2016 IEEE Workshop on Multimedia Signal Processing

21-23 September 2016, Montreal, Canada

mmsp 2016

Book of Abstracts

Wednesday, Sept. 21, 2016

Oral session: Image/video quality assessment (10:00-11:20am)

Chair: Hantao Liu, Cardiff University

104 - Saliency in Objective Video Quality Assessment: What is the Ground Truth?

Hantao Liu, Cardiff University; Wei Zhang, Cardiff University*

Abstract: Finding ways to be able to objectively and reliably assess video quality as would be perceived by humans has become a pressing concern in the multimedia community. To enhance the performance of video quality metrics (VQMs), a research trend is to incorporate visual saliency aspects. Existing approaches have focused on utilising a computational saliency model to improve a VQM. Since saliency models still remain limited in predicting where people look in videos, the benefits of inclusion of saliency in VQMs may heavily depend on the accuracy of the saliency model used. To gain an insight into the actual added value of saliency in VQMs, ground truth saliency obtained from eye-tracking instead of computational saliency is an essential prerequisite. However, collecting eye-tracking data within the context of video quality is confronted with a bias due to the involvement of massive stimulus repetition. In this paper, we introduce a new experimental methodology to alleviate such potential bias and consequently, to be able to deliver reliable intended data. We recorded eye movements from 160 human observers while they freely viewed 160 video stimuli distorted with different distortion types at various degradation levels. We analyse the extent to which ground truth saliency as well as computational saliency actually benefit existing state of the art VQMs. Our dataset opens new challenges for saliency modelling in video quality research and helps better gauge progress in developing saliency-based VQMs.

60 - SIQ288: A Saliency Dataset for Image Quality Research

Hantao Liu, Cardiff University; Wei Zhang, Cardiff University*

Abstract: Saliency modelling for image quality research has been an active topic in multimedia over the last five years. Saliency aspects have been added to many image quality metrics (IQMs) to improve their performance in predicting perceived quality. However, challenges to optimising the performance of saliency-based IQMs remain. To make further progress, a better understanding of human attention deployment in relation to image quality through eye-tracking experimentation is indispensable. Collecting substantial eye-tracking data is often confronted with a bias due to the involvement of massive stimulus repetition that typically occurs in an image quality study. To mitigate this problem, we proposed a new experimental methodology with dedicated control mechanisms, which allows collecting more reliable eye-tracking data. We recorded 5760 trials of eye movements from 160 human observers. Our

dataset consists of 288 images representing a large degree of variability in terms of scene content, distortion type as well as degradation level. We illustrate how saliency is affected by the variations of image quality. We also compare state of the art saliency models in terms of predicting where people look in both original and distorted scenes. Our dataset helps investigate the actual role saliency plays in judging image quality, and provides a benchmark for gauging saliency models in the context of image quality.

146 - Boosting in Image Quality Assessment

Dogancan Temel, Georgia Tech; Ghassan AlRegib, Georgia Institute of Technology*

Abstract: In this paper, we analyze the effect of boosting in image quality assessment through multi-method fusion. Existing multi-method studies focus on proposing a single quality estimator. On the contrary, we investigate the generalizability of multi-method fusion as a framework. In addition to support vector machines that are commonly used in the multi-method fusion, we propose using neural networks in the boosting. To span different types of image quality assessment algorithms, we use quality estimators based on fidelity, perceptually-extended fidelity, structural similarity, spectral similarity, color, and learning. In the experiments, we perform k-fold cross validation using the LIVE, the multiply distorted LIVE, and the TID 2013 databases and the performance of image quality assessment algorithms are measured via accuracy-, linearity-, and ranking-based metrics. Based on the experiments, we show that boosting methods generally improve the performance of image quality assessment and the level of improvement depends on the type of the boosting algorithm. Our experimental results also indicate that boosting the worst performing quality estimator with two or more additional methods leads to statistically significant performance enhancements independent of the boosting technique and neural network-based boosting outperforms support vector machine-based boosting when two or more methods are fused.

103 - Perceptual Video Quality Assessment: Spatiotemporal Pooling Strategies for Different Distortions and Visual Maps

Mohammed Aabed, Georgia Inst. of Technology; Ghassan AlRegib, Georgia Institute of Technology*

Abstract: In this paper, we investigate the challenge of distortion map feature selection and spatiotemporal pooling in perceptual video quality assessment (PVQA). We analyze three distortion maps representing different visual features spatially and temporally: squared error, local pixel-level SSIM, and absolute difference of optical flow magnitudes. We examine the performance of each of these maps with different spatial and temporal pooling strategies across three databases. We identify the most effective statistical pooling strategies spatially and temporally with respect to PVQA. We also show the most significant spatial and temporal features correlated with perception for every distortion/feature map. Our results show that varying the pooling strategy and distortion maps yields a significant improvement in perceptual quality estimation. We also deduce insights from our results to better understand the sensitivity of human vision to distortions. We aim for these findings to provide perceptual cues and guidelines to researchers during metric design, perceptual feature selection, HVS modeling and pooling selection/optimization. We further show that the same distortions across databases can yield different results in terms of PVQA evaluation and verification.

Poster session: Image processing, compression and applications (11:20-12:40pm)

Chair: Ghassan AlRegib, Georgia Institute of Technology

33 - Learning Graph Fusion for Query and Database Specific Image Retrieval
Chih-Kuan Yeh, National Taiwan University; Wei-Chieh Wu, National Taiwan University; Y.-C. Frank Wang, Academia Sinica*

Abstract: In this paper, we propose a graph-based image retrieval algorithm via query-database specific feature fusion. While existing feature fusion approaches exist for image retrieval, they typically do not consider the image database of interest (i.e., to be retrieved) for observing the associated feature contributions. In the offline learning stage, our proposed method first identifies representative features for describing images to be retrieved. Given a query input, we further exploit and integrate its visual information and utilize graph-based fusion for performing query-database specific retrieval. In our experiments, we show that our proposed method achieves promising performance on the benchmark database of UKbench, and performs favorably against recent fusion-based image retrieval approaches.

66 - Content-adaptive Non-parametric Texture Similarity Measure
Motaz Alfarraj, Georgia Institute of Technolog; Yazeed Alaudah, Georgia Institute of Technology; Ghassan AlRegib, Georgia Institute of Technology*

Abstract: In this paper, we introduce a non-parametric texture similarity measure based on the singular value decomposition of the curvelet coefficients followed by a content-based truncation of the singular values. This measure focuses on images with repeating structures and directional content such as those found in natural texture images. Such textural content is critical for image perception and its similarity plays a vital role in various computer vision applications. In this paper, we evaluate the effectiveness of the proposed measure using a retrieval experiment. The proposed measure outperforms the state-of-the-art texture similarity metrics on CUReT and PerTex texture databases, respectively.

161 - Image Coding Using Parametric Texture Synthesis
Uday Thakur, RWTH; Bappaditya Ray, RWTH*

Abstract: Visual textures like grass, water etc. consist of dense and random variations in contrast that are perceptually indistinguishable by a human eye. Such textures are costly to encode using image and video codecs. For example, in the state of the art compression standard High Efficiency Video Coding (HEVC), detailed textures

typically show relatively strong blurring artifacts at low rates (high QP). Texture synthesis is a process whereby one can obtain a reconstruction of a visually equivalent texture with decent visual quality, given a set of parameters. In this paper, texture synthesis is used as a tool in combination with HEVC, exploiting Human Visual Perception (HVP) properties by creating an artificial textured content using model parameters at the decoder side. A novel scheme for compression (prediction and quantization) of parameters for complex wavelet based texture synthesis is introduced. The compressed parameters are sufficient to synthesize high quality texture content at the decoder side. Simulation results have shown, that with same rates, both the subjective and the objective quality is enhanced, compared to HEVC.

57 - Optimizing Tone Mapping Operators for Keypoint Detection under Illumination Changes

Aakanksha Rana, Telecom Paristech; Giuseppe Valenzise, CNRS Telecom ParisTech; Frederic Dufaux, Telecom ParisTech, France*

Abstract: Tone mappings operators (TMO) have recently raised interest for their capability to handle illumination changes. However, these TMOs are optimized with respect to perception rather than image analysis tasks like keypoint detection. Moreover, no work has been done to analyze the factors affecting the optimization of TMOs for such tasks. In this paper, we investigate the influence of two factors--Correlation Coefficient (CC) and Repeatability Rate (RR) of the tone mapped images for the optimization of classical Retinex based models to enhance keypoint detection under illumination changes. CC-based optimized models aim at increasing the similarity of the tone mapped images. Conversely, RR-based optimized models quantify the optimal detection performance gains. By considering two simple Retinex based models, i.e., Gaussian and bilateral filtering, we show that estimating as precisely as possible the illumination, CC-based optimized models do not necessarily bring to optimal keypoint detection performance. We conclude that, instead, other criteria specific to RR-based optimized models should be taken into account. Moreover, large gains in performance with respect to existing popular TMOs motivate further research towards optimal tone mapping technique for computer vision applications.

17 - Robust Propagated Filtering with Applications to Image Texture Filtering and Beyond

Dennis Wen, Academia Sinica; Y.-C. Frank Wang, Academia Sinica*

Abstract: Extracting meaningful structures from an image is an important task and benefits a wide range of image application tasks. However, it is typically very challenging to distinguish between noisy or textural patterns from image structures, especially when such patterns do not exhibit regularity (e.g., irregular textural patterns or those with varying scales). While existing edge-preserving image filters like bilateral, guided, or propagation filters aim at observing strong image edges, they cannot be easily applied to solve the above texture filtering tasks. In this paper, we propose robust propagated filter, which is an extension to propagation filters while exhibiting excellent ability in eliminating the aforementioned textural patterns when performing filtering. We show in our experimental results that our filter provides promising results on image filtering. Additional experiments on inverse image halftoning and detail enhancement further verify the effectiveness of our proposed method.

133 - Manipulating 4D Light Fields for Creative Photography

Jana Ehmann, LG Electronics*

Abstract: We demonstrate how 4D light field images can be used to create artistic photographic effects. Traditionally, 4D light fields have been used in photography to perform digital image refocusing, rendering of 3D models, and creating anaglyphs for 3D viewing. By having information about the paths of light within a camera, we can simulate different optical aberrations to achieve results similar to those of novelty lenses, such as the ones from the LensBaby family. The images can be created realistically, with minimum human intervention in adjusting the parameters for desired types and degrees of effects.

23 - Multi-Image Super-Resolution Using a Locally Adaptive Denoising-Based Refinement

Michel Bätz, Friedrich-Alexander University; Ján Koloda, Friedrich-Alexander University; Andrea Eichenseer, Friedrich-Alexander University; Andre Kaup, FAU*

Abstract: Spatial resolution enhancement is of particular interest in many applications such as entertainment, surveillance, or automotive systems. Besides using a more expensive, higher resolution sensor, it is also possible to apply super-resolution techniques on the low resolution content. Super-resolution methods can be basically classified into single-image and multi-image super-resolution. In this paper, we propose the integration of a novel locally adaptive denoising-based refinement step as an intermediate processing step in a multi-image super-resolution framework. The idea is to be capable of removing reconstruction artifacts while preserving the details in areas of interest such as text. Simulation results show an average gain in luminance PSNR of up to 0.2 dB and 0.3 dB for an upscaling of 2 and 4, respectively. The objective results are substantiated by the visual impression.

24 - Reliability-based Mesh-to-Grid Image Reconstruction

Ján Koloda, Friedrich-Alexander University; Jürgen Seiler; Andre Kaup, FAU*

Abstract: This paper presents a novel method for the reconstruction of images from samples located at non-integer positions, called mesh. This is a common scenario for many image processing applications, such as super-resolution, warping or virtual view generation in multi-camera systems. The proposed method relies on a set of initial estimates that are later refined by a new reliability-based content-adaptive framework that employs denoising in order to reduce the reconstruction error. The reliability of the initial estimate is computed so stronger denoising is applied to less reliable estimates. The proposed technique can improve the reconstruction quality by more than 2 dB (in terms of PSNR) with respect to the initial estimate and it outperforms the state-of-the-art denoising-based refinement by up to 0.7 dB.

122 - Efficient Detail-enhanced Exposure Correction Based on Auto-fusion for LDR Image

Jiayi Chen, Xi'an Jiaotong University; Xuguang Lan, Xi'an Jiaotong University; Meng Yang, Xi'an Jiaotong University*

Abstract: We consider the problem of how to simultaneously and well correct the over- and under-exposure regions in a single low dynamic range (LDR) image. Recent methods typically focus on global visual quality but cannot well-correct much potential details in extremely wrong exposure areas, and some are also time consuming. In this paper, we propose a fast and detail-enhanced correction method based on automatic fusion which combines a pair of complementarily corrected images, i.e. backlight & highlight correction images (BCI &HCI). A BCI with higher visual quality in details is quickly produced based on a proposed faster multi-scale retinex algorithm; meanwhile, a HCI is generated through contrast enhancement method. Then, an automatic fusion algorithm is proposed to create a color-protected exposure mask for fusing BCI and HCI when avoiding potential artifacts on the boundary. The experiment results show that the proposed method can fast correct over/under-exposed regions with higher detail quality than existing methods.

89 - Image Compression using Adaptive Sparse Representations over Trained Dictionaries

Ali Akbari, ISEP; Maria Trocan, Institut Supérieur d'Électronique de Paris; Bertrand Granado, upmc*

Abstract: Sparse representation is a common approach for reducing the spatial redundancy by modelling an image as a linear combination of few atoms taken from an analytic or trained dictionary. This paper introduces a new image codec based on adaptive sparse representations wherein the visual salient information is considered into the rate allocation process. Firstly, the regions of the image that are more conspicuous to the human visual system are extracted using a classical graph-based method. Further, block-based sparse representation over a trained dictionary coupled with an adaptive sparse representation is proposed, such that the adaptivity is achieved by appropriately assigning more atoms of the dictionary to the blocks belonging to the salient regions. Experimental results show that the proposed method outperforms the existing image coding standards, such as JPEG and JPEG2000, which use an analytic dictionary, as well as the state-of-the-art codecs based on trained dictionaries.

35 - Adaptive Frequency Prior for Frequency Selective Reconstruction of Images from Non-Regular Subsampling

Jürgen Seiler; Andre Kaup, FAU*

Abstract: Image signals typically are defined on a rectangular two-dimensional grid. However, there exist scenarios where this is not fulfilled and where the image information only is available for a non-regular subset of pixel position. For processing, transmitting or displaying such an image signal, a resampling to a regular grid is required. Recently, Frequency Selective Reconstruction (FSR) has been proposed as a very effective sparsity-based algorithm for solving this under-determined problem. For this, FSR iteratively generates a model of the signal in the Fourier-domain. In this context, a fixed frequency prior inspired by the optical transfer function is used for favoring low-frequency content. However, this fixed prior is often too strict and may lead to a reduced reconstruction quality. To resolve this weakness, this paper proposes an adaptive frequency prior which takes the local density of the available samples into account. The proposed adaptive prior allows for a very high reconstruction quality, yielding gains of up to 0.6 dB PSNR over the fixed prior, independently of the density of the available samples. Compared to other state-of-the-art algorithms, visually noticeable gains of several dB are possible.

168 - Efficient Imaging through Scattering Media by Random Sampling

Yifu Hu, Tsinghua University; Xin Jin, Tsinghua University; Qionghai Dai, Tsinghua University*

Abstract: Imaging through scattering media is a tough task in computational imaging. A recent breakthrough technique based on speckle scanning was proposed with outstanding imaging performance. However, to achieve high imaging quality, dense sampling of the integrated intensity matrix is needed, which leads to a time-consuming scanning process. In this paper, we propose a method that exploits spatial redundancy of the integrated intensity matrix and reconstructs the complete matrix from few random samples. A reconstruction model that jointly penalizes total variation and weighted sum of nuclear norm of local patches is built with improved reconstruction quality. Experiments are performed to verify the effectiveness of the proposed method and results demonstrate that the proposed method can achieve a same imaging quality with 80% reduction of the data acquisition complexity.

Oral (Special) Session: Physiology-based Quality-of-Experience assessment and user affect-aware interfaces (2:40-4:00pm)

Chair: Sebastian Möller, TUB and Tiago H. Falk, INRS-EMT

40 - Using Cardio-Respiratory Signals to Recognize Emotions Elicited by Watching Music Video Clips

Leila Mirmohamadsadeghi, EPFL; Ashkan Yazdani, EPFL; Jean-Marc Vesin, EPFL*

Abstract: The automatic recognition of human emotions from physiological signals is of increasing interest in many applications. Images with high emotional content have been shown to alter signals such as the electrocardiogram (ECG) and the respiration among many other physiological recordings. However, recognizing emotions from multimedia stimuli, such as music video clips, which are growing in numbers in the digital world and are the medium of many recommendation systems, has not been adequately investigated. This study aims to investigate the recognition of emotions elicited by watching music video clips, from features extracted from the ECG, the respiration and several synchronization aspects of the two. On a public dataset, we achieved higher classification rates than the state-of-the-art using either the ECG or the respiration signals alone. A feature related to the synchronization of the two signals achieved even better performance.

97 - Effect of content features on short-term video quality in the visual periphery

Ahmed Aldahdooh, ; Yashas Rai, University of Nantes; Suiyi Ling, University of Nantes; Marcus Barkowsky, University of Nantes; Patrick Le Callet, University of Nantes*

Abstract: The area outside our central field of vision, also referred to as the visual periphery, captures most information in a visual scene, although much less sensitive than the central Fovea. Vision studies in the past have stated that there is reduced sensitivity of texture, color, motion and flicker (temporal harmonic) perception in this area, that bears an interesting application in the domain of quality perception. In this work, we particularly analyze the perceived subjective quality of videos containing H.264/AVC transmission impairments, incident at various degrees of retinal eccentricities of subjects. We relate the perceived drop in quality, to five basic types of features that are important from a perceptive standpoint: texture, color, flicker, motion trajectory distortions and also the semantic importance of the underlying regions. We are able to observe that the perceived drop in quality across the visual periphery, is closely related to the Cortical Magnification fall-off characteristics of the V1 cortical region. Additionally, we see that while object importance and low frequency spatial distortions are important indicators of quality in the central foveal region, temporal flicker and color distortions are the most important determinants of quality in the periphery. We therefore conclude that, although human observers are more forgiving of distortions they viewed peripherally, they are nevertheless not totally blind towards it: the effects of flicker and color distortions being particularly important.

157 - Affective States Classification using EEG and Semi-supervised Deep Learning Approaches

Haiyan Xu, University of Toronto; Konstantinos Plataniotis, University of Toronto*

Abstract: Affective states of a user provide important information for many applications such as, personalized information (e.g., multimedia content) retrieval/delivery or intelligent human-computer interface design. In recently years, physiological signals, Electroencephalogram (EEG) in particular, have been shown to be very effective in estimating a user's affective states during social interaction or under video or audio stimuli. However, due to the large number of parameters associated with the neural expression of emotion, there is still a lot of unknowns on the specific spatial and spectral correlation of the EEG signal and the affective states expression. To investigate on such correlation, two types of semi-supervised deep learning approaches, stacked denoising autoencoder (SDAE) and deep belief networks (DBN), were applied as application specific feature extractors for the affective states classification problem using EEG signals. To evaluate the efficacy of the proposed semi-supervised approaches, a subject-specific affective states classification experiment were carried out on the DEAP database to classify 2-dimensional affect states. The DBN based model achieved averaged F1 scores of 86.67%, 86.60% and 86.69% for arousal, valence and liking states classification respectively, which has significantly improved the state-of-art classification performance. By examining the weight vectors at each layer, we were also able to gain insights on the spatial or spectral locations of the most discriminating features. Another main advantage of applying the semi-supervised learning methods is that only a small fraction of labeled data, e.g., 1/6 of the training samples, were used in this study.

170 - Physiological Quality-of-Experience Assessment of Text-to-Speech Systems

Rishabh Gupta, INRS-EMT; Tiago Falk, INRS-EMT*

Abstract: With the emergence of various text-to-speech (TTS) systems, developers have to provide superior user experience in order to remain competitive. To this end, quality-of-experience (QoE) perception modelling and measurement has become a key priority. QoE models rely on three influence factors: technological, contextual and human. Existing solutions have typically relied on using individual physiological modalities, such as electroencephalography (EEG), to model human influence factors (HIFs). In this paper, we show that fusion of physiological modalities, such as EEG, functional near infrared spectroscopy (fNIRS) and heart rate, provide gains of up to 18.4% relative to utilizing only technological factors and 4% relative to using the best performing individual physiological modality.

Thursday, Sept. 22, 2016

Oral session: Fast/parallel HEVC (10:00-11:20am)

Chair: Stephane Coulombe, ETS

105 - Efficient HEVC Decoder for Heterogeneous CPU with GPU Systems

Biao Wang, Institut für Technische Informatik und Mikroelektronik, Technische Universität B; Diego De Souza, INESC-ID, IST, ULisboa; Mauricio Alvarez-Mesa, Technische Universität Berlin; Chi Chi Ching, Technische Universität Berlin; Ben Juurlink, Technische Universität Berlin; Aleksandar Ilic, INESC-ID, Instituto Superior Técnico, Universidade de Lisboa; Nuno Roma, INESC-ID, Instituto Superior Técnico,* *Universidade de Lisboa; Leonel Sousa, INESC-ID, Instituto Superior Técnico, Universidade de Lisboa*

Abstract: The High Efficiency Video Coding (HEVC) standard provides higher compression efficiency than other video coding standards but at the cost of increased computational load, which makes it hard to achieve real-time encoding/decoding of high-resolution, high-quality video sequences. In this paper, we investigate how Graphics Processing Units (GPUs) can be employed to accelerate HEVC decoding. GPUs are known to provide massive processing capability for throughput computing kernels, but the HEVC entropy decoding kernel cannot be executed efficiently on GPUs. We therefore propose a complete HEVC decoding solution for heterogeneous CPU+GPU systems, in which the entropy decoder is executed on the CPU and the remaining kernels on the GPU. Furthermore, the decoder is pipelined such that the CPU and the GPU can decode different frames in parallel. The proposed CPU+GPU decoder achieves an average frame rate of 150 frames per second for Ultra HD 4K video sequences when four CPU cores are used with an NVIDIA GeForce Titan X GPU.

74 - Slice-Based Parallelization in HEVC Encoding: Realizing the Potential through Efficient Load Balancing

Maria Koziri, University of Thessaly; Panos Papadopoulos, University of Thessaly; Nikos Tziritas, CAS/SIAT, China; Antonios Dadaliaris, University of Thessaly, Greece; Thanasis Loukopoulos, University of Thessaly; Samee Khan, North Dakota State University, Fargo, USA*

Abstract: The new video coding standard HEVC (High Efficiency Video Coding) offers the desired compression performance in the era of HDTV and UHDTV, as it achieves nearly 50% bit rate saving compared to H.264/AVC. To leverage the involved computational overhead, HEVC offers three parallelization potentials namely: wavefront parallelization, tile-based and slice-based. In this paper we study slice-based parallelization of HEVC using OpenMP on the encoding part. In particular we delve on the problem of proper slice sizing to reduce load imbalances among threads. Capitalizing on existing ideas for H.264/AVC we develop a fast dynamic approach to decide on load distribution and compare it against an alternative in the HEVC literature. Through experiments with commonly used video sequences, we highlight the merits and drawbacks of the tested heuristics. We then improve upon them for the case of Low-Delay by exploiting GOP structure. The resulting algorithm is shown to clearly outperform its counterparts achieving less than 10% load imbalance in many cases.

143 - Fast Mode Decision for HEVC Intra Coding With Efficient Mode Skipping and Improved RMD

Xin Lu, Harbin Institute of Technology; Nan Xiao; Yue Hu; Zhilu Wu; Graham Martin.*

Abstract: HEVC employs a quad-tree based Coding Unit (CU) structure to achieve a significant improvement in coding efficiency compared with previous standards. However, the computational complexity is greatly increased. We proposed a fast

mode decision algorithm to reduce intracoding complexity. Firstly, an initial candidate list of intra modes is constructed for each Prediction Unit (PU). The prediction mode correlation between adjacent quad-tree coding levels and between temporal neighbouring frames is used to predict the most likely coding mode. The number of prediction mode that need to be evaluated in residual quad-tree (RQT) process is further reduced by taking the Hadamard cost of prediction
mode into consideration. Simulation results show that the proposed algorithm saves encoding time by up to 51% compared with the HM 13.0 implementation, while having a negligible impact on rate distortion.

31 - Coding Unit Splitting Early Termination for Fast HEVC Intra Coding Based on Global and Directional Gradients

Mohammadreza Jamali, École de technologie supérieur; Stephane Coulombe, ETS*

Abstract: High efficiency video coding (HEVC) doubles the compression ratio as compared to H.264/AVC, for the same quality. To achieve this improved coding performance, HEVC presents a new content-adaptive approach to split a frame into coding units (CUs), along with an increased number of prediction modes, which results in significant computational complexity. To lower this complexity with intra coding, in this paper, we develop a new method based on global and directional gradients to terminate the CU splitting procedure early and prevent processing of unnecessary depths. The global and directional gradients determine if the unit is predicted with high accuracy at the current level, and where that's the case, the CU is deemed to be non-split. Experimental results show that the proposed method reduces the encoding time by 52% on average, with a small quality loss of 0.07 dB (BD-PSNR) for all-intra scenarios, as compared to the HEVC reference implementation, HM 15.0.

Poster session: HEVC, video coding/transcoding (11:20-12:40pm)

Chair: Ricardo de Queiroz, Universidade de Brasilia

167 - Single-Input-Multiple-Ouput Transcoding For Video Streaming

Chengzhi Wang, SJTU; Bo Li, Shanghai Jiaotong University; Jie Wang, Central South University; Hao Zhang, Central South University; Hao Chen, Shanghai Jiaotong University; Yiling Xu, Cooperative MediaNet Innovation Center, Shanghai Jiao Tong University; Zhan Ma, Nanjing University*

Abstract: In this work, a single input multiple output (SIMO) transcoding architecture is proposed. SIMO will benefit the mobile edge computing (such as HTTP Live Streaming requiring multiple copies of the video streams at different quality levels) without resorting to the legacy transcoding that video stream is completed decoded and encoded multiple times without exploring the compressed information. Leveraging the information encoded in the existing video streams, we could reduce the search candidates when transcoding the high quality bitstream to other versions with reduced quality level. As the first step, we have demonstrated the SIMO idea with bit

rate shaping (i.e., bit rate transcoding) only scenario. It has shown more than 2x complexity reduction without quality loss using the common test conditions.

159 - A Drift Compensated Reversible Watermarking Scheme for H.265/HEVC

Sibaji Gaj; Shuvendu Rana; Arijit Sur; Prabin Bora - IIT Guwahati*

Abstract: In this paper, a compressed domain drift compensated reversible watermarking scheme is proposed with a high embedding capacity and the least amount of visual quality degradation for H.265/HEVC videos. Using compressed domain syntax elements, such as motion vector and transformed residual, a set of 4x4 Transform Blocks (TB) of similar texture are chosen from consecutive I Frames for watermark embedding. Due to texture similarity of these selected TBs, the differences between the transformed coefficients are equal or close to zero. Utilizing this difference statistics, a multilevel watermarking is inserted in the compressed video by altering near zeros values in the difference transformed coefficients. A comprehensive set of experiments has been carried out to justify the efficacy of the proposed scheme over existing literature.

131 - Optimised Selection of Structure of Pictures for Video Coding

Vignes Poobalasingam, Queen Mary University of London; Saverio Blasi, BBC; Marta Mrak, BBC; Ebroul Izquierdo, Queen Mary University of London*

Abstract: Encoders based on the High Efficiency Video Coding (HEVC) standard consider an input sequence as a succession of slices grouped in Structures of Pictures (SOP). The SOP used while encoding specifies many parameters, such as the coding order of frames, or the reference frames used during inter-prediction. Reference encoders typically make use of a fixed SOP structure of a given size, which is periodically repeated throughout the whole sequence. In this paper, the usage of unconventional SOP structures is first analysed, showing that most sequences benefit from usage of larger SOPs, and that the selection of the optimal SOP is highly content dependent. As a result, an algorithm is proposed to automatically select the optimal SOP size based on a low-complexity texture analysis of neighbouring frames. The algorithm is capable of adaptively changing SOP size during the encoding. Extensive evaluation shows that consistent bitrate reductions are reported at the same objective quality as an effect of using the proposed algorithm.

126 - A New AL-FEC Coding Scheme for Mobile Video Broadcasting with Limited Feedback

Wei Huang, Shanghai Jiaotong University; Hao Chen, Shanghai Jiaotong University; Yiling Xu, Cooperative MediaNet Innovation Center, Shanghai Jiao Tong University; Zhu Li, Dept of CSEE, University of Missouri, Kansas City; Lianghui Ding, Shanghai Jiaotong University; Wenjun Zhang, Shanghai Jiatong University*

Abstract: For the next generation mobile video broadcasting, especially in-band solutions that serves the mobile devices, a limited feedback scheme via cellular channel polling is feasible to give accurate real-time information on the broadcast receivers' channel erasure rate, and decoding buffer status. In this work, We analyze the problems in traditional fountain codes especially Raptor codes and come up with the relationship between the buffer status and the optimal coding degree. Then we propose an application layer forward error correction (AL-FEC) coding degree scheme based on this feedback, to achieve a better decode efficiency and save the code redundancy. Simulation results demonstrate the effectiveness of this solution both in decoding speed and redundancy saving, and open up new opportunities in the next generation broadcasting system design.

39 - Adaptive Color Space Transforms For 4:4:4 Video Coding Considering Uncorrelated Noise Among Color Components

Kodai Kikuchi, NHK; Takeshi Kajiyama, NHK; Kei Ogura, NHK; Eiichi Miyashita, NHK*

Abstract: We propose a color space transform for 4:4:4 video coding. The uncorrelated noise contained in specific color component deteriorates compression efficiency. The proposed transform can limit the propagation of the uncorrelated noise of the color component to other color components by introducing zero coefficients in a transform matrix. Simulation results show that the proposed transform improves image quality after compression using extended JPEG in certain natural images. To achieve high image quality for any kind of image, an adaptive selection among conventional and the proposed color space transforms were performed by calculating the independency among color-difference components.

47 - Daala: Building A Next-Generation Video Codec From Unconventional Technology

Jean-Marc Valin; Timothy Terriberry; Nathan Egge; Thomas Daede; Yushin Cho; Christopher Montgomery; Michael Bebenita - Mozilla.*

Abstract: Daala is a new royalty-free video codec that attempts to compete with state-of-the-art royalty-bearing codecs. To do so, it must achieve good compression while avoiding all of their patented techniques. We use technology that is as different as possible from traditional approaches to achieve this. This paper describes the technology behind Daala and discusses where it fits in the newly created AV1 codec from the Alliance for Open Media. We show that Daala is approaching the performance level of more mature, state-of-the art video codecs and can contribute to improving AV1.

130 - Motion Classification-based Fast Motion Estimation for HEVC

Rui Fan; Bo Li; Yongfei Zhang; Yang Liu, Beijing Key Laboratory of Digital Media, Beihang University, Beijing Institute of Graphics.*

Abstract: Motion estimation (ME) is one of the most efficient but time-consuming coding tools in the new high efficiency video coding (HEVC) standard. Test zone search (TZS) is adopted as the fast algorithm for ME in the reference software of HEVC. However, the high complexity of TZS still prevents HEVC from being applied in real-time applications. Several faster algorithms have been proposed to speed up the ME process, which however result in severe performance loss due to simple search patterns or inaccurate early termination. To reduce the ME complexity while maintaining the coding performance, a motion classification-based fast integer-pixel ME algorithm is proposed in this paper. First, prediction units are classified into two categories according to the motion information of neighboring prediction units. Second, different search strategies are implemented for each PU according to the motion characteristics of different categories. Experimental results demonstrate that the proposed algorithm saves 12.16% total encoding time on average when compared with TZS with almost no performance loss and outperforms the state-of-the-art fast ME algorithms by achieving a better RD performance with approximate encoding time saving.

156 - Hybrid Video Object Tracking in H.265/HEVC Video Streams

Serhan Gül, Fraunhofer HHI; Jan Timo Meyer, Fraunhofer HHI; Cornelius Hellge, Fraunhofer HHI; Thomas Schierl, Fraunhofer HHI; Wojciech Samek, Fraunhofer HHI*

Abstract: In this paper we propose a hybrid tracking method which detects moving objects in videos compressed according to H.265/HEVC standard. Our framework largely depends on motion vectors (MV) and block types obtained by partially decoding the video bitstream and occasionally uses pixel domain information to distinguish between two objects. The compressed domain method is based on a Markov Random Field (MRF) model that captures spatial and temporal coherence of the moving object and is updated on a frame-to-frame basis. The hybrid nature of our approach stems from the usage of a pixel domain method that extracts the color information from the fully-decoded I frames and is updated only after completion of each Group-of-Pictures (GOP). We test the tracking accuracy of our method using standard video sequences and show that our hybrid framework provides better tracking accuracy than a state-of-the-art MRF model.

100 - Lossless Compression In HEVC With Integer-To-Integer Transforms

Fatih Kamisli, METU*

Abstract: Many approaches have been proposed to support lossless coding within video coding standards that are primarily designed for lossy coding. The simplest approach is to just skip transform and quantization and directly entropy code the prediction residual, which is the approach used in HEVC version 1. However, this simple approach is inefficient for compression. More efficient approaches include processing the residual with DPCM prior to entropy coding, which is the approach used in HEVC version 2. This paper explores an alternative approach based on processing the residual with integer-to-integer (i2i) transforms. I2i transforms map integers to integers, however, unlike the integer transforms used in HEVC for lossy coding, they do not increase the dynamic range at the output and can be used in lossless coding. Experiments with the HEVC reference software show competitive results.

88 - Background Simplification For ROI-Oriented Low Bitrate Video Coding

Benoit Boyadjis, Thales; Cyril Bergeron; Béatrice Pesquet-Popescu; Frederic Dufaux, Telecom ParisTech, France*

Abstract: Low-bitrate video compression is a challenging task, particularly with the increasing complexity of video sequences. Re-shaping video data before its compression with modern hybrid encoders has provided interesting results in the low and ultra-low bitrate domains. In this work, we propose a novel saliency guided preprocessing approach, which combines adaptive re-sampling and background texture removal, to achieve efficient ROI-oriented compression. Evaluated with HEVC, we show that our solution improves the ROI encoding over a wide range of resolutions and bitrates whilst maintaining a high background intelligibility level.

Oral session: 3D image/video analysis (2:40-4:00pm)

Chair: Frédéric Dufaux, Telecom ParisTech

82 - An Embedded 3D Geometry Score For Mobile 3D Visual Search

Hanwei Wu, KTH; Haopeng Li, KTH Royal Institute of Technology; Markus Flierl, KTH Royal Institute of Technology*

Abstract: The scoring function is a central component in mobile visual search. In this paper, we propose an embedded 3D geometry score for mobile 3D visual search (M3DVS). In contrast to conventional mobile visual search, M3DVS uses not only the visual appearance of query objects, but utilizes also the underlying 3D geometry. The proposed scoring function interprets visual search as a process that reduces uncertainty among candidate objects when observing a query. For M3DVS, the uncertainty is reduced by both appearance-based visual similarity and 3D geometric similarity. For the latter, we give an algorithm for estimating the query-dependent threshold for geometric similarity. In contrast to visual similarity, the threshold for geometric similarity is relative due to the constraints of image-based 3D reconstruction. The experimental results show that the embedded 3D geometry score improves the recall-data rate performance when compared to a conventional visual score or 3D geometry-based re-ranking.

155 - Segmentation Based 3D Depth Watermarking using SIFT

Shuvendu Rana, IIT Guwahati; Sibaji Gaj, IIT Guwahati; Arijit Sur, IIT Guwahati; Prabin Bora, IIT Guwahati*

Abstract: In this paper, a 3D image watermarking scheme is proposed to embed the watermark with the depth of the 3D image for depth image based rendering (DIBR) 3D image representation. To make the scheme invariant to view synthesis process, watermark is inserted with the scale invariant feature transform (SIFT) feature point locations obtained from the original image. Moreover, embedding zone for watermarking has been selected in such a way that no watermark can be inserted in the foreground object to avoid perceptible artefacts. Also, a novel watermark embedding policy is used to insert the watermark with the depth of the 3D image to resist the image processing attacks. A comprehensive set of experiments are carried out to justify the robustness of the proposed scheme.

158 - Detection of Fake 3D Video Using CNN

Shuvendu Rana, IIT Guwahati; Sibaji Gaj, IIT Guwahati; Arijit Sur, IIT Guwahati; Prabin Bora, IIT Guwahati*

Abstract: In this paper, a novel automatic fake and the real 3D video recognition scheme is proposed to distinguish the 3D video converted from the 2D video using 2D to 3D conversion process (say fake 3D) from the 3D video captured using direct capturing of the 3D camera (say real 3D). To identify the real and fake 3D, pre-filtration is done using the dual tree complex wavelet transform to emerge the edge and vertical and horizontal parallax characteristics of real and fake 3D videos. Convolution neural network (CNN) is used to train the 3D characteristics to distinguish the fake 3D videos from the real ones. A comprehensive set of experiments has been carried out to justify the efficacy of the proposed scheme over the existing literature.

43 - 3D Interest Point Detection Based on Geometric Measures and Sparse Refinement

Xinyu Lin, University of Electronic Scien; Ce Zhu, University of Electronic Science and Technology of China; Qian Zhang, University of Electronic Science and Technology of China; Yipeng Liu, University of Electronic Science and Technology of China*

Abstract: Three dimensional (3D) interest point detection plays a fundamental role in computer vision. In this paper, we introduce a new method for detecting 3D interest points of 3D mesh models based on geometric measures and sparse refinement (GMSR). The key point of our approach is to calculate the 3D saliency measure using two novel geometric measures, which are defined in multi-scale space to effectively distinguish 3D interest points from edges and flat areas. Those points with local maxima of 3D saliency measure are selected as the candidates of 3D interest points. Finally, we utilize an L0 norm based optimization method to refine the candidates of 3D interest points by constraining the number of 3D interest points. Numerical experiments show that the proposed GMSR based 3D interest point detector outperforms current six state-of-the-art methods for different kinds of 3D mesh models.

Poster session: Audio/speech processing (4:20-5:40pm)

Chair: Douglas O'Shaughnessy, INRS-EMT

8 - Experimental Analysis of Stave Recognition of Musical Score using Projection, Hough Transform and Hit-or-miss Transforms

GenFang Chen, Hangzhou Normal University; Mei-Xia Zhu; Liang Zheng.*

Abstract: An Optical Music Recognition system comprises six stages: image pre-processing, segmentation, feature extraction, symbol recognition, musical semantics, and MIDI representation. This paper focuses on the stave recognition stage of symbol recognition, and obtains the musical information using Hough transform, projection, and mathematical morphology. It explains three major methods of stave recognition in detail, and it selects 152 musical score images to experiment for the methods. The result of experiment indicates that three methods all can detect stave, the recognition rate of Mathematical Morphology is high then other two methods, but its runtime also is long then other two methods.

118 - Assessment of sound source localization of an intra-aural audio wearable device for audio augmented reality applications

Narimene Lezzoum, Ecole de technologie supérieur; Jérémie Voix, Ecole de technologie supérieure*

Abstract: This paper presents a study on the effect of an intra-aural Audio Wearable Device (AWD) on sound source localization. The AWD used in this study is equipped with a miniature outer-ear microphone, a miniature digital signal processor and a miniature internal loudspeaker in addition to other electronics. It is aimed for audio augmented reality applications, for example to play 3D audio sounds and stimuli while keeping the wearer protected from loud or unwanted ambient noise. This AWD is evaluated in terms of ambient source localization using three localization cues computed using signals played from different positions in the horizontal and sagittal planes and recorded in the ear canals of an artificial head with and without a pair of AWDs. The localization cues are: the interaural time difference, the interaural level difference, and the head related transfer functions. Results showed that the used AWD does barely affect the localization of low frequency sounds with localization error around 2°, and only affects the localization of higher frequency sound depending on their position and frequency range.

149 - Advanced Residual Coding for MPEG Surround Encoder

Ikhwana Elfitri, Andalas University; Muhammad Sobirin; Fadlur Rahman; Rahmadi Kurnia.*

Abstract: MPEG Surround (MPS) has been widely known as both efficient technique for encoding multi-channel audio signals and rich-features audio standard. However, the generation of residual signal in the basis of a single module in MPS encoder can be considered as not optimal to compensate for error due to down-mixing process. In this paper, an improved residual coding method is proposed in order to ensure the down-mixing error can be optimally minimized particularly for MPS operation at high bit-rates. The distortion introduced during MPS encoding and decoding processes is first studied which then motivates for developing an approach which is more accurate in determining residual signals for better compensation for the distortion. A subjective test demonstrates that the MPS with improved residual coding can be competitive to Advanced Audio Coding (AAC) multi-channel for encoding 5-channel audio signals at bit-rates of 256 and 320 kb/s.

53 - Two-layer Large-scale Cover Song Identification System Based on Music Structure Segmentation

Kang Cai, Institute of Computer Science ; Deshun Yang, Institute of Computer Science & Technology, Peking University, Beijing, China; Xiaoou Chen, Institute of Computer Science & Technology, Peking University, Beijing, China*

Abstract: This paper focuses on cover song identification over a large-scale dataset. Identifying all covers of a query song from music collection is a challenging task since covers vary in multiple aspects, such as tempo, key, and structure. For the large-scale dataset, cover song identification is more challenging and few works have been published. Previous works usually use a single representation for a whole song, such as 2D Fourier transform and chord profile, which cannot reflect the property that covers are largely determined by a local similarity. To address this problem, we propose a novel cover song identification method based on music structure segmentation. The proposed structural method identifies cove songs on section level instead of song level. The experimental results show that the structural method improves the mean average precision of 2D Fourier transform method from 9.5% to 12.1% on SecondHandsSong dataset. In addition, we also propose a two-layer cover song identification system to improve the efficiency.

9 - On the Enhancement of Dereverberation Algorithms Using Multiple Perceptual-Evaluation Criteria

Rafael Zambrano López, COPPE-UFRJ; Thiago de Moura Prego, ; Amaro Azevedo de Lima, ; Sergio Lima Neto, UFRJ*

Abstract: This paper describes an enhancement strategy based on several perceptual-assessment criteria for dereverberation algorithms. The complete procedure is applied to an algorithm for reverberant speech enhancement based on single-channel blind spectral subtraction. This enhancement was implemented by combining different quality measures, namely the so-called QAreverb, the speech-to-reverberation modulation energy ratio (SRMR) and the perceptual evaluation of speech quality (PESQ). Experimental results, using a 4211-signal speech database, indicate that the proposed modifications can improve the word error rate (WER) of speech recognition systems an average of 20%.

41 - A study of the perceptual relevance of the burst phase of stop consonants with implications in speech coding

Vincent Santini, Université de Sherbrooke; Philippe Gournay, Université de Sherbrooke; Roch Lefebvre, Université de Sherbrooke*

Abstract: Stop consonants are an important constituent of the speech signal. They contribute significantly to its intelligibility and subjective quality. However, because of their dynamic and unpredictable nature, they tend to be difficult to encode using conventional approaches such as linear predictive coding and transform coding. This paper presents a system to detect, segment, and modify stop consonants in a speech signal. This system is then used to assess the following hypothesis: Muting the burst phase of stop consonants has a negligible impact on the subjective quality of speech. The muting operation is implemented and its impact on subjective quality is evaluated on a database of speech signals. The results show that this apparently drastic alteration has in reality very little perceptual impact. The implications for speech coding are then discussed.

21 - A Quantitative Real Time Data Analysis in Vehicular Speech Environment with Varying SNR

Sai Gadde, Lawrence Technological University; Mahdi Ali, Hyundai Motor Company; Philip Olivier, Lawrence Technological University; Rakan Chabaan, Hyundai America Technical Center, Inc.; Scott Bone, Hyundai America Technical Center, Inc.; Sam Tabaja, itsystems; Nabih Jaber, Lawrence Technological University*

Abstract: The purpose of this paper is to compare the performance of two common filters operating on noisy speech recorded in automobiles travelling at various speeds. The filters are based on Spectral Subtraction (SS) and Kalman Filtering (KF). The literature contains studies based on simulated data whereas this paper uses real time data collected in car's in search of an optimal solution. The comparisons were based on real recorded samples containing noisy speech signals with durations of approximately 2 minutes each. Different cases of noise levels which represent the most common situations experienced by drivers were created. The different settings used include varying car speeds (e.g., 40 mph, 70 mph), varying fan power, and window positions settings. The study was carried out using three different car models. The measured noisy voice signals were filtered using the different filtering techniques and the resulting filtered signals were compared in the time domain and the frequency domain, both quantitatively and psychometrically. Furthermore, the quantitative analysis approach was applied to the results for more accurate interpretation. Results show that SS outperforms KF in noise reduction, and with much less speech distortion at the different Signal to Noise Ratios (SNRs) tested. The audio test results subjected to human listening are comparable with the simulation results. Overall, SS showed superior performance over KF in vehicular hands-free speech applications.

54 - Robust Sound Event Classification by Using Denoising Autoencoder

Jianchao Zhou, Peking University; Liqun Peng, Peking University; Xiaoou Chen, Institute of Computer Science & Technology, Peking University, Beijing, China; Deshun Yang, Institute of Computer Science & Technology, Peking University, Beijing, China*

Abstract: Over the last decade, a lot of research has been done on sound event classification. But a main problem with sound event classification is that the performance sharply degrades in the presence of noise. As spectrogram-based image features and denoising autoencoder reportedly have superior performance in noisy conditions, this paper proposes a new robust feature called denoising autoencoder image feature (DIF) for sound event classification which is an image feature extracted from an image-like representation produced by denoising autoencoder. Performance of the feature is evaluated by a classification experiment using a SVM classifier on audio examples with different noise levels, and compared with that of baseline features including mel-frequency cepstral coefficients (MFCC) and spectrogram image feature (SIF). The proposed DIF demonstrates better performance under noise-corrupted conditions.

171 - Novel Affective Features for Multiscale Prediction of Emotion in Music

Naveen Kumar, University of Southern Califor; Tanaya Guha, IIT Kanpur; Colin Vaz, University of Southern California; Che-Wei Huang, Univ. of Southern California; Shrikanth Narayanan, University of Southern California*

Abstract: The majority of computational work on emotion in music concentrates on developing machine learning methodologies to build new, more accurate prediction systems, and usually relies on generic acoustic features. Relatively less effort has been put to the development and analysis of features that are particularly suited for the task. The contribution of this paper is twofold. First, the paper proposes two features that can efficiently capture the emotion-related properties in music. These features are named compressibility and sparse spectral components. These features are designed to capture the overall affective characteristics of music (global features). We demonstrate that they can predict emotional dimensions (arousal and valence) with high accuracy as compared to generic audio features. Secondly, we investigate the relationship between the proposed features and the dynamic variation in the emotion ratings. To this end, we propose a novel Haar transform-based technique to predict dynamic emotion ratings using only global features.

95 - On the applicability of the SBC codec to support super-wideband speech in Bluetooth handsfree communications

Nathan Souviraà-Labastie, Orange Labs; Stéphane Ragot, Orange Labs*

Abstract: With the recent standardization of the Enhanced Voice Services (EVS) codec in 3GPP, mobile operators can upgrade their voice services to offer super-wideband (SWB) audio quality (with 32kHz sampling rate). There is however one important use case which is currently limited by existing standards: handsfree communication with wireless headsets, car kits, or connected audio devices often rely on Bluetooth, and the handsfree-profile (HFP) in Bluetooth is currently limited to narrowband and wideband speech. Following the approach used to extend HFP to support wideband, we study in this paper the applicability of the SBC codec to further extend HFP to SWB. An evaluation of performance is provided taking into account Bluetooth system constraints.

164 - Improved sparse component analysis based on ant K-means clustering for underdetermined blind source separation

Shuang Wei, Hohai University; Xinchao Peng, Hohai University; Chungui Tao, Hohai University; Feng Wang, ; Defu Jiang, Hohai University*

Abstract: This paper proposes an improved sparse component analysis (SCA) approach to improve the performance of underdetermined blind source separation for the acoustic/speech sources. First, the pre-processing to build a sparse model for acoustic/speech sources is described. Then, the proposed SCA approach designs an improved K-means clustering algorithm based on ant colony algorithm to estimate the mixture matrix, and utilize a method based on orthogonal matching pursuit algorithm to recover the signals. Experiment results demonstrate that the improved clustering algorithm can enhance the global searching ability which benefits for the proposed SCA approach to achieve a higher estimation accuracy of mixture matrix, and the improved method in the second step can make the proposed SCA approach work well for recovering the multi-channel blind source signals.

99 - Audiovisual Quality Study For Videotelephony On Ip Networks

Ines Saidi, Orange Labs; Lu Zhang, INSA Rennes; Vincent Barriac, Orange Labs; Olivier Deforges, INSA Rennes*

Abstract: In this paper, an audiovisual quality assessment experiment was conducted on audiovisual clips collected using a PC-based video telephony application connected via a local IP network. The analyses of experimental results provided a better understanding of the influence of network impairments (packet loss, jitter, delay) on perceived audio and video qualities, as well as their interaction effect on the overall audiovisual quality in video telephony applications. We updated the human perception acceptability limits of audio-video synchronization for video telephony. Further, we investigated the contribution of this synchronization to the audiovisual quality independently and accompanied with network impairments. Finally, we proposed an integration model to estimate the audiovisual quality in the studied context.

Friday, Sept. 23, 2016

Oral (Special) Session: Multimodal Interaction with Digital Information in Smart Cities (10:00-11:20am)

Chair: Abed El Saddik, UOttawa

110 - Human Gesture Recognition via Bag of Angles for 3D Virtual City Planning in CAVE Environment

Nour Eldin Elmadany, Ryerson University; Yifeng He, Ryerson university; Ling Guan, Ryerson*

Abstract: Automatic Virtual Environment (CAVE) provides an immersive virtual enviroment for 3D city planning. However, the user in the cave has to wear wearable markers for the interaction with the 3D models. To develop a natural interaction, depth camera like Kinect has to be considered. In this paper, we propose a new skeleton joint representation called Bag of Angles (BoA) for human gesture recognition. We evaluated our proposed BoA representation on two dataset UTD-MHAD and UTD-MHAD-KinectV2. The evaluation results demonstrated that the proposed BoA representation can achieve a higher recognition accuracy compared to the other existing representation methods.

83 - An Intelligent Floor Surface for Foot-based Exploration of Geospatial Data

Naoto Hieda, McGill University; Jan Anlauff, McGill University; Severin Smith, ; Yon Visell, University of California, Santa Barbara; Jeremy Cooperstock, McGill*

Abstract: The combination of an expanding set of geospatial databases and continued advances in sensing technologies provide us with a wealth of data of relevance to smart cities, including weather and traffic, panoramic imagery (Google Street View), real-time video (from fixed surveillance cameras and citizen smartphone devices), and high-resolution 3D point cloud reconstruction of city infrastructure (LiDAR). Modern display technologies, such as a CAVE-like environment or a head-mounted display, allow for immersive rendering of geospatial data. In such displays, we consider the potential for navigating and exploring geospatial data through the use of natural gestures. We focus especially on foot gestures since hands are often occupied for other controls. For this purpose, we use an intelligent floor interface with force sensors and actuators that support two levels of interaction: foot-gesture navigation and immersive exploration which renders information related to the environment through visual, auditory and haptic feedback. This is built on an existing game engine framework to allow rapid prototyping of a multimodal, immersive environment, such as integration of smart city sensor data to provide visual and haptic cues.

56 - MUDVA: A Multi-Sensory Dataset for the Vehicular CPS Applications

Kazi Masudul Alam, University of Ottawa; Mohammad Hariz, University of Ottawa; Seyed Vahid Hosseini, University of Ottawa; Mukesh Saini, University of Ottawa; Abed El Saddik, UOttawa*

Abstract: Vehicular Cyber-Physical System (VCPS) is a new trend in the research of the intelligent transport systems (ITS). In VCPS, vehicles work as a hub of sensors to collect interior and exterior information about the vehicle. Vehicles can use ad-hoc networking or 3G/LTE communication technology to share useful information with their neighboring vehicles or with the infrastructures to accomplish user safety, comfort, and entertainment tasks. In order to facilitate efficient sensor-services fusion in the VCPS applications, we need real life vehicular sensory datasets. While there has been many datasets containing vehicle mobility traces, there is hardly any that contains sensory information to be shared on the network. In this paper, we present a scenario specific modular dataset architecture along with some multi-sensory dataset modules. One of the dataset modules provides time synchronized multi-vehicle data including multi-view video, multi-directional sound, GPS, accelerometer, gyroscope, and magnetic field sensors. Each of the three vehicles recorded front, back, left, and right videos while moving closely in the suburban areas to let explore vehicular cooperative applications. Another module presents necessary tools and datasets to identify vehicular events such as acceleration, deceleration, turn, and no-turn events. We also present development details of a safety application using the presented datasets along with a list of other possible applications.

98 - Comparison of Feature-level and Kernel-level Data Fusion Methods in Multi-Sensory Fall Detection

Che-Wei Huang, Univ. of Southern California; Shrikanth Narayanan, University of Southern California*

Abstract: In this work, we studied the problem of fall detection using signals from tri-axial wearable sensors. In particular, we focused on the comparison of methods to combine signals from multiple tri-axial accelerometers which were attached to different body parts in order to recognize human activities. To improve the detection rate while maintaining a low false alarm rate, previous studies developed detection algorithms by cascading base algorithms and experimented on each sensory data separately. Rather than combining base algorithms, we explored the combination of multiple data sources. Based on the hypothesis that these sensor signals should provide complementary information to the characterization of human's physical activities, we benchmarked a feature level and a kernel level fusions to learn the kernel that incorporates multiple sensors in the support vector classifier. The results show that given the same false alarm rate constraint, the detection rate improves when using signals from multiple sensors, compared to the baseline where no fusion was employed.

Poster session: Applications: Human (face, skin, body, neural movement) analysis (11:20-12:40pm)

Chair: Tiago H. Falk, INRS-EMT

6 - Hierarchical Differential Image Filters for Skin Analysis

*Parham Aarabi *, ModiFace Inc. ; Jingyi Zhang , ModiFace Inc.*

Abstract: In this paper we present a framework for analyzing skin parameters from portrait images and videos. Using a series of Hierarchical Differential Image Filters (HDIF), it becomes possible to detect different skin features such as wrinkles, spots, and roughness. These detected features are used to compute skin ratings that are compared to actual ratings by dermatologists. Analyzing a database of 49 images with ratings by a panel of dermatologists, the proposed HDIF method is able to detect skin roughness, dark spots, and deep wrinkles with an average rating error of 11.3%, 17.6%, and 15.6%, respectively, as compared to individual dermatologist rating errors of 8.2%, 7.4%, and 6.5%. Although dermatologist ratings are more accurate than the proposed HDIF method, the ratings are close enough that the HDIF ratings can be a viable solution where dermatologist ratings are not readily available.

48 - Discrimination between Diabetic Macular Edema and Normal Retinal OCT B-scan Images based on Convolutional Neural Networks

Reza Rasti, Isfahan University of Medical Sciences; Hossein Rabbani, Isfahan University of Medical Sciences; Alireza Mehri, Isfahan University of Medical Sciences; Rahele Kafieh, Isfahan University of Medical Sciences*

Abstract: This paper presents potential of a new fully automated approach for detection of cases suffering from diabetic macular edema (DME) via optical coherence tomography imaging. The proposed method is independent from any segmentation stage. The algorithm utilizes a simplified Convolutional Neural Network (CNN) model as an adaptive feature extraction and classification method. In this study, two different retinal OCT datasets are used for evaluation of the method based on Cross-Validation approach. The first set constitutes 10 normal subjects and 6 patients with diabetic macular edema (DME) from Topcon device. The second set consists 15 normal subjects and 15 patients with DME from Heidelberg device. After applying a block matching and 3D-filtering (BM3D) denoising method on overall OCT b-scans, proposed classifier has the accuracy of 82.21% (AUC = 0.86) on dataset1 and 84.23% (AUC = 0.88) on dataset2 for 2D B-scan images diagnosis according to Normal and DME slices.

58 - Multiple Human Detection in Depth Images

Muhammad Hassan Khan, University of Siegen; Kimiaki Shirahama, University of Siegen, Germany; Marcin Grzegorzek, University of Siegen, Germany; Muhammad Shahid Farid, Univesity of the Punjab, Pakistan*

Abstract: Most human detection algorithms in depth images perform well in detecting and tracking the movements of a single human object. However, their performance is rather poor when the person is occluded by other objects or when there are multiple humans present in the scene. In this paper, we propose a novel human detection technique which analyzes the edges in depth image to detect multiple people. The proposed technique detects a human head through a fast template matching algorithm and verifies it through a 3D model fitting technique. The entire human body is extracted from the image by using a simple segmentation scheme comprising a few morphological operators. Our experimental results on three large human detection datasets and the comparison with the state-of-the-art method showed an excellent performance achieving a detection rate of 94.53% with a small false alarm of 0.82%.

36 - Learning Patch-Based Anchors for Face Hallucination

Wei-Jen Ko, National Taiwan University; Y.-C. Frank Wang, Academia Sinica; Shao-Yi Chien, National Taiwan University*

Abstract: With the goal of increasing the resolution of face images, recent face hallucination methods advance learning techniques which observe training low and high-resolution patches for recovering the output image of interest. Since most existing patch-based face hallucination approaches do not consider the location information of the patches to be hallucinated, the resulting performance might be limited. In this paper, we propose an anchored patch-based hallucination method, which is able to exploit and identify image patches exhibiting structurally and spatially similar information. With these representative anchors observed, improved performance and computation efficiency can be achieved. Experimental results demonstrate that our proposed method achieves satisfactory performance and performs favorably against recent face hallucination approaches.

169 - Laplacian One Class Extreme Learning Machines for Human Action Recognition

Vasileios Mygdalis, Aristotle University of Thessa; Alexandros Iosifidis, Tampere University of Technology; Anastasios Tefas, AUTH; Ioannis Pitas, University of Thessaloniki*

Abstract: A novel OCC method for human action recognition namely the Laplacian One Class Extreme Learning Machines is presented. The proposed method exploits local geometric data information within the OC-ELM optimization process. It is shown that emphasizing on preserving the local geometry of the data leads to a regularized solution, which models the target class more efficiently than the standard OC-ELM algorithm. The proposed method is extended to operate in feature spaces determined by the network hidden layer outputs, as well as in ELM spaces of arbitrary dimensions. Its superior performance against other OCC options is consistent among five publicly available human action recognition datasets.

64 - Facial Expression Recognition with Dynamic Gabor Volume Feature

Junkai Chen, Hong Kong PolyU; Zheru Chi; Hong Fu.*

Abstract: Facial expression recognition is a long standing problem in affective computing community. A key step is extracting effective features from face images. Gabor filters have been widely used for this purpose. However, a big challenge for Gabor filters is its high dimensionality. In this paper, we propose an efficient feature called dynamic Gabor volume feature (DGVF) based on Gabor filters while with a lower dimensionality for facial expression recognition. In our approach, we first apply Gabor filters with multi-scale and multi-orientation to extract different Gabor faces. And these Gabor faces are arranged into a 3-D volume and Histograms of Oriented Gradients from Three Orthogonal Planes (HOG-TOP) are further employed to encode the 3-D volume in a compact way. Finally, SVM is trained to perform the classification. The experiments conducted on the Extended Cohn-Kanade (CK+) Dataset show that the proposed DGVF is robust to capture and represent the facial appearance features. And our method also achieves a superior performance compared with the other state-of-the-art methods.

80 - Face Video Based Touchless Blood Pressure and Heart Rate Estimation

Monika Jain, IIIT - Delhi; Sujay Deb, Indraprastha Institute of Information Technology, Delhi (IIIT-D); A Subramanyam, Indraprastha Institute of Information Technology, Delhi (IIIT-D)*

Abstract: Hypertension (high blood pressure) is the leading cause for increasing number of premature deaths due to cardiovascular diseases. Continuous hypertension screening seems to be a promising approach in order to take appropriate steps to alleviate hypertension-related diseases. Many studies have shown that physiological signal like Photoplethysmogram (PPG) can be reliably used for predicting the Blood Pressure (BP) and Heart Rate (HR). However, the existing approaches use a transmission or reflective type wearable sensor to collect the PPG signal. These sensors are bulky and mostly require an assistance of a trained medical practitioner; which preclude these approaches from continuous BP monitoring outside the medical centers. In this paper, we propose a novel touchless approach that predicts BP and HR using the face video based PPG. Since the facial video can easily be captured using a consumer grade camera, this approach is a convenient way for continuous hypertension monitoring outside the medical centers. The approach is validated using the face video data collected in our lab, with the ground truth BP and HR measured using a clinically approved BP monitor – OMRON HBP1300. Accuracy of the method is measured in terms of normalized mean square error, mean absolute error and error standard deviation; which complies with the standards mentioned by Association for the Advancement of Medical Instrumentation. Two-tailed dependent sample t-test is also conducted to verify that there is no statistically significant difference between the BP and HR predicted using the proposed approach and the BP and HR measured using OMRON.

154 - Global Anomaly Detection in Crowded Scenes based on Optical Flow Saliency

Ang Li, Beijing Jiaotong University; Yigang Cen; Xiao-Ping Zhang; Li-Hong Cui; Zhenjiang Miao.

Abstract: In this paper, an algorithm of global anomaly detection in crowded scenes using the saliency in optical flow field is proposed. First, the scale invariant feature transforms (SIFT) method is utilized to get the saliency map of optical flow field. Then, the histogram of maximal optical flow projection (HMOFP) is extracted. On the basis of the HMOFP feature of normal frames, an online dictionary learning algorithm is used to train an optimal dictionary with proper redundancy after selecting the training samples. This dictionary is better than the dictionary simply composed by the HMOFP feature of the whole training frames. We use the l1-norm of the sparse reconstruction coefficients (i.e., the sparse reconstruction cost, SRC) to detect the anomaly of the testing frame. The experiment results on UMN dataset and the comparison to the state-of-the-art methods show that our algorithm is superior and effective.

101 - Fast Circlet Based Framework For Optic Disk Detection

Omid Sarrafzadeh, Isfahan University of Medical Sciences; Hossein Rabbani, Isfahan University of Medical Sciences; Alireza Mehri Dehnavi.*

Abstract: Optic Disc (OD) detection in Retinal Images (RI) of fundus is crucial to automate a screening system for diabetic reti-nopathy. Most researches for automatic localization of OD benefit the regions of vessels. We present in this paper a fast and novel method based on the Circlet Transform (CT) to detect OD in digital retinal fundus images that doesn't need the location of the vessels. First, each R, G and B band is enhanced using Contrast Limited Adaptive Histogram Equalization. Then, the enhanced image is converted from RGB to L*a*b color space. Next, the CT is applied on the L* band and finally the circlelet coefficients are analyzed to find the location of the OD. The proposed algorithm is implemented on DRIVE dataset and the experimental results show a very well OD localization. The average processing time of running algorithm for an image (565×584 pixels) is 1.56 seconds.

42 - Laughter Detection based on the Fusion of Local Binary Patterns, Spectral and Prosodic Features

Stefany Bedoya, INRS ; Tiago Falk, INRS-EMT*

Abstract: Today, great focus has been placed on context-aware human-machine interaction, where systems are aware not only of the surrounding environment, but also about the mental/affective state of the user. Such knowledge can allow for the interaction to become more human-like. To this end, automatic discrimination between laughter and speech has emerged as an interesting, yet challenging problem. Typically, audio- or video-based methods have been proposed in the literature; humans, however, are known to integrate both sensory modalities during conversation and/or interaction. As such, this paper explores the fusion of support vector machine classifiers trained on local binary pattern (LBP) video features, as well as speech spectral and prosodic features as a way of improving laughter detection performance. Experimental results on the publicly-available MAHNOB

Laughter database show that the proposed audio visual fusion scheme can achieve a laughter detection accuracy of 93.3%, thus outperforming systems trained on audio or visual features alone.

93 - Robust MRI Reconstruction Via Re-Weighted Total Variation And Non-Local Sparse Regression

Mingli Zhang, Ecole de Technologie Superieure; Christian Desrosiers, École de Technologie Supérieure*

Abstract: Total variation (TV) based sparsity and non-local self-similarity have been shown to be powerful tools for the reconstruction of magnetic resonance (MR) images. However, due to the uniform regularization of gradient sparsity, standard TV approaches often over-smooth edges in the image, resulting in the loss of important details. This paper presents a novel compressed sensing method for the reconstruction of MRI data, which uses a regularization strategy based on re-weighted TV to preserve image edges. This method also leverages the redundancy of non-local image patches through the use of a sparse regression model. An efficient strategy based on the Alternating Direction Method of Multipliers (ADMM) algorithm is used to recover images with the proposed model. Experimental results on a simulated phantom and real brain MR data show our method to outperform state-of-the-art compressed sensing approaches, by better preserving edges and removing artifacts in the image.

163 - Magnetic resonance image classification using nonnegative matrix factorization and ensemble tree learning techniques

Javier Ramirez, UGR; Juan Gorriz, University of Granada; Francisco Martínez-Murcia, University of Granada; Fermín Segovia, University of Granada; Diego Salas-Gonzalez, University of Granada*

Abstract: This paper shows a magnetic resonance image (MRI) classification technique based on nonnegative matrix factorization (NNMF) and ensemble tree learning methods. The system consists of a feature extraction process that applies NNMF to gray matter (GM) MRI first-order statistics of a number of subcortical structures and a learning process of an ensemble of decision trees. The ensembles are trained by means of boosting and bagging while their performance is compared in terms of the classification error and the received operating characteristics curve (ROC) using k-fold cross validation. The results show that NNMF is well suited for reducing the dimensionality of the input data without a penalty on the performance of the ensembles. The best performance was obtained by bagging in terms of convergence rate and minimum residual loss, especially for high complexity classification tasks (i.e. NC vs. MCI and MCI vs. AD).

Poster session: Video processing, applications, and dimensionality reduction (1:40-3:00pm)

Chair: Enrico Magli, Politecnico de Torino

30 - Mobile Live Streaming: Insights from the Periscope Service

Leonardo Favario, Politecnico di Torino; Matti Siekkinen, Aalto University; Enrico Masala, Politecnico di Torino*

Abstract: Live video streaming from mobile devices is quickly becoming popular through services such as Periscope, Meerkat, and Facebook Live. Little is known, however, about how such services tackle the challenges of the live mobile streaming scenario. This work addresses such gap by investigating in details the characteristics of the Periscope service. A large number of publicly available streams have been captured and analyzed in depth, in particular studying the characteristics of the encoded streams and the communication evolution over time. Such an investigation allows to get an insight into key performance parameters such as bandwidth, latency, buffer levels and freezes, as well as the limits and strategies adopted by Periscope to deal with this challenging application scenario.

111 - A Simple Approach Towards Efficient Partial-Copy Video Detection

Zobeida Guzman-Zavaleta, INAOE; Claudia Feregrino-Uribe, INAOE*

Abstract: Video copy detection is still an open problem as current approaches are not able to carry out the detection with enough efficacy and efficiency. These are desirable features in modern video-based applications requiring real-time processing in large scale video databases and without compromising detection performance, especially when facing non-simulated video attacks. These characteristics are also desirable in partial-copy detection, where the detection challenges increase when the video query contains short segments corresponding to a copied video, this is, partial-copies at frame level. Motivated by these issues, in this work we propose a video fingerprinting approach based on the extraction of a set of low-cost and independent binary global and local fingerprints. We tested our approach with a video dataset of real-copies and the results show that our method outperforms robust state-of-the-art methods in terms of detection scores and computational efficiency. The latter is achieved by processing only short segments of 1 second length, which takes a processing time of 44 ms.

162 - Movie Shot Selection Preserving Narrative Properties

Ioannis Mademlis, Aristotle University of Thessa; Anastasios Tefas, AUTH; Nikos Nikolaidis; Ioannis Pitas, University of Thessaloniki*

Abstract: Automatic shot selection is an important aspect of movie summarization that is helpful both to producers and to audiences, e.g., for market promotion or browsing purposes. However, most of the related research has focused on shot selection based on low-level video content, which disregards semantic information, or on narrative properties extracted from text, which requires the movie script to be available. In this work, semantic shot selection based on the narrative prominence of movie characters in both the visual and the audio modalities is investigated, without the need for additional data such as a script. The output is a movie summary that only contains video frames from selected movie shots. Selection is controlled by a user-provided shot retention parameter, that removes key-frames/key-segments from the skim based on actor face appearances and speech instances. This novel process (Multimodal Shot Pruning, or MSP) is algebraically modelled as a multimodal matrix Column Subset Selection Problem, which is solved using an evolutionary computing approach.

123 - Two-Way Real Time Multimedia Stream Authentication Using Physical Unclonable Functions

Mehrdad Zaker Shahrak, University of Nebraska-Lincoln; Mengmei Ye, University of Nebraska-Lincoln; Viswanathan Swaminathan, Adobe; Sheng Wei, University of Nebraska-Lincoln*

Abstract: Multimedia authentication is an integral part of multimedia signal processing in many real-time and security sensitive applications, such as video surveillance. In such applications, a full-fledged video digital rights management (DRM) mechanism is not applicable due to the real time requirement and the difficulties in incorporating complicated license/key management strategies. This paper investigates the potential of multimedia authentication from a brand new angle by employing hardware-based security primitives, such as physical unclonable functions (PUFs). We show that the hardware security approach is not only capable of accomplishing the authentication for both the hardware device and the multimedia stream but, more importantly, introduce minimum performance, resource, and power overhead. We justify our approach using a prototype PUF implementation on Xilinx FPGA boards. Our experimental results on the real hardware demonstrate the high security and low overhead in multimedia authentication obtained by using hardware security approaches.

38 - Automatic Camera Self-Calibration for Immersive Navigation of Free Viewpoint Sports Video

Qiang Yao; Hiroshi Sankoh; Keisuke Nonaka; Sei Naito - KDDI R&D Laboratories, Inc.*

Abstract: In recent years, the demand of immersive experience has triggered a great revolution in the applications and formats of multimedia. Particularly, immersive navigation of free viewpoint sports video has become increasingly popular, and people would like to be able to actively select different viewpoints when watching sports videos to enhance the ultra realistic experience. In the practical realization of immersive navigation of free viewpoint video, the camera calibration is of vital importance. Especially, automatic camera calibration is very significant in real-time implementation and the accuracy of camera parameter directly determines the final experience of free viewpoint navigation. In this paper, we propose an automatic camera self-calibration method based on a field model for free viewpoint navigation in sports events. The proposed method is composed of three parts, namely, extraction of field lines in a camera image, calculation of crossing points, determination of the

optimal camera parameter. Experimental results show that the camera parameter can be automatically estimated by the proposed method for a fixed camera, dynamic camera and multi-view cameras with high accuracy. Furthermore, immersive free viewpoint navigation in sports events can also be completely realized based on the camera parameter estimated by the proposed method.

55 - Video Temporal Super-Resolution Using Nonlocal Registration and Self-Similarity

Matteo Maggioni; Pier Luigi Dragotti - Imperial College London*

Abstract: In this paper we present a novel temporal super-resolution method for increasing the frame-rate of single videos. The proposed algorithm is based on motion-compensated 3-D patches, i.e., a sequence of 2-D blocks following a given motion trajectory. The trajectories are computed through a coarse-to-fine motion estimation strategy embedding a regularized block-wise distance metric that takes into account the coherence of neighbouring motion vectors. Our algorithm comprises two stages. In the first stage, a nonlocal search procedure is used to find a set of 3-D patches (targets) similar to a given patch (reference), subsequently all targets are registered at sub-pixel precision with respect to the reference in an upsampled 3-D FFT domain, and finally all registered patches are aggregated at their appropriate locations in the high-resolution video. The second stage is used to further improve the estimation quality by correcting each 3-D patch of the video obtained from the first stage with a linear operator learned from the self-similarity of patches at a lower temporal scale. Our experimental evaluation on color videos shows that the proposed approach achieves high quality super-resolution results from both an objective and subjective point of view.

84 - Color-guided Depth Refinement Based on Edge Alignment

Hu Tian, Fujitsu R&D Center; Fei Li*

Abstract: Depth maps captured by consumer-level depth cam- eras such as Kinect usually suffer from the problem of corrupted edges and missing depth values. In this paper, an effective approach with the support of guided color images is proposed to tackle this problem. Firstly, an effective two-pass alignment algorithm is used to reliably align the depth edges with color image edges. Then, a new depth map with refined edges is gen- erated based on interpolated drift vectors. Finally, a constrained maximal bilateral filter is proposed to fill the holes. Compared with existing methods, our approach can better refine the depth edges and avoid blurred depths in areas of depth discontinuities, as demonstrated by experiments on real Kinect data.

119 - Adaptive Enhancement Filtering for Motion Compensation

Xiaoyu Xiu, InterDigital Communications; Yuwen He, InterDigital Communications; Yan Ye, InterDigital Communications*

Abstract: This paper proposes an enhanced motion compensated prediction algorithm for hybrid video coding. The algorithm is built upon the concept of applying adaptive enhancement filtering at the motion compensation stage. For luma, a high-pass filter is applied to the motion compensated prediction signal to recover distorted high-frequency information. For chroma, cross-plane filters are applied to enhance the motion compensated signals by restoring the blurred edges and textures of the chroma planes using the high-frequency information of the luma plane. To verify the effectiveness, the proposed algorithm is implemented on the HM Key Technology Area (HM-KTA) 1.0 platform. Experimental results show that compared to the anchor, the proposed algorithm achieves average Bjøntegaard delta (BD) rate savings of 0.4%, 8.8% and 7.4% for Y, Cb and Cr components, respectively.

14 - Temporally Consistent High Frame-Rate Upsampling with Motion Sparsification

Dominic Ruefenacht, UNSW; David Taubman, UNSW*

Abstract: This paper continues our work on occlusion-aware temporal frame interpolation (TFI) that employs piecewise-smooth motion with sharp motion boundaries. In this work, we propose a triangular mesh sparsification algorithm, which allows to trade off computational complexity with reconstruction quality. Furthermore, we propose a method to create a background motion layer in regions that get disoccluded between the two reference frames, which is used to get temporally consistent interpolations among frames interpolated between the two reference frames. Experimental results on a large data set show the proposed mesh sparsification is able to reduce the processing time by 75%, with a minor drop in PSNR of 0.02 dB. The proposed TFI scheme outperforms various state-of-the-art TFI methods in terms of quality of the interpolated frames, while having the lowest processing times. Further experiments on challenging synthetic sequences highlight the temporal consistency in traditionally difficult regions of disocclusion.

85 - Generalized Dirichlet Mixture Matching Projection for Supervised Linear Dimensionality Reduction of Proportional Data

Walid Masoudimansour, Concordia University; Nizar Bouguila, Concordia University*

Abstract: In this paper, a novel effective method to reduce the dimensionality of labeled proportional data is introduced. Most well-known existing linear dimensionality reduction methods rely on solving the generalized eigenvalue problem which fails in certain cases such as sparse data. The proposed algorithm is a linear method and uses a novel approach to the problem of dimensionality reduction to solve this problem while resulting higher classification rates. Data is assumed to be from two different classes where each class is matched to a mixture of generalized Dirichlet distributions after projection. Jeffrey divergence is then used as a dissimilarity measure between the projected classes to increase the inter-class variance. To find the optimal projection that yields the largest mutual information, genetic algorithm is used. The method is especially designed as a preprocessing step for binary classification, however, it can handle multi-modal data effectively due to the use of mixture models and therefore can be used for multi-class problems as well.

49 - Low-power distributed sparse recovery testbed on wireless sensor networks

Riccardo De Lucia, Politecnico di Torino; Sophie Fosson, Politecnico di Torino; Enrico Magli, POLITO*

Abstract: Recently, distributed algorithms have been proposed for the recovery of sparse signals in networked systems, e.g. wireless sensor networks. Such algorithms allow large networks to operate autonomously without the need of a fusion center, and are very appealing for smart sensing problems employing low-power devices. They exploit local communications, where each node of the network updates its estimates of the sensed signal also based on the correlated information received from neighboring nodes. In the literature, theoretical results and numerical simulations have been presented to prove convergence of such methods to accurate estimates. Their implementation, however, raises some concerns in terms of power consumption due to iterative inter-node communications, data storage, computation capabilities, global synchronization, and faulty communications. On the other hand, despite these potential issues, practical implementations on real sensor networks have not been demonstrated yet. In this paper we fill this gap and describe a successful implementation of a class of randomized, distributed algorithms on a real low-power wireless sensor network testbed with very scarce computational capabilities. We consider a distributed compressed sensing problem and we show how to cope with the issues mentioned above. Our tests on synthetic and real signals show that distributed compressed sensing can successfully operate in a real-world environment.

Oral session: Multimedia Communication and Networking (3:00-4:20 pm)

Chair: TBD

116 - Delay-rate-distortion Optimization for Cloud-based Collaborative Rendering

Xiaoming Nan, Ryerson University; Yifeng He, Ryerson university; Ling Guan, Ryerson*

Abstract: Cloud rendering is emerged as a new cloud service to satisfy user's desire for running sophisticated graphics applications on thin devices. However, traditional cloud rendering approaches, both remote rendering and local rendering, have limitations. Remote rendering shifts intensive rendering tasks to cloud server and streams rendered frames to client, which suffers from high delay and bandwidth usage. Local rendering sends graphics data to client and performs rendering on local devices, which requires initial buffering delay and demands high computation capacity at client. In this paper, we propose a novel cloud based collaborative rendering framework, which integrates remote rendering and local rendering. Based on the proposed framework, we study the delay-Rate-Distortion (d-R-D) optimization problem, in which the source rates are optimally allocated for streaming encoded video frames and graphics data to minimize the overall distortion under the bandwidth and response delay constraints. Experiment results demonstrate that the proposed collaborative rendering framework can effectively allocate source rates to achieve the minimal distortion compared to the traditional remote rendering and local rendering.

117 - Joint Optimization of Resource Allocation and Workload Scheduling for Cloud based Multimedia Services

Xiaoming Nan, Ryerson University; Yifeng He, Ryerson university; Ling Guan, Ryerson*

Abstract: With the development of cloud technology, cloud computing has been increasingly used as distributed platforms for multimedia services. However there are two fundamental challenges for service providers: one is resource allocation, and the other is workload scheduling. Due to the rapidly varying workload and strict response time requirement, it is difficult to optimally allocate virtual machines (VMs) and assign workload. In this paper, we study the resource allocation and workload scheduling problem for cloud based multimedia services. Specifically, we introduce a queuing model to quantify the resource demands and service performance, and a directed acyclic graph (DAG) model to characterize the precedence constraints among jobs. Based on the proposed models, we jointly optimize the allocated VMs and the assigned workload to minimize the total resource cost under the response time constraints. Since the formulated problem is mixed integer non-linear programming, a heuristic is proposed to efficiently allocate resources for practical services. Experimental results show that the proposed scheme can effectively allocate VMs and schedule workload to achieve the minimal resource cost.

46 - Novel UEP Product Code Scheme with Protograph-based Linear Permutation and Iterative Decoding for Scalable Image Transmission

Huihui Wu, McMaster University; Sorina Dumitrescu, Department of Electrical and Computer Engineering, McMaster University*

Abstract: This paper introduces a linear permutation module before the inner encoder of the iteratively decoded product coding structure, for the transmission of scalable bitstreams over error-prone channels. This can improve the error correction ability of the inner code when some source bits are known from the preceding outer code decoding stages. The product code consists of a protograph low-density parity-check code (inner code) and Reed-Solomon (RS) codes of various strengths (outer code). Further, an algorithm relying on protograph-based extrinsic information transfer analysis is devised to design good base matrices from which the linear permutations are constructed. In addition, an analytical formula for the expected fidelity of the reconstructed sequence is derived and utilized in the optimization of the RS codes redundancy assignment. The experimental results reveal that the proposed approach consistently outperforms the scheme without the linear permutation module, reaching peak improvements of 1.98 dB and 1.30 dB over binary symmetric channels (BSC) and additive white Gaussian noise (AWGN) channels, respectively.

29 - Layer-Based Temporal Dependent Rate-Distortion Optimization in Random-Access Hierarchical Video Coding

Yanbo Gao, , University of Electronic Science and Technology of China; Ce Zhu, University of Electronic Science and Technology of China; Shuai Li; Tianwu Yang.*

Abstract: Rate-distortion optimization (RDO) plays an important part in improving the coding efficiency of High Efficiency Video Coding (HEVC), especially for the hierarchical coding structure defined in the Random-Access (RA) configuration, noted as Random-Access Hierarchical Video Coding (RA-HVC), where different frames are assigned to different temporal layers and further coded with different coding parameters. Due to the inter-frame prediction, coding result of one unit may affect the coding performance of the following temporally related units. Therefore, the temporal dependency among units needs to be considered in the coding process. However, the RDO process in the current video codec is performed without considering the varying temporal dependency, thus compromising the rate-distortion performance significantly. To address this problem, a layer-based temporal dependent RDO method is proposed in this paper where the temporal dependency among different frames in the same or different layers is examined. By reformulating the temporal dependent RDO for the RA-HVC, we show that it can be implemented in a way of simply refining the Lagrange multiplier. Experimental results show that the proposed method achieves, in average, about 1.4% BD-rate savings with a negligible increase in encoding time for the random-access configuration.