Improved Detection Method for Zero Quantized Blocks on H.264/AVC**

*Authors: Bo-Jhih Chen, Shen-Chuan Tai*

*Abstract:* An improved detection method for observing the zero quantized block (ZQB) is proposed. The additional computational cost would be reduced due to 4X4 ZQB's being detected prior to the forward transform and the quantization processes. We report a new criterion based on the statistical analysis by considering the energy conservation theorem. Experiments are also carried out to validate the present method. The results indicate that the present method has both a better detection rate with the negligible PSNR degradation and a reasonable error and/or false detection comparing to the prevalent methods. Particularly, computation savings are obtained as well.

#### **Title: Robust Motion-Compensated Orthogonal Video Coding Using EB-COT**

*Authors: Ousmane Barry, Du Liu, Stefan Richter, Markus Flierl*

*Abstract:* This paper proposes a rate-distortion control for motion-compensated orthogonal video coding schemes and evaluates its robustness to packet loss as faced in, e.g., IP networks. The robustness of standard hybrid video coding is extensively studied in the literature. In contrast, motion-compensated orthogonal sub bands offer important advantages and new features for robust video transmission. In this work, we utilize so-called unidirectional motion-compensated orthogonal transforms in combination with entropy coding similar to EBCOT known from JPEG2000. The approach provides a flexible embedded structure and allows flexible rate-distortion optimization. Moreover, it may even permit separate encoding and rate control. The proposed rate-distortion control takes channel coding into account and obtains a preemptively protected representation. Our implementation is based on repetition codes, adapted to the channel condition, and improves the PSNR significantly. The optimization requires an estimate of the packet loss rate at the encoder and shows moderate sensitivity to estimation errors.

## Afternoon

**01:20PM-02:20PM    Short Oral 4    Image/Video Coding**

*Title: Group Segmentation During Object Tracking Using Optical Flow Discontinuities*

*Authors: Simon Denman, Clinton Fookes, Sridha Sridharan*

*Abstract:* Within a surveillance video, occlusions are commonplace, and accurately resolving these occlusions is key when seeking to accurately track objects. The challenge of accurately segmenting objects is further complicated by the fact that within many real-world surveillance environments, the objects appear very similar. For example, footage of pedestrians in a city environment will consist of many people wearing dark suits. In this paper, we propose a novel technique to segment groups and resolve occlusions using optical flow discontinuities. We demonstrate that the ratio of continuous to discontinuous pixels within a region can be used to locate the overlapping edges, and incorporate this into an object tracking framework. Results on a portion of the ETISEO database show that the proposed algorithm results in improved tracking performance overall, and improved tracking within occlusions.

*Title: Modeling and Detection of Blurred Illumination Edges*

*Authors: Cheng Wen, Chih-Hung Huang, Kuang-Chiung Chang*

*Abstract:* Luminance edges are classified into reflectance edges and illumination edges. Most of the edge detection papers published focus on reflectance edges detections, whereas only few of them deal with illumination edges. In this paper, a new approach to detect illumination edges in grayscale images is proposed. Differing from the conventional illumination edge detection techniques, which were mostly based on the grayscale intensity changes in the image being processed, instead, an error-function-based formulation is introduced to model blurred illumination edges, using a factor called the illumination ratio defined in this paper. Then, the profile of the illumination ratio is fitted with the proposed error-function model to determine the

locations of all illumination edges. Although the frame-differencing method using complementary patterns is generally acknowledged to be the most accurate illumination edge detector, the proposed technique could perform better than it in both speed and accuracy. This technique benefits the fields of visual detection, machine vision, and especially the coded structured-light-based 3D scanning systems.

*Title: High-quality Non-blind Image Deconvolution*

*Authors: Jong-Ho Lee, Yo-Sung Ho*

*Abstract:* The most annoying artifacts in image deconvolution are ringing and amplified noise. These artifacts can be reduced significantly by regularization using the Maximum a Posteriori (MAP) method that exploits not only the likelihood but also the image prior in image deconvolution. Although ringing and noise can be reduced significantly with strong regularization, image details are also reduced, so the deconvolved image is neither sharp nor clear. In this paper, we propose a non-blind image deconvolution method with adaptive regularization that can reduce ringing and noise more noticeable in a smooth region and preserve image details in a textured region. For adaptive regularization, after we make a quick estimate of the reference image that can indicate the strength of regularization, we perform regularization adaptively according to the local characteristics. Experimental results show that ringing and noise are suppressed significantly, while preserving image details effectively.

