# Proceedings of the 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference

# (APSIPA ASC 2012)

## Hollywood, California, USA
## 3 - 6 December 2012

## Pages 1-792

IEEE Catalog Number: CFP1214U-PRT
ISBN: 978-1-4673-4863-8

# Technical Program Abstracts

## Tuesday, December 4, 2012

## 10:20 - 12:00

### OS.1-IVM.1 High Efficiency Video Coding (HEVC)

Session Chairs: Siwei Ma, Oscar Au, Jiaying Liu — Location: Doheny

**An Efficient NEON-based Quarter-pel Interpolation Method for HEVC**

Hao Lv *Peking University*, Ronggang Wang *Peking University*, Jie Wan *Peking University*, Huizhu Jia *Peking University*, Xiaodong Xie *Peking University*, Wen Gao *Peking University*

SIMD (Single Instruction Multiple Data) instructions have been widely used for digital signal processing and multimedia applications, especially video codec. This paper proposes the quarter-pel interpolation acceleration method of the HEVC (High Efficiency Video Coding), which is implemented with ARM SIMD instructions. Data level parallelism is utilized to use the SIMD capability of NEON effectively. Experiment results show that the implementation of the proposed method is approximately five times faster than that of the HEVC reference software for the HEVC quarter-pel interpolation operation.

**Efficient SIMD Optimization of HEVC Encoder over X86 Processors**

Keji Chen *Peking University*, Yizhou Duan *Peking University*, Leju Yan *Peking University*, Jun Sun *Peking University*, Zongming Guo *Peking University*

High Efficient Video Coding (HEVC) is the next generation video coding standard in progress. Based on the traditional hybrid coding framework, HEVC implements enhanced tools to improve compression efficiency at the cost of far more computational payload than the capacity of real-time video applications. In this paper, we focus on the fast implementation of the HEVC encoder over modern Intel x86 processors. First, we identify the most time-consuming modules of HM 6.2 encoder, represented by motion compensation, Hadamard transform, sum of difference (SAD/SSD) calculation and integer transform. Then the single-instruction-multiple-data (SIMD) methods are proposed to optimize the computational performance of these modules. Experimental results show that the optimized encoder achieves 56% - 85% time saving compared with the HM 6.2 encoder over Intel i5-750 processor.

**Lossy and Lossless Intra Coding Performance Evaluation: HEVC, H.264/AVC, JPEG 2000 and JPEG LS**

Qi Cai *Shanghai Jiao Tong University*, Li Song *Shanghai Jiao Tong University*, Guichun Li *Santa Clara University*, Nam Ling *Santa Clara University*

High Efficiency Video Coding (HEVC), the latest international standard of video coding under development, has shown a major breakthrough with regards to compression efficiency. But most of the currently published studies were intended to evaluate the overall R-D performance of HEVC in comparison to prior H.264/AVC video coding standard. In this paper, we present sufficient rate-distortion performance comparisons of image coding between the HEVC and previous image and intra-only video coding standards, including JPEG 2000, JPEG LS and H.264/AVC intra high profile. In addition, some recently reported performances of HEVC are also reviewed and compared. The coding simulations are conducted on a set of recommended video sequences during the development of the HEVC standard. Experimental results show that HEVC can offer consistent performance gains over a wide range of bitrates on natural video sequences as expected. Besides, we also present the comparison results of all these standards in the scenario of lossless image coding.

**An Adaptive Frame Complexity Based Rate Quantization Model for Intra-Frame Rate Control of High Efficiency Video Coding (HEVC)**

Lin Sun *HKUST*, Oscar C. Au *HKUST*, Wei Dai *HKUST*, YuanFang Guo *HKUST*, Ruobing Zou *HKUST*

An efficient and accurate R-Q model is greatly important for intra-frame rate control of the latest High Efficiency video Coding (HEVC) standard. However, previous methods pay more attention to the gradient based rate quantization (R-Q) model for the intra bit rate control. In this paper, we analyze the drawbacks of the gradient based frame complexity measure when applied different Quantization Parameters (QPs). Then we propose a novel edge based frame complexity measure using the Gaussian Gradient operator with properly selected parameters. In order to tackle the problems that the gradient based rate quantization model fails when using the different QPs, based on these two complexity measures we propose an adaptive frame complexity based R-Q model for intra bit rate control. Simulations have been conducted based on HM6.2 which is the latest reference software of HEVC. Note that we may be the first to do this work in HEVC, so we do not have the classical methods which have been implemented in HEVC to compare with. So we implement the traditional gradient based rate quantization model and the Cauchy distribution based rate quantization model in HEVC. Then we compare the bit rate mismatch ratio between their methods and our proposed method. The simulation results show that by using our proposed scheme, better bit rate estimation for intra frames can be achieved. Up to 33.1% mismatch ratio reduction compared with the Cauchy distribution based model and up to 13% mismatch ratio reduction compared with the gradient based model.

**Early Termination of Coding Unit Splitting for HEVC**

Qin Yu *Peking University*, Xinfeng Zhang *Peking University*, Siwei Ma *Peking University*, Shiqi Wang *Peking University*

The emerging high-efficiency video coding (HEVC) standard employs a new coding structure characterized by coding unit (CU), prediction unit (PU) and transform unit (TU). It improves the coding efficiency significantly, but also introduces great computation complexity on the decision of optimal CU, PU and TU sizes. To reduce the encoding complexity, we propose a CU splitting early termination scheme for inter frame coding. In the proposed scheme, the characteristics of prediction residuals are utilized to early terminate the CU splitting. Specifically, the Mean Square Error (MSE) between the prediction block and the origin block for each CU level is obtained and then compared with an adaptive threshold. The recursive CU splitting process is early terminated according to the threshold. Experimental results demonstrate that, the proposed algorithm achieves up to 34.83% total encoding time reduction with less than 0.25% BD-rate increase on average.