*Title: SAR Image Speckle Noise Suppression Based on DFB Hidden Markov Models Using Immune Clonal Selection Thresholding*

*Authors: Haiyan Jin, Xueming Sun*

*Abstract:* Synthetic aperture radar (SAR) images are inherently affected by multiplicative speckle noise, which is due to the coherent nature of the scattering phenomenon. This paper proposes a novel DFB-based algorithm with hidden Markov modeling, which reduces speckle in SAR images while preserving the structural features and textural information of the scene, and introduces evolutionary computation theory - immune clonal selection (ICS) method to optimize threshold avoiding the drawback of experiential threshold. We compare our proposed method to wavelets techniques applied on real SAR imagery and we quantify the achieved performance improvement.

*Title: Hierarchal Predefined Codebook for High-Frequency Component Augmentation Image Enlargement*

*Authors: Hakaru Tamukoh, Hideaki Kawano, Noriaki Suetake, Masatoshi Sekine, Byungki Cha, Takashi Aso*

*Abstract:* A high-quality image enlargement is achieved by an estimation of high-frequency component. A codebook based method, which consists of a relationship of high and low-frequency image components, is known as one of the high-quality image enlargement. However, it takes too much calculation time, and is not appropriate for a real-time application such as a moving image enlargement. In this paper, we propose a hierarchal predefined codebook which achieves both of a fast calculation and a high-quality image enlargement. First, we define 1,904 code words generated from fundamental edge patterns. Then, the vector quantization is applied to the generated code words, and dividing 1,904 code words into 225 clusters. Finally, we obtain the hierarchal predefined codebook based on the divided code words. Experimental results show that the proposed method achieved both of a fast calculation

and a high-quality image enlargement.

**Title: Blurred Image Registration by Combined Invariant of Legendre Moment and Harris-Laplace Detector**

**Authors:** Xiubin Dai, Hui Zhang, Huazhong Shu, Limin Luo, Tianliang Liu  
**Abstract:** Since real imaging systems are imperfect, the acquired images are always distorted by blur, scale and rotation transformation. Then the registration of these degraded images has become an important task in many applications in which the moment invariants are usually efficient tools. However, the existing methods can only deal with the slightly distorted images and have the problems of information redundancy. Besides, some methods have overlapping constraint that the images to be aligned should be fully included into the reference images. In this paper, we proposed a novel method in which a new set of combined invariants based on Legendre moment holding for blur, rotation and scale degradation simultaneously were constructed as feature descriptors, and scale-invariant Harris-Laplace detector was applied to exact feature points. The experimental results show that our method can work well without overlapping constraint, especially when the distortion is great.

**Title: A Hierarchical Multi-classifier Framework for Landform Segmentation Using Multi-spectral Satellite Images - A Case Study over the Indian Subcontinent**

**Authors:** Uthara Gosa Mangai, Suranjana Samanta, Sukhendu Das, Pinaki Roy Chowdhury, Koshy Varghese, Manisha Kalra  
**Abstract:** There is an increasing need for automatically segmenting the regions of different landforms from a multispectral satellite image. The problem of Landform classification using data only from a 3-band optical sensor (IRS-series), in the absence of DEM (Digital Elevation Model) data, is complex due to overlapping and confusing spectral reflectance from several different landform classes. We propose a hierarchical method for landform classification for identifying a wide variety of landforms occurring over parts of the Indian subcontinent. At the first stage, the image is classified into one of three broad categories: Desertic, Coastal or Fluvial, using decision fusion of three SVMs (Support Vector Machine). In the second stage, the image is then segmented into different regions of landforms, specifically belonging to the class (category) identified at stage 1. To show the improvement in accuracy of our classification method, the results are compared with two other methods of classification.

**Title: Fast, High-Quality Line Antialiasing by Prefiltering with an Optimal Cubic Polynomial**

**Authors:** Jan Wassenberg  
**Abstract:** Scan-converting lines for raster-based displays is a basic building block of many computer graphics tasks. While many fast line drawing algorithms are known, most produce thin and 'jagged' lines due to aliasing. Wu's algorithm includes a crude approximation of antialiasing, which still includes noticeable step edges. Even hardware multisampling cannot entirely eliminate aliasing. Instead, the proper solution is to remove high-frequency components by pre-filtering the lines. We improve upon previous ad-hoc filters by deriving the optimal (in the sense of minimizing aliasing)

cubic polynomial filter. When combined with our new, optimized variant of the Gupta-Sproull line drawing algorithm, this outperforms Wu's fast approximation while delivering much higher-quality results.

**Title: Adaptive Weighting Scheme for Edge-based Line Interpolation**

**Authors:** Sang-Jun Park, Changki Min, Jechang Jeong, Gwanggil Jeon  
**Abstract:** The adaptive weighting based on the intensity variations of the neighbor pixels and the distance between the neighbor pixels and a pixel to be interpolated is proposed for the edge-based deinterlacing method. The local edge direction is first determined by the modified edge-based line averaging (MELA) method. Then, according to the local edge direction, the inverse distance and intensity variation based weighting is applied to the fourth-order linear interpolation. The experimental results show that the proposed method provides an improvement over the other existing deinterlacing methods including the MELA method. Also, the proposed method is proved to be suitable for the real-time applications due to its short processing time.

**Title: Fast and Enhanced Algorithm for Exemplar Based Image Inpainting**

**Authors:** Anupam, Pulkit Goyal, Sapan Diwakar  
**Abstract:** Image inpainting is the art of filling in missing data in an image. The purpose of inpainting is to reconstruct missing regions in a visually plausible manner so that it seems reasonable to the human eye. There have been several approaches proposed for the same. In this paper, we present an algorithm that improves and extends a previously proposed algorithm and provides faster in painting. Using our approach, one can in paint large regions (e.g. remove an object etc.) as well as recover small portions (e.g. restore a photograph by removing cracks etc.). The inpainting is based on the exemplar based approach. The basic idea behind this approach is to find examples (i.e. patches) from the image and replace the lost data with it. This technique can be used in restoring old photographs or damaged film. It can also remove superimposed text like dates, subtitles etc., or even entire objects from the image like microphones or wires to produce special effects. We obtained good quality results quickly using our approach.

**Title: DWT and TSQV-based Semi-fragile Watermarking Scheme for Tampering Detection in Remote Sensing Images**

**Authors:** Jordi Serra-Ruiz, David Megias  
**Abstract:** A semi fragile watermarking scheme for remote sensing images is presented. The suggested scheme uses the signature of the multispectral or hyper spectral image to embed the mark and detects a forgery of the marked image, e.g. a tampered region. The original image is segmented in three-dimensional blocks and, for each block, a discrete Wavelet transform (DWT) and a tree structured vector quantizer is built. These trees are manipulated using an iterative algorithm until the resulting image satisfies some selected conditions. Each tree is partially modified accordingly to a secret key in order to avoid tampering attacks. This key determines the internal structure of the tree and, also, the resulting distortion. In order to make the marked image robust against near-loss less compression, the trees are built using only the LL sub-band of the DWT. The results show that the method works correctly with remote sensing images and detects copy-and-replace attacks from

segments of the same image, whilst allowing for JPEG2000 compression.

**Title: Color to Gray: Attention Preservation**

**Authors:** Yezhou Yang, Mingli Song, Jiajun Bu, Chun Chen, Cheng Jin

**Abstract:** In this paper, we propose an approach to preserve a crucial visual cue in color to grayscale transformation: attention. The main contributions are three folds: 1) preserving visual attention is more biological plausible than preserving other low level cues, which makes our method more reasonable in theory from both biological and psychological aspects, 2) We treat the saliency map from visual attention analysis as a classifier and aim to preserve attention area in the output grayscale image. 3) A simple minimizing function toward this specific problem is established and can be easily solved. Experimental results on test images indicate that our method is both practical and effective in preserving visual attention from the original color image to corresponding grayscale image.

#### 02:20PM-03:40PM Oral 4 Object Detection and Recognition

**Title: Cascade of Complementary Features for Fast and Accurate Pedestrian Detection**

**Authors:** Alaa Leithy, Mohamed N. Moustafa, Ayman Wahba

**Abstract:** We propose a cascade of two complementary features classifiers to detect pedestrians from static images quickly and accurately. Co-occurrence Histograms of Oriented Gradients (CoHOG) descriptors have a strong classification capability but are extremely high dimensional. On the other hand, Haar-like features are computationally efficient but not highly discriminative for extremely varying texture and shape information such as pedestrians with different clothing and stances. Therefore, the combination of both features enables fast and accurate pedestrian detection. Our framework comprises a cascade of Haar based weak classifiers followed by a CoHOG-SVM classifier. Additionally, we propose reducing CoHOG descriptor dimensionality using Independent Component Analysis (ICA). The experimental results on the Daimler Chrysler and INRIA benchmark datasets show that we can reach very close accuracy to the most accurate CoHOG-only classifier but in less than 1/1000 of its computational cost.

**Title: Object Recognition in 3D Scenes with Occlusions and Clutter by Hough Voting**

**Authors:** Federico Tombari, Luigi Di Stefano

**Abstract:** In this work we propose a novel Hough voting approach for the detection of free-form shapes in a 3D space, to be used for object recognition tasks in 3D scenes with a significant degree of occlusion and clutter. The proposed method relies on matching 3D features to accumulate evidence of the presence of the objects being sought in a 3D Hough space. We validate our proposal by presenting a quantitative experimental comparison with state-of-the-art methods as well as by showing how our method enables 3D object recognition from real-time stereo data.

**Title: Object Extraction Using Novel Region Merging and Multidimensional Features**

**Authors:** Tanima Dutta, Debi Prasad Dogra, Biswapati Jana

**Abstract:** Understanding natural images is a difficult task. One method to accomplish that can be, first, segment the image into regions of similar characteristics and then apply some object extraction scheme. Alternatively, extraction of characteristics of the desired objects can be initiated at the beginning. In this paper, we propose a scheme that adopts the former approach. An image is first segmented and then a novel region processing algorithm is applied. It is found that the proposed algorithm is capable of removing a high percentage of insignificant regions. Following region processing, a multidimensional feature extraction scheme is used. A set of five primitive and non-primitive features are used to create the feature vectors. The image dataset is divided into two parts, i.e. training and test sets. Results show the effectiveness of the proposed feature vector in extracting known objects present in natural and specific domain images.

**Title: Head Pose Classification from Low Resolution Images Using Pair-wise Non-Local Intensity and Color Differences**

**Authors:** Teera Sinteerakul, Daisuke Sugimura, Yoichi Sato

**Abstract:** In this work, we propose novel image descriptors for identifying head poses in low resolution images. The key novelty of our method is to exploit two types of non-local metric for estimating head poses: non-local intensity difference feature (iDF) and non-local color difference feature (cDF). Unlike the existing methods that one pixel can only represent one head pose information, our proposed features are designed to capture geometry of head pose image via relative information of two-randomly picked pixels. The iDF is designed to capture relative head image regions represented by the two pixels (e.g. one pixel represent hair while the other represent skin ) without explicitly labeling any of the pixels. On the other hand, the cDF is designed to capture information about whether or not the two randomly-selected pixels belong to the same head image regions, again, without explicitly labeling any of the regions. Our experimental results demonstrate that our descriptors using pair wise differences in intensity and color outperform current state-of-the-art for head pose estimation from extremely low-resolution images.

#### 03:40PM-03:55PM SP 2 Multimedia Browsing and Understanding

**Title: Attention Based Album Slideshow**

**Authors:** Huiying Liu, Qingming Huang, Shuqiang Jiang

**Abstract:** Slideshow is a popular way to display a digital album automatically. However, it is time consuming and boring when the album is large. Actually, while browsing an album manually, an audience may spend different length of time on different photos. Considering this, in this paper, we propose an approach to estimate the attention amount attracted by each photo and further predict its duration. Three types of attention related factors are considered. 1) The subjective factor, described through photo content. 2) The objective factor, which means the photographic quality of a photo. 3) The album context, represented through a photo's novelty in the album. Then Support Vector Regression is employed to estimate the duration. The

proposed slideshow method displays the photos according to the estimated duration. Both objective and subjective evaluations show the effectiveness of the proposed method.

**Title: Compact Codebook Generation Towards Scale-Invariance**

**Authors:** Si Liu, Shuicheng Yan, Changsheng Xu, Hanqing Lu

**Abstract:** In this paper, we present a novel visual codebook learning approach towards compactness and scale-invariance for dense patch image encoding. Firstly, each image is described as a bag of orderless gridding local patches, each of which is expressed in three scales. Then a unified objective function is proposed to simultaneously enforce the codebook compactness and select the optimal scale for each local patch, and a convergency provable iterative procedure is utilized for optimization. A direct advantage of the new codebook is that each local patch is essentially described by its best scale, and thus shares certain characteristic of SIFT yet not constrained to any salient point detectors. The experiments on PASCAL 07 dataset validate the effectiveness and efficiency of our proposed method for image classification task.

**Title: Video Concept Detection Using Support Vector Machine with Augmented Features**

**Authors:** Xinxing Xu, Dong Xu, Ivor W. Tsang

**Abstract:** In this paper, we present a direct application of Support Vector Machine with Augmented Features (AFSVM) for video concept detection. For each visual concept, we learn an adapted classifier by leveraging the pre-learned SVM classifiers of other concepts. The solution of AFSVM is to re-train the SVM classifier using augmented feature, which concatenates the original feature vector with the decision value vector obtained from the pre-learned SVM classifiers in the Reproducing Kernel Hilbert Space (RKHS). The experiments on the challenging TRECVID 2005 dataset demonstrate the effectiveness of AFSVM for video concept detection.

**Title: Automatic Image Annotation Using Multi-object Identification**

**Authors:** Yin-Fu Huang, Hsin-Yun Lu

**Abstract:** Due to the prevalence of digital cameras, it is easy to retrieve digital images from the Internet. With the rapid development of digital image processing, databases, and Internet technologies, how to efficiently manage a large amount of digital images is very important. In this paper, we proposed a novel approach for automatic image annotation. We extract color, texture, and shape features from a set of training images to build the main object classifier and background object models by using Support Vector Machine (SVM). We apply JSEG to segment background objects out of images, and then extract the feature vectors from the segmented objects for identification. In order to prevent over-segmenting the main object, the combination of Active Contour Model and JSEG is proposed to improve the system performance. Since the images in the same class have background consistency, we exploit Gaussian mixture model (GMM) to explore the relationship between image classes and image backgrounds, and build the association knowledge base. After classifying test images, we only need to compare the backgrounds with the related models for classification. Finally, the experimental results show that the proposed method has high

effectiveness for image annotation.

**17-Nov-2010**

**Morning**

**09:00AM-10:20AM Oral 5 Super-resolution and Image Enhancement**

**Title: Selecting High Quality Frames for Super Resolution Reconstruction Using Perceptual Quality Metrics**

**Authors:** Amit Prabhudesai, Ajit Bopardikar, Chandra Sekhar Reddy

**Abstract:** Super-resolution involves the use of signal processing techniques to estimate the high-resolution (HR) version of a scene from multiple low-resolution (LR) observations. It follows that the quality of the reconstructed HR image would depend on the quality of the LR observations. The latter depends on multiple factors like the image acquisition process, encoding or compression and transmission. However, not all images are equally affected by a given type of impairment. A proper choice of the LR observations for reconstruction, should yield a better estimate of the HR image, over a naive method using all images. We propose a simple, model-free approach to improve the performance of super-resolution systems based on the use of perceptual quality metrics. Our approach does not require, or even assume, a specific realization of the SR system. Instead, we select the image subset with high perceptual quality from the available set of LR images. Finally, we present the logical extension of our approach to select the perceptually significant regions in a given LR image, for use in SR reconstruction.

**Title: Blind Super-resolution for Single Image Reconstruction**

**Authors:** Fei Han, Xiangzhong Fang, Ci Wang

**Abstract:** Image super-resolution reconstructions (SR) require image degradation model (DM) as the prior, however, the actual DM is often unknown in practical applications. In this work, a novel framework is proposed for single image SR, where the explicit DM is unknown. Based on Bayesian MAP, an iteration scheme is adopted to update the reconstructed SR image and the DM estimate. During reconstruction, MRF-Gibbs image prior is incorporated for regularization and example-based machine learning technique is employed to draw the DM estimations back to the potential DM space. The SR resulted images by the proposed method are superior to the ones produced by bicubic interpolation and conventional SR algorithm with incorrect DM, in both aesthetic and quantitative aspects.

**Title: Image Enlargement with Lost High-Frequency Components Estimation Using Clustered Eigenspace-BPLP**

**Authors:** Hideaki Orii, Hideaki Kawano, Noriaki Suetake, Hiroshi Maeda

**Abstract:** In this paper, in order to realize the image enlargement with high performance, we propose a new enlargement method, which restores high-frequency components lost in the interpolation-based methods, by using high-frequency estimation.

The estimation is based on the eigenspace method proposed by Amano et al.. The proposed enlargement method can generate lost high-frequency components by using pairs of low- and high-frequency components of the original image, and is referred as back projection for lost high-frequency components. Experimental results show the proposed method can achieve promising enlargement.

**Title: Adaptive Histogram Analysis for Image Enhancement**

*Authors: Qieshi Zhang, Hiroshi Inaba, Sei-ichiro Kamata*

*Abstract:* One image processing application is to reconstruct the original scene from the low quality images. Considering the idea histogram distribution can reflect good vision effect. So many histogram analyzing based methods have been studied recently. However, some methods require users to set some parameters or condition, and cannot get the optimal results automatically. To overcome those short come, this paper presents an Adaptive Histogram Separation and Mapping (AHSM) method for Backlight image enhancement. First, we separate the histogram by binary tree structure with the proposed Adaptive Histogram Separation Unit (AHSU). And then mapping the Low Dynamic Range (LDR) histogram partition into High Dynamic Range (HDR). By doing this, the excessive or scarcity enhancement can be avoid. The experimental results show that the proposed method can gives better enhancement results, also compared with some histogram analyzing based methods and get better results.

**10:35AM-11:55AM Oral 6 Image Enhancement and Image Filter**

**Title: On the Transfer of Painting Style to Photographic Images through Attention to Colour Contrast**

*Authors: Xiaoyan Zhang, Martin Constable, Ying He*

*Abstract:* This paper proposes a way to transfer the visual style of a painting as characterised by colour contrast to a photographic image by manipulating the visual attributes in terms of hue, saturation and lightness. We first extract the visual characteristics in hue, saturation and lightness from a painting. Then, these characteristics are transferred to a photographic image by histogram matching in saturation and lightness and dominant hue spread and relative position mapping along the RYB colour wheel. We evaluated the proposed transfer method on a number of paintings and photographs. The results are encouraging.

**Title: Color Image Enhancement in HSV Space Using Nonlinear Transfer Function and Neighborhood Dependent Approach with Preserving Details**

*Authors: Deepak Ghimire, Joonwhoan Lee*

*Abstract:* This paper proposes a method for enhancing the color images based on nonlinear transfer function and pixel neighborhood by preserving details. In the proposed method, an input RGB color image is converted into an HSV color image. The image enhancement is applied only on the V (luminance value) component of the color image, because change in the H and S component could change the color balance between HSV components. The V component is enhanced in two steps. At

first the V channel is divided into smaller blocks and in each block dynamic range compression is carried out using nonlinear transfer function. In the second step each pixels in each block are further enhanced for the adjustment of the image contrast depending upon the centre pixel and its neighborhood. The aim behind dividing the image into blocks is to preserve the details. The original H and S component image and enhanced V component image are converted back to the RGB image. The experimental results show that the proposed method yields better performance by preserving details and no change in color in comparison with other methods.

**Title: Threshold Tuning of Switching Median Filter Employing Distribution Distance**

*Authors: Ryosuke Kubota, Noriaki Suetake*

*Abstract:* In this paper, we propose a new threshold tuning method of a switching median filter by using the distribution distance. The switching median filter detects noise-corrupted pixels based on a threshold. Then it restores only the detected pixels. The present switching median filter deals with the random-valued impulse noises, whose distribution is ideally assumed as a uniform distribution. In the present method, the threshold is tuned by evaluating a distribution distance between the assumed and the detected noises. The effectiveness of the proposed method is verified by some experiments.

**Title: Descreening Using HOG-based Adaptive Smoothing Filter**

*Authors: Kyu-Sung Hur, Yeul-Min Baek, Whoi-Yul Kim*

*Abstract:* In this paper, a novel descreening method using a histogram of oriented gradients (HOG) based on an adaptive smoothing filter is proposed. Conventional descreening methods, which are used for recovering continuous-tone from scanned halftone images, are based on low-pass filtering approach. The low-pass filter has been applied to spatial domain, frequency domain, or both to remove the high frequency halftone noise. However, clipping high frequency components tends to blur images because image details also show high frequency characteristics. Moreover, due to the large variance in halftone noise, it is difficult to preserve edges by conventional adaptive smoothing methods. Therefore, to improve the descreening effect, more care has to be taken in distinguishing between edges and halftone noise. The proposed method employs HOG to distinguish edges. The amount of smoothing to be performed on the halftone image is then calculated according to the magnitude of the HOG in the edge and edge normal orientation. The proposed method was tested on various scanned halftone materials, and the results show that it removes halftone noise as effectively as the Moire pattern while still preserving image details.

**Afternoon**

**12:55PM-02:15PM Oral 7 Video Surveillance**

**Title: Tracking Groups of People in Presence of Occlusion**

*Authors: Hamed Kiani Galoogahi*

*Abstract:* This paper addresses the problem of people group tracking in presence of

occlusion as people form groups, interact within groups or leave groups. Foreground objects (a person or a group of people) from two consecutive frames are matched based on appearance (RGB histogram) and object location (2D region) similarity. While tracking, this method determines and handles some events such as objects merging and splitting using forward and backward matching matrices. The experimental results show that the proposed algorithm is efficient to track group of people in cluttered and complex environments even when total or partial occlusion occurs.

**Title: Motion Detection for Rapidly Moving Cameras in Fully 3D Scenes**

**Authors:** Dong Zhang, Ping Li

**Abstract:** This paper presents a novel method for detecting motion regions in image sequences obtained by rapidly moving cameras in fully 3-D scenes. The proposed method mainly focuses on the situations that the backgrounds of the image sequences change rapidly. It has three innovations over existing methods: First, it presents a new initialization method to fast and sparsely gather information of the background model, while traditional methods utilize a complicated training step. Second, a novel model updating scheme is proposed for establishing the on-line sparse background model iteratively. This is the main contribution of the proposed method and this enables the method to work in a 3-D scene which totally changes through the image sequence, while most other methods can only work with a pre-modeled scene or with a camera that moving in limited scope. Third, a novel two-stage model of the background and foreground motion regions is proposed and the foreground motion regions are detected using maximum a posterior estimation. The method is tested on various challenging image sequences captured by freely moving cameras and results show that it is very effective and robust.

**Title: Video Surveillance Traffic Analysis Using Scene Geometry**

**Authors:** M. Brulin, H. Nicolas, C. Maillet

**Abstract:** We present a system to perform video analysis in the context of traffic surveillance's application. A training step is performed to estimate the scene's geometry and global information about the motion that occurs in the scene. Lanes boundaries, depth and motion information given by the initialization step are used to assist the vehicles' segmentation and to correct eventual errors.

**Title: Moving Objects Detection and Tracking Framework for UAV-based Surveillance**

**Authors:** Aryo Wiman Nur Ibrahim, Pang Wee Ching, Gerald Seet G.L., Michael Lau W.S., Witold Czajewski

**Abstract:** Automated motion detection and tracking of ground moving objects using aerial platforms is challenging due to the small object size in comparison with objects such as buildings, as well as the fact that flying cameras can undergo rapid translations and rotations. As such, our objectives are to develop a system for gathering useful information from aerial images by mapping visited areas through image mosaicking and to detect moving objects in the captured video. To do so, the Moving Objects Detection and Tracking (MODAT) framework has been developed to facilitate the application and combination of various relevant computer vision and image processing techniques in order to achieve our objectives.

**02:15PM-03:35PM SP 3 Image/Video Technology Meets Art**

**Title: Artistic Emulation - Filter Blending for Painterly Rendering**

**Authors:** Crystal Valente, Reinhard Klette

**Abstract:** This paper looks at painterly rendering using filter blending to create a novel range of artistic effects. We look into several techniques in the field of painterly rendering and combine these different rendering styles together in a user defined way to create a new filter. This creates a user defined painting style based on aspects of different painting styles and processes. The final application uses a triangular region for human-computer interaction where each corner represents an artistic filter and the middle of this triangle represents the original image. A point within the triangle is chosen to determine the filters that are used for blending and their contributing strengths. Artistic effects are based on three base filters that can be tailored by the user to suit the specific subject matter of the image.

**Title: What Makes Motion Meaningful? Affective Properties of Abstract Motion**

**Authors:** Lyn Bartram, Ai Nakatani

**Abstract:** Motion provides a rich modality for communicating emotion and creating affect, and developing techniques for analyzing, manipulating and enhancing animations in video and computer graphics is an active research area. However, we have little empirical evidence of how particular attributes of motion elicit particular impressions. In this paper we describe a study of how perceptual attributes of abstract motion contribute to affective interpretation. Our results show that even small simple motions can reliably evoke certain impressions given particular motion attributes. These results contribute to our understanding of computational aesthetics and have implications for the design of media experiences, visualization, interactive art and immersive environments. They also provide insight into what properties of motion might prove most evocative in video and animation textures.

**Title: Depth-Aware Video Abstraction**

**Authors:** Jianbing Shen, Ying He

**Abstract:** In this paper, we present a novel depth-aware abstraction approach for videos. By introducing the additional depth cues, our approach is able to enhance the depth perception of the artistic-abstraction video. We first recover consistent depth information from the input video using multiview stereo and energy optimization algorithm. Then, the coherent line is extracted by employing the an isotropic difference-of-Gaussians filter, and the Gaussian Kuwahara filter is further employed to produce the effect of artistic-abstraction. Finally, the depth aware video artistic-abstraction is achieved by unsharp masking the consistent video depth. Our algorithm is effective for artistically representing the depth perception of artistic abstraction video. Experimental results on a variety of videos show that the proposed approach successfully produces a visually satisfying depth perception for the artistic-abstraction video.

**Title: Depth-based Analyses of Landscape Paintings and Photographs According to Itten's Contrasts**

**Authors:** *Martin Constable, Xiaoyan Zhang*

**Abstract:** Using Itten's Color Contrasts as a starting point, we performed a depth-based analysis of a set of paintings by the Hudson River school of landscape painters. This was compared to a similar analyses of a collection of contemporary 'snap shot' landscape photographs. Differences between the two groups were observed with the paintings being clearly more organized. This organization of contrasts can be considered a style that is representative of a school of painters. Photographs or optically acquired imagery can be rendered according to this organization to acquire an aesthetic that has been informed by this style.

**03:50PM-05:10PM SP 4 Graphics and Visualization**

**Title: Sophisticated Construction and Search of 2D Motion Graphs for Synthesizing Videos**

**Authors:** *Jun Kobayashi, Chongke Bi, Shigeo Takahashi*

**Abstract:** This paper presents an intuitive method for synthesizing videos by directly manipulating video objects without using 3D models. The proposed method extracts a video object from each video frame and creates locally consistent video sequences using a 2D motion graph, where its node corresponds to the extracted video object and its edge represents a motion transition between a pair of nodes. Our primary contribution lies in a sophisticated construction of the 2D motion graph using shape matching techniques, and its search that allows us to intuitively synthesize a new video sequence by manipulating feature points extracted from the video objects through the 2D screen space. The method further employs a deformation technique to interpolate between video objects with relatively different shapes, and thus can increase available motion transitions by inserting intervening video objects into the 2D motion graph. Several examples have been generated to demonstrate that this approach can create the user-intended motions of the video objects easily by clicking and dragging the feature points.

**Title: Real-Time GPU-Aided Lung Tumor Tracking**

**Authors:** *Yin Yang, Zichun Zhong, Guodong Rong, Xiaohu Guo, Jing Wang, Timothy Solberg, Weihua Mao*

**Abstract:** A real time solution of tracking daily lung tumor motion is proposed in this paper in order to achieve an accurate dose delivery for radiation therapy as recently developed cone-beam computed tomography (CBCT) technique is not able to catch tumor motions due to the patient's respiration. We develop a novel GPU-based fast digitally reconstructed radiograph (DRR) generation algorithm which enables an instant DRR computation and rendering from patients' radiation therapy treatment planning CTs. In the meantime, classic image correlation algorithm is extended as the main method to locate tumors in X-ray 2D projections, the raw data of CBCT scans. With the GPU-aided implementation, this algorithm is capable of capturing movement of lung tumors as fast as the CBCT image acquisition in real time, which greatly facilitates the radioactive treatment.

**Title: Subband Architecture Based Exposure Fusion**

**Authors:** *Ying Zhao, Jianbing Shen, Ying He*

**Abstract:** In this paper, we present a novel detail-preserving fusion approach from multiple exposure images using sub band architecture. Given a sequence of different exposures, the Quadrature Mirror Filters (QMFs) based sub band architecture is first employed to decompose the original sequence into different frequency sub bands. After that, the importance weight maps are calculated according to the image appearance measurements, such as exposure, contrast and saturation. In order to remove the nonlinear distortion introduced by the sub band decomposition stage, we compute the gain control maps and modify these sub bands based on the calculated gain maps. Finally, the coefficients of sub bands are blended into a high-quality detail-preserving fusion image. Experimental results demonstrate that the proposed approach successfully creates a visually pleasing exposure fusion image.

**Title: Detail Preserving 3D Motion Compression Based on Local Transformation**

**Authors:** *Chang Liu, Zhaoqiang Lai, Jiayi Hu, Jing Hua*

**Abstract:** With the advance of 3D acquisition techniques, more and more 3D motion data with fine details become available. In order to efficiently represent the data, we explored different motion compensation algorithms through conventional techniques and analyzed their performance both in theory and by experiments. A novel motion compression framework based on local transformation is proposed to solve the existing problems. Experiments show that our algorithm can achieve very high compression rate while preserving the fine details. The algorithm is easy to implement and is versatile to different applications including non-skeleton driven motion compression.