**Hardware Oriented Re-design and Matrix Approximation Analysis for Transform in High Efficiency Video Coding (HEVC)**

Lin Sun *HKUST*, Oscar C. Au *HKUST*, Jiali Li *HKUST*, Ruobing Zou *HKUST*, Wei Dai *HKUST*

In this paper, we propose an adaptive truncate k-bit re-configurable approximation (aTra) method which can achieve similar video coding efficiency as the original transform but substitute all low efficient multiplication by conformed shifting and addition operations, lifting the utilization of the hardware implementation and making simple hardware implementation and high efficiency pipeline design possible.

Also, we may be the first to propose three mathematical constraints for the hardware matrix approximation to make the final performance controllable. The proposed method can achieve regular data flow and massive operation reduction balancing the rate distortion (RD) performance and data throughput. Particularly for the current secondary transform, rotational transform (ROT), we obtain the hardware friendly ROT through our proposed method based on one of the constraints. The simulation results implemented in the HEVC reference software HM1.0 present the validity of our method. Our method achieves similar performance when compared with the original one but it is much better than the simple hardware implementation.

## OS.2-IVM.2 Camera-based Human Centric Computing: Technology and Applications

Session Chairs: Haowei Liu, Weiyao Lin, YingLi Tian, Yi Wu Location: Beachwood

### RGBD Camera-based Activity Analysis

Chenyang Zhang *The City College of New York*, Yingli Tian *The City College of New York*

In this paper, we propose a new activity analysis framework to facilitate the independence of elderly adults living in the community, reduce risks, and enhance the quality of life at home by using RGB-D cameras. Our contributions include two aspects: 1) recognizing 5 activities related to falling including standing, fall from standing, fall from sitting, sit on chair, and sit on floor. The main analysis is based on the depth information due to the advantages of handling illumination changes and identity protection. If the monitored person is out of the range of 3D camera, RGB-based video analysis module is employed to continue the activity monitoring. 2) Identifying the monitored person if there are multiple people in camera view by combining depth and RGB information. We have collected a dataset under different lighting conditions and ranges. Experimental results demonstrate the effectiveness of the proposal framework.

### Recognizing Object Manipulation Activities Using Depth and Visual Cues

Haowei Liu *Intel*, Matthai Philipose *Microsoft Research*, Ming-Ting Sun *University of Washington*

We present the design of an approach to recognize human activities that involve manipulating objects. Our proposed approach identifies objects being manipulated and models high-level tasks being performed accordingly. Realistic settings for such tasks pose several problems for computer vision, including sporadic occlusion by subjects, non-frontal poses, and objects with few local features. We show how size and segmentation information derived from depth data can address these challenges using simple and fast techniques. In particular, we show how to robustly and without supervision find the manipulating hand, properly detect/recognize objects and properly use the temporal information to fill in the gaps between sporadically detected objects, all through careful inclusion of depth cues. We evaluate our approach on a challenging dataset of 12 kitchen tasks that involve 24 objects performed by 2 subjects. The entire system yields 82%/84% precision (74%/83%recall) for task/object recognition. Our techniques outperform the state-of-the-art significantly in activity/object recognition rates.

### Virtual Mirror By Fusing Multiple RGB-D Cameras

Ju Shen *University of Kentucky*, Sen-ching Samson Cheung *University of Kentucky*, Jian Zhao *Microsoft*

Mirror is possibly the most common optical device in our everyday life. Rendering a virtual mirror using a joint camera-display system has a wide range of applications from cosmetics to medicine. Existing works focus primarily on simple modification of the mirror images of body parts and provide no or limited range of viewpoint dependent rendering. In this paper, we propose a framework for rendering mirror images from a virtual mirror based on 3D point clouds and color texture captured from a network of structured-light RGB-D cameras. We validate our models by comparing the results with a real mirror. Commodity structured-light cameras often have missing and erroneous depth data which directly affect the quality of the rendering. We address this problem via a novel probabilistic model that accurately separates foreground objects from background scene before correcting the erroneous depth data. We experimentally demonstrate that our depth correction algorithm outperforms other state-of-the-art techniques.

### Periodic Motion Detection With ROI-Based Similarity Measure And Extrema-Based Reference-Frame Selection

Xintong Han *Shanghai Jiao Tong University*, Gaojian Li *Fudan University*, Weiyao Lin *Shanghai Jiao Tong University*, Xiaoqiong Su *Shanghai Jiao Tong University*, Hongxiang Li *University of Louisville*, Hua Yang *Shanghai Jiao Tong University*, Hui Wei *Fudan University*

This paper presents a new algorithm for detecting and analyzing the periodic motions in video sequences. Different from the previous methods which detect periodic motions from the entire frame, we propose a convex-hull-based process to automatically determine the regions of interest (ROI) of the motions and utilize an ROI-based similarity measure to detect the motion periods. Furthermore, we also propose an extrema-based method to select the optimal reference frame for further improving the periodic detection performance. Our proposed algorithm can not only effectively detect motion periods with both constant and variable period lengths, but also have obvious advantage when handling periodic motion with slight movements. Experimental results demonstrate the effectiveness of our proposed method.

### Abnormal Crowd Behavior Detection Based on Local Pressure Model

Hua Yang *Shanghai Jiao Tong University*, Yihua Cao *Shanghai Jiao Tong University*, Shuang Wu *Shanghai Jiao Tong University*, Weiyao Lin *Shanghai Jiao Tong University*, Shibao Zheng *Shanghai Jiao Tong University*, Zhenghua Yu *Shanghai Jiao Tong University*

Abnormal crowd behavior detection is an important issue in crowd surveillance. In this paper, a novel local pressure model is proposed to detect the abnormality in large-scale crowd scene based on local crowd characteristics. These characteristics include the local density and velocity which are very significant parameters for measuring the dynamic of crowd. A gird of particles is placed over the image to reduce the computation of the local crowd parameters. Local pressure is generated by applying these local characteristics in pressure model. Histogram is utilized to extract the statistical property of the magnitude and direction of the pressure. The crowd feature vector of the whole frame is obtained through the analysis of Histogram of Oriented Pressure (HOP). SVM and median filter are then adopted to detect the anomaly. The performance of the proposed method is evaluated on publicly available datasets from UMN. The experimental results show that the proposed method can achieve a higher accuracy than that of the previous methods on detecting abnormal crowd behavior.

### Combining RGB and Depth Features for Human Activity Recognition

Yang Zhao *UESTC*, Zicheng Liu *Microsoft*, Lu Yang *UESTC*, Hong Cheng *UESTC*

We study the problem of human activity recognition from RGB-D sensors when the skeletons are not available. The skeleton tracking in Kinect SDK works well when the human subject is facing the camera and there are no occlusions. In surveillance or senior home monitoring scenarios, however, the camera is usually mounted higher than human subjects and there may be serious occlusions. Consequently, the skeleton tracking may not work well. In RGB image based activity recognition, a popular approach that can handle cluttered background and partial occlusions is the interest point based approach. When both RGB and depth channels are available, one can still use the interest point based approach. But there are questions on whether we should detect interest points from RGB channel or from depth channel, and what descriptor to use for the depth channel. The goal of this paper is to compare the performances of different ways of extracting interest points. In addition, we have developed a depth map based descriptor which outperforms the HOGHOF descriptor on the depth video. We show that the best performance is achieved when we extract interest points from RGB channel, and combine the RGB based descriptor and depth map based descriptor.

## OS.3-SLA.1 Speech Processing and Its Applications

Session Chair: Jen-Tzung Chien Location: Runyon

### Open Answer Scoring for S-CAT Automated Speaking Test System Using Support Vector Regression

Yutaka Ono *Chiba University*, Misuzu Otake *Chiba University*, Takahiro Shinozaki *Chiba University*, Ryuichi Nisimura *Wakayama University*, Takeshi Yamada *University of Tsukuba*, Kenkichi Ishizuka *University of Tsukuba*, Yasuo Horiuchi *Chiba University*, Shingo Kuroiwa *Chiba University*, Shingo Imai *University of Tsukuba*

We are developing S-CAT computer test system that will be the first automated adaptive speaking test for Japanese. The speaking ability of examinees is scored using speech processing techniques without human raters. By using computers for the scoring, it is possible to largely reduce the scoring cost and provide a convenient means for language learners to evaluate their learning status. While the S-CAT test has several categories of question items, open answer question is technically the most challenging one since examinees freely talk about a given topic or argue something for a given material. For this problem, we proposed to use support vector regression (SVR) with various features. Some of the features rely on speech recognition hypothesis and others do not. SVR is more robust than multiple regression and the best result was obtained when 390 dimensional features that combine everything were used. The correlation coefficients between human rated and SVR estimated scores were 0.878, 0.847, 0.853, and 0.872 for fluency, accuracy, content, and richness measures, respectively.

### Singing Voice Conversion Method Based on Many-to-Many Eigenvoice Conversion and Training Data Generation Using a Singing-to-Singing Synthesis System

Hironori Doi *Nara Institute of Science and Technology*, Tomoki Toda *Nara Institute of Science and Technology*, Tomoyasu Nakano *National Institute of Advanced Industrial Science and Technology*, Masataka Goto *National Institute of Advanced Industrial Science and Technology*, Satoshi Nakamura *Nara Institute of Science and Technology*

The voice quality (identity) of singing voices is usually fixed in each singer. To overcome this limitation and enable singers to freely change their voice quality using signal-processing technologies, we propose a singing voice conversion method based on many-to-many eigenvoice conversion (EVC) that can convert the voice quality of an arbitrary source singer into that of another arbitrary target singer. Previous EVC-based methods required parallel data consisting of song pairs of a single reference singer and many prestored target singers for training a voice conversion model, but it was difficult to record such data. Our proposed method therefore uses a singing-to-singing synthesis system called VocaListener to generate parallel data by imitating singing voices of many prestored target singers with the system's singing voices. Experimental results show that our method succeeded in enabling people to sing a song with the voice quality of a different target singer even if only an extremely small amount of the target singing voice is available.

### Response Generation based on Statistical Machine Translation for Speech-Oriented Guidance System

Kazuma Nishimura *Nara Institute of Science and Technology*, Hiromichi Kawanami *Nara Institute of Science and Technology*, Hiroshi Saruwatari *Nara Institute of Science and Technology*, Kiyohiro Shikano *Nara Institute of Science and Technology*

An example-based response generation is a robust and practical approach for a real-environment information guidance system. However, this framework cannot reflect differences in nuance, because the set of answer sentences are fixed beforehand. To overcome this issue, we have proposed response generation using a statistical machine translation technique. In this paper, we make use of N-best speech recognition candidates instead of manual transcription used in our previous study. As a result, the generation rate of appropriate response sentences was improved by using multiple recognition hypothesis.

### Multi-Stream Acoustic Model Adaptation for Noisy Speech Recognition

Tamura Satoshi *Gifu University*, Hayamizu Satoru *Gifu University*

In this paper, a multi-stream-based model adaptation method is proposed for speech recognition in noisy or real environments. The proposed scheme comes from our experience about audio-visual model adaptation. At first, an acoustic feature vector is divided into several vectors (e.g. static, first-order and second-order dynamic vectors), namely streams. While adaptation, a stream performing relatively high recognition performance is updated for the stream only. Alternatively, a stream having less recognition power is adapted using all the streams that are superior to the stream. In order to evaluate the proposed technique, recognition experiments were conducted using every streams, and then adaptation experiments were also investigated for various types of combination of streams.

## OS.4-SPS.1 Complexity-efficient Design and Implementation for Signal Processing Systems

Session Chairs: Shang-Ho (Lawrence) Tsai, Chih-Hung Kuo Location: Laurel

### A Highly Parallel Design for Irregular LDPC Decoding on GPGPUs

Tsou-Han Chiu *National Chiao Tung University*, Hsien-Kai Kuo *National Chiao Tung University*, Bo-Cheng Charles Lai *National Chiao Tung University*

Low-Density Parity-Check (LDPC) code is a powerful error correcting code. It has been widely adopted by many communication systems. Finding a fast and efficient design of LDPC has been an active research area. This paper proposes a high performance design for irregular LDPC decoding on a general purpose graphic processing unit (GPGPU). A GPGPU is a many-core architecture which enables massively parallel computing. In this paper, a high degree of computation parallelism has been exposed by decoding multiple LDPC code-words concurrently. An innovative data structure is proposed to more efficiently leverage memory coalescing for the irregular data accesses of LDPC decoding. Data spatial locality is maximized by keeping more reusable data within the on-chip cache of a GPGPU. The data communication overhead between a host and a GPGPU is minimized through a single word copy for the convergence check. The experiment results show that the proposed design can achieve up to 55.68X runtime improvement, when compared with a sequential LDPC program on a CPU.

**Vehicular Signal Transmission Using Power Line Communications**

Yen-Chang Chen *National Chiao Tung University*, Shang-Ho Tsai *National Chiao Tung University*, Kai-jiun Yang *National Chiao Tung University*, Ping-Fan Ho *Industrial Technology Research Institute*, Kuo-Feng Tseng *National Chiao Tung University*, Ho-Shun Chen *National Chiao Tung University*

We propose a power line communication (PLC) system for transmitting control signals in vehicle through internal power lines, such that the internal wires can be reduced and the vehicles can be lighter. The signals from different devices are multiplexed and modulated to the power lines by the transmitter. In the receiver, first the noise within power lines are filtered. Afterwards the multiplexed signals are selectively extracted by specific codes which are minimally correlated. Finally the control signals are restored from the error-control coded bits which make the information more robust. The maximum data rate of the chip is 50 kbps, and the die area is 3.74 mm^2 using a TSMC 0.18 um standard cell library with power consumption 22.49 mW.

**High-Performance Turbo-MIMO System Design with Iterative Soft-Detection and Decoding**

Der-Wei Yang *National Cheng Kung University*, Jing-Shiun Lin *National Cheng Kung University*, Shih-Hao Fang *National Cheng Kung University*, Chia-Fen Lin *National Cheng Kung University*, Ming-Der Shieh *National Cheng Kung University*

In turbo-multiple-input multiple-output (Turbo-MIMO) systems, the soft-output MIMO detector can provide the priori information to the turbo decoder. Unfortunately, if Rayleigh fading channels are applied, the induced unreliable priori information would cause the system performance degradation. In this paper, we proposed an iterative method to acquire the high reliability priori information from MIMO soft-detector in Turbo-MIMO systems. Similar to the conventional updating rules in the turbo decoding algorithm, we utilize the extrinsic information from the turbo decoder to update the log-likelihood ratios (LLRs) based on log-MAP algorithm in the list sphere decoding (LSD) algorithm. To reduce the overall computational complexity, different iteration profiles are also discussed. Simulation results show that the proposed Turbo-MIMO system can significantly improve the system performance compared to that of the conventional Turbo-MIMO system.

**Hardware Architecture Design of Hybrid Distributed Video coding with Frame Level Coding Mode Selection**

Chieh-Chuan Chiu *National Taiwan University*, Hsin-Fang Wu *National Taiwan University*, Shao-Yi Chien *National Taiwan University*, Chia-han Lee *Academia Sinica*, V. Srinivasa Somayazulu *Intel Corporation*, Yen-Kuang Chen *Intel Corporation*

Distributed video coding (DVC), a new video coding paradigm based on Slepian-Wolf and Wyner-Ziv theories, is a promising solution for implementing low-power and low-cost distributed wireless video sensors since most of the computation load is moved from the encoder to the decoder. In this paper, the hardware architecture design of an efficient distributed video coding system, hybrid DVC with frame-level coding mode selection, is proposed. With the fully block-pipelined architecture, coding mode pre-decision, and specially-designed LDPC engine, the proposed hardware is an efficient solution for distributed video sensors with high rate-distortion performance.

**Hardware-Efficient EVD Processor Architecture in FastICA for Epileptic Seizure Detection**

Yi-Hsin Shih *National Chiao Tung University*, Tsan-Jieh Chen *National Chiao Tung University*, Chia-Hsiang Yang *National Chiao Tung University*, Herming Chiueh *National Chiao Tung University*

Independent component analysis (ICA) is a key signal processing technique to improve the detection accuracy of epileptic seizures. It separates artifacts and epileptic signals, which facilitates the succeeding signal processing for seizure detection. FastICA is an efficient algorithm to compute ICA through proper pre-processing. In the preprocessing stage of the FastICA, eigenvalue decomposition (EVD) is applied to reduce the convergence time of iterative calculation of weights for demultiplexing received multi-channel signals. To calculate EVD efficiently, the Jacobi method is preferable since an array structure is proposed to decompose matrix efficiently by leveraging givens rotations. Multiple diagonal and off-diagonal processing elements run in parallel to calculate EVD. The micro-rotations can be realized efficiently by coordinate rotation digital computer (CORDIC), which calculates trigonometric functions using only addition, shift, and table lookup without dedicated multipliers. In this work, an approximate Jacobi is adopted instead to reduce the number of iterations significantly. Optimized rotation angles can be calculated efficiently using shift-add operations for multiplications with coefficients of power of 2 in the diagonal processing elements. Normalization operation in the original mathematical formulation can be omitted due to signal re-scaling in both diagonal and off-diagonal processing elements. The number of processing cycles is reduced by 6 times for each sweep due to the reduced number of pipelining stages in the critical path. The approximate Jacobi method provides a 6x speedup (185-252 cycles instead of 1440 cycles) for a 6-channel EVD. An overall 77.2% area reduction is achieved due to arithmetic simplification and hardware reduction. The hardware architecture is verified by testing the human electroencephalogram (EEG) signals from the Freibur

**Tracking Performance Analysis of the Set-Membership NLMS Adaptive Filtering Algorithm**

Reza Arablouei *University of South Australia*, Kutluyil Dogancay *University of South Australia*

In this paper, we analyze the tracking performance of the set-membership normalized least mean squares (SM-NLMS) adaptive filtering algorithm using the energy conservation argument. The analysis leads to a nonlinear equation whose solution gives the steady-state mean squared error (MSE) of the SM-NLMS algorithm in a nonstationary environment. We prove that there is always a unique positive solution for this equation. The results predicted by the analysis show good agreement with the simulation experiments.

## OS.5-IVM.3 Recent Topics in Computer Vision and Image Processing

Session Chair: Salina Abdul Samad Location: Trousdale Estates

**Facial Image Prediction Using Exemplar-based Algorithm and Non-negative Matrix Factorization**

Hsuan-Ting Chang *National Yunlin University of Science and Technology*, Hsiao-wei Peng *National Yunlin University of Science and Technology*

Human aging face prediction is a popular research topic because of its various useful applications such as security system, missing persons search system, etc. In this study, we propose Exemplar-based Algorithm whose property considers the environment of human growth. Moreover, both the non-negative matrix factorization and linear interpolation methods are used to perform the prediction for six facial ROIs. In the proposed method, we employ the family images, in which each family member has more than one images at different ages. And we predict the image ROIs to replace the original ones to obtain the prediction result. However, it is difficult to collect the facial image ROI of families at various age, we also refer the databases from the internet. In experimental results, the correlation coefficient between the real and predicted images can reach 0.82. However, the factor such as expression and light in the reference images could result in lower correlation coefficient.

### Video Prediction Block Structure and the Emerging High Efficiency Video Coding Standard

Shan Liu *MediaTek USA Inc.*, Shawmin Lei *MediaTek USA Inc.*

In the ISO/IEC 14496-10 |ITU-T H.264 advanced video coding (AVC) standard, the prediction block sizes can be 16x16, 8x8 and 4x4 for Intra prediction; 16x16, 16x8, 8x16, 8x8, 8x4, 4x8 and 4x4 for Inter prediction. In the first HEVC test model (HM1.0), each 2Nx2N Intra CU may consist of either one 2Nx2N prediction unit (PU) or four NxN prediction units; while each 2Nx2N Inter CU may consist of one 2Nx2N PU, two 2NxN or Nx2N PUs or four NxN PUs. Since then, some investigations have been made to the prediction block structure based on HM1.0, including the removal of NxN prediction partition mode for all coding units (CU) except the smallest CU, and the removal of 4x4 Inter prediction. Experimental results show that with both these simplifications, the encoder complexity can be greatly reduced at the minimum cost of coding efficiency. Therefore both of these two suggestions were adopted by the HEVC standard.

### Classification of Beverages Using Electronic Nose and Machine Vision Systems

Mazlina Mamat *Universiti Kebangsaan*, Salina Abdul Samad *Universiti Kebangsaan*

In this work, the classification of beverages was conducted using three approaches: by using the electronic nose alone, by using the machine vision alone and by using the combination of electronic nose and machine vision. A total of two hundred and twenty eight beverages from fifteen different brands were used in this classification problem. A supervised Support Vector Machine was used to classify beverages according to their brand. Results show that by using the electronic nose alone and the machine vision alone were able to classify 73.7% and 92.9% of the beverages correctly. When combining the electronic nose and the machine vision, the classification accuracy was increased to 96.6%. Based on the results, it can be concluded that the combination of the electronic nose and the machine vision is able to extract more information from the sample, hence improving the classification accuracy.

### A Subjective Comparison of Depth Image Based Rendering and Frame Compatible Stereo for Low Bit Rate 3D Video Coding

Peshala Pahalawatta *DDD Inc.*, Kevin Stec *DDD Inc.*

Frame compatible stereo video delivery has become a de-facto standard because it enables the delivery of stereoscopic information over legacy devices that can currently only decode a 2D signal. At the cost of reducing spatial resolution of the images, frame compatible delivery also reduces the bandwidth requirements for signaling stereoscopic 3D video. The new generations of playback devices are less constrained than legacy devices in that they are increasingly becoming capable of decoding multiple video streams in parallel. Bandwidth, however, remains an issue especially in mobile wireless and real-time streaming environments. This paper explores the use of texture and depth data to render 3D views, and compares the bandwidth requirements of the depth based rendering method to frame compatible stereo. Some interesting subjective observations that affect the comparison are discussed along with the results of a formal subjective evaluation. The relative merits and drawbacks of each method are detailed both in terms of compression efficiency and overall quality of experience.

### Joint Perceptually-Based Intra Prediction and Quantization for HEVC

Guoxin Jin *Northwestern University*, Robert Cohen *MERL*, Anthony Vetro *MERL*, Huifang Sun *MERL*

This paper proposes a new coding scheme which jointly applies perceptual quality metrics to prediction, quantization and rate-distortion optimization (RDO) within the High Efficiency Video Coding (HEVC) framework. A new prediction approach which uses template matching is introduced. The template matching uses a structural similarity metric (SSIM) and a Just-Noticeable Distortion (JND) model. The matched candidates are linearly filtered to generate a prediction. We also modify the JND model and use Supra-threshold Distortion (StD) as the distortion measurement in RDO. Experimental results showing improvements for coding textured areas are presented as well.

## OS.6-BioSPS.1 Brain-body Physiological Networks Connectivity and Synchrony Analysis

Session Chairs: Tomasz M. Rutkowski, Zbigniew R. Struzik Location: Franklin Hills

### Linear and Nonlinear Features for Automatic Artifacts Removal from MEG Data Based on ICA

Montri Phothisonothai *University of Tokyo*, Hiroyuki Tsubomi *University of Tokyo*, Aki Kondo *University of Tokyo*, Yoshio Minabe *Kanazawa University*, Mitsuru Kikuchi *University of Tokyo*, Kastumi Watanabe *University of Tokyo*

This paper presents the automatic method to remove physiological artifacts from magnetoencephalogram (MEG) data based on independent component analysis (ICA). The proposed features including kurtosis (K), probability density (PD), central moment of frequency (CMoF), spectral entropy (SpecEn), and fractal dimension (FD) were used to identify the artifactual components such as cardiac, ocular, muscular, and sudden high-amplitude changes. For an ocular artifact, the frontal head region (FHR) thresholding was proposed. In this paper, ICA method was on the basis of FastICA algorithm to decompose the underlying sources in MEG data. Then, the corresponding ICs responsible for artifacts were identified by means of appropriate parameters. Comparison between MEG and artifactual components showed the statistical significance at $p<0.001$ for all features. The output artifact-free MEG waveforms showed the applicability of the proposed method in removing artifactual components.

### Sonification of Muscular Activity in Human Movements using the Temporal Patterns in EMG

Masaki Matsubara *University of Tsukuba*, Hiroko Terasawa *University of Tsukuba*, Hideki Kadone *University of Tsukuba*, Kenji Suzuki *University of Tsukuba/JAT*, Shoji Makino *University of Tsukuba*

Biofeedback is currently considered as an effective method for medical rehabilitation. It aims to increase the awareness and recognition of the body's motion by feeding back the physiological information to the patients in real time. Our goal is to create an auditory biofeedback that aids understanding of the dynamic motion involving multiple muscular parts, with the ultimate aim of clinical rehabilitation use. In this paper, we report the development of a real-time sonification system using EMG, and we propose three sonification methods that represent the data in pitch, timbre, and the combination of polyphonic timbre and loudness. Our user evaluation test involves the task of timing and order identification and a questionnaire about the subjective comprehensibility and the preferences, leading to a discussion of the task performance and usability. The results show that the subjects can understand the order of the muscular activities at 63.7% accuracy on average. And the sonification method with polyphonic timbre and loudness provides an 85.2% accuracy score on average, showing its effectiveness. Regarding the preference of the sound design, we found that there is not a direct relationship between the task performance accuracy and the preference of sound in the proposed implementations.

### Higher-order PLS for Classification of ERPs with Application to BCIs

Qibin Zhao *Brain Science Institute, RIKEN*, Liqing Zhang *Shanghai Jiao Tong University*, Cao Jianting *Saitama Institute of Technology*, Andrzej Cichocki *Brain Science Institute, RIKEN*

The EEG signals recorded during Brain Computer Interfaces (BCIs) are naturally represented by multi-way arrays in spatial, temporal, and frequency domains. In order to effectively extract the underlying components from brain activities which correspond to the specific mental state, we propose the higher-order PLS approach to find the latent variables related to the target labels and then make classification based on latent variables. To this end, the low-dimensional latent space can be optimized by using the higher-order SVD on a cross-product tensor, and the latent variables are considered as shared components between observed data and target output. The EEG signals recorded under the P300-type affective BCI paradigm were used to demonstrate the effectiveness of our new approach.

### Spatial Auditory BCI Paradigm Utilizing N200 and P300 Responses

Zhenyu Cai *University of Tsukuba*, Tomek Rutkowski *University of Tsukuba*

The paper presents our recent results obtained with a new auditory spatial localization based BCI paradigm in which the ERP shape differences at early latencies are employed to enhance the traditional P300 responses in an oddball experimental setting. The concept relies on the recent results in auditory neuroscience showing a possibility to differentiate early anterior contralateral responses to attended spatial sources. Contemporary stimuli-driven BCI paradigms benefit mostly from the P300 ERP latencies in so called aha-response settings. We show the further enhancement of the classification results in spatial auditory paradigms by incorporating the N200 latencies, which differentiate the brain responses to lateral, in relation to the subject head, sound locations in the auditory space. The results reveal that those early spatial auditory ERPs boost online classification results of the BCI application. The online BCI experiments with the multi-command BCI prototype support our research hypothesis with the higher classification results and the improved information-transfer-rates.

### EEG Steady-State Synchrony Patterns Sonification

Teruaki Kaniwa *University of Tsukuba*, Masaki Matsubara *University of Tsukuba*, Tomek Rutkowski *University of Tsukuba*, Hiroko Terasawa *University of Tsukuba*

This paper describes an application of multichannel EEG sonification approach. We present results obtained with a multichannel sonification method tested with steady-state EEG responses. We elucidate brain synchrony patterns in auditory do- main with utilization of EEG coherence measure. The transitions in the synchrony patterns are represented as timbre (i.e. spectro- temporal) deviation and as spatial movement of the sound cluster. Our final sonification evaluation experiment with six subjects confirms the validity of the proposed brain synchrony elucidation approach.

## OS.7-WCN.1 Cooperative and Coordinated Wireless Communications

Session Chairs: Sau-Hsuan Wu, Wan-Jen Huang, Feng-Tsun Chien Location: Whitley Heights

### Ergodic Mutual Information of Amplify-and-Forward MIMO Relay Channels with LOS Components

Chung-Kai Hsu *National Sun Yat-sen University*, Chao-Kai Wen *National Sun Yat-sen University*, Jung-Chieh Chen *National Kaohsiung Normal University*, Wan-Jen Huang *National Sun Yat-sen University*, Pangan Ting *Industrial Technology Research*

In this paper, we address the ergodic mutual information of amplify_and_forward multiple_input multiple_output two_hop relay channels. In these channels, the source terminal, relay terminal, and destination terminal are equipped with a number of correlated antennas, and there presents a line_of_sight component on each link. The models have widel applications in the field of machine_type communication devices, such as meters and sensors. Given channel matrices with Gaussian entries, the mean of mutual information is derived under the large_system regimen, in which the number of antennas at the transmitter and the receiver go to infinity with a fixed ratio. Simulation results demonstrate that even for a moderate number of antennas at each end, the proposed analytical results provide undistinguishable results from those obtained by Monte_Carlo simulations. In addition, the well approximation property holds even if the entries of the channel matrices are non_Gaussian.

### Relay Selection in Multiuser Two-Way Cooperative Relaying Systems

Yi-Ru Liao *National Chiao Tung University*, Feng-Tsun Chien *National Chiao Tung University*, Min-Kuan Chang *National Chung Hsing University*

In this paper, we study a relay selection (RS) problem in multi-user two-way cooperative relaying systems. We consider a more practical scenario in which multiple users, multiple relays and a single destination are involved in the two-way network. In this paper, the code division multiple access (CDMA) system with non-orthogonal spreading sequences is employed to handle the multiuser interference. Relay selection based on maximizing the SINR of the worse link is proposed in this research. Besides, aiming at mitigating the interference, we consider the design of linear filter at each relay such that the minimum SINR of the worst link in the two-way transmission is maximized. The result shows that the linear filter is similar to minimum mean-square error (MMSE) detector. Furthermore, we simulate the proposed scheme with several different parameters such as the numbers of users and relays, and the length of spreading sequences. Also, we compare the proposed RS method with random RS approach, and the result shows that our proposed method has better performance in terms of the bit error rate (BER).

### Game Theoretic Channel Allocation for the Delay-Sensitive Cognitive Radio Networks

Wenson Chang *National Cheng Kung University*, Yun-Li Yang *Qisda Corporation*

In this paper, we propose several channel allocation schemes via the Game theoretical approaches for the distributed CR networks. Distinguished from the literature, more important factors are taken into account when designing the potential games for the interweave and underlay CR networks, i.e. the queueing delay, complete inter-system interference and protection of primary users (PUs). Particularly, in the underlay CR networks, a PU can be protected by adaptively adjusting the cost for secondary users (SUs) to share a subchannel with him. Consequently, SUs and PUs can achieve higher end-to-end throughput and maintain the desired signal-to-noise-and-interference ratio, respectively. Moreover, for proving the convergence of the proposed schemes, the associated potential functions are also defined. Via the simulation results, the proposed schemes are proved to be capable of effectively reducing the queuing delay at the cost of the slightly decreased throughput.

### Robust Linear Beamformer Designs for CoMP AF Relaying in Downlink Multi-Cell Networks

Chun-I Kuo *MStar Semiconductor, Inc.*, Sau-Hsuan Wu *National Chiao Tung University*, Chun-Kai Tseng *National Chiao Tung University*

Robust beamforming methods are studied to support relay-assisted coordinated multi-point (CoMP) retransmissions in downlink multi-cell networks. Linear beamformers (BFers) for relay stations of different cells are jointly designed to maintain, in a CoMP amplify-and-forward (AF) relaying manner, the target signal to interference-plus-noise ratios (SINR) at the cellular boundaries of this type of networks. Considering the feasibility in realizations, BFer designs are only allowed to use the channel state information (CSI) feedbacks of the wireless links inside a network. This kind of designs turns out to be a challenging optimization problem when attempting to maintain the SINR under the estimation and quantization errors in CSI. A conservative criterion and solution method is proposed for this robust design problem.

Despite the conservativeness, the proposed method appears to provide an effective BFer design for CoMP AF relaying, either from the perspective of power consumption or from the viewpoints of BFers' complexity and \emph{feasibility} in syntheses. Simulations also show that when applying the proposed CoMP AF relaying method in Automatic Retransmission reQuest (ARQ), data throughput can be efficiently increased for users close to the joint cellular boundaries inside a multi-cell network.

**Zero-Forcing Design of Precoders and Decoders in Multiuser CDMA Cooperative Networks**
Li-Chung Lo *National Sun Yat-sen University*, Wan-Jen Huang *National Sun Yat-sen University*, Chun-Ting Liu *National Sun Yat-sen University*

Consider a multiuser cooperative CDMA networks where multiple sources transmit signals toward their respective destinations with assistance of multiple relays. We propose joint designs of precoders at relays and decoders at the destinations to eliminate MAI and improve system performance. Specifically, two sub-optimal designs of precoders are developed to maximize SNR averaged over all users and to maximize SNR of the worst user respectively. It shows through computer simulations that the precoder maximizing average SNR favors the best user, while the precoder maximizing the minimal SNR balances radio usage of relays such that all users can achieve near-optimal diversity order.

**Distributed Beamforming with Compressed Feedback in Time-Varying Cooperative Networks**
Miao-Fen Jian *National Sun Yat-sen University*, Wan-Jen Huang *National Sun Yat-sen University*, Chao-Kai Wen *National Sun Yat-sen University*

In this paper, we investigate distributed beamforming with limited feedback for time varying cooperative networks with multiple amplify-and-forward (AF) relays. With perfect channel state information, transmit beamforming has been shown to achieve significant diversity and coding gain in both MIMO or cooperative systems. However, it requires large amount of overhead for receiver to feed back channel information or beamforming coefficients, which makes it impractical. To perform transmit beamforming with limited feedback, the destination can choose the best codeword as beamforming vector from a predetermined codebook. In this work, we adopt the generalized Lloyd algorithm (GLA) to optimize codebook in terms of maximal average SNR. Furthermore, the feedback message can be compressed by exploiting temporal correlation of channel states. Specifically, We model channel states as a first-order finite-state Markov chain, and propose two compression methods according to the property of the transition probabilities among different channel states. Simulations show that distributed beamforming with compressed feedback performs closely to the case with infinite feedback.

## OS.8-SLA.2 Speech Processing (I)

Session Chair: Jianwu Dang Location: Mt. Olympus

**Mandarin Vowel Synthesis Based on 2D and 3D Vocal Tract Model by Finite-Difference Time-Domain Method**
Yuguang Wang *Tianjin University*, Hongcui Wang *Tianjin University*, Jianguo Wei *Tianjin University*, Jianwu Dang *JAIST/Tianjin University*

Finite-difference time-domain (FDTD) method is an effective numerical method to do acoustic simulation. This paper focused on the details of Mandarin vowel synthesis based on 2D and 3D vocal tract model by FDTD method. To do so, a 3D vocal tract shape and vocal tract area function were extracted from the MRI volumetric images during Chinese vowel production. 3D and 2D model with staggered FDTD mesh were constructed based on the vocal tract and the area function, respectively. Finally, vowels were synthesized by simulating wave sound propagation in the vocal tract using FDTD method with the two-mass vocal folds model. The formant frequencies of synthesized vowels were compared to those of real speech sounds. It is found that the mean absolute errors of formant frequencies were 7.77% and 6.07% for 2D and 3D model, respectively. Results suggested that both 2D and 3D model are capable of producing speech formants in about the same accuracy. However, 3D method exhibits more realistic phenomenon in high frequency region because it was based on complete 3D vocal tract model. It is also observed that the bandwidths of real speech can be achieved by setting the normal sound absorption coefficient within a proper range.

**Speaker Adaptation Intensively Weighted on Mis-Recognized Speech Segments**
Takahiro Oku *NHK*, Yuya Fujita *NHK*, Akio Kobayashi *NHK*, Toru Imai *NHK*

A “re-speak method” is an effective speech recognition method for simultaneous closed-captioning of live broadcasting programs picked up in noisy environments featuring spontaneous or emotional commentary. An acoustic model of the re-speaker needs to be constantly adapted according to the re-speaker's daily health condition or level of fatigue. In this paper, we propose efficient speaker adaptation for the re-speak method. Conventional speaker adaptation is performed uniformly over entire speech segments. In comparison, our proposed speaker adaptation determines intensive adaptation segments corresponding to recognition error parts by comparing speech recognition results and manually error-corrected results. These results are provided in real time by the simultaneous closed-captioning process. Then, the frame-level statistics for speaker adaptation are multiplied by larger weights in proportion to the degree of the recognition errors more over the intensive adaptation segments than they are over the other segments. In an experiment on an information variety program in Japanese broadcasting, our speaker adaptation method reduced the word error rate relatively by 3.4% compared with the conventional uniform adaptation method.

**Introduction of False Detection Control Parameters in Spoken Term Detection**
Yuto Furuya *University of Yamanashi*, Satoshi Natori *University of Yamanashi*, Hiromitsu Nishizaki *University of Yamanashi*, Yoshihiro Sekiguchi *University of Yamanashi*

This paper describes spoken term detection (STD) with false detection control. Our STD method uses phoneme transition network (PTN) derived by multiple automatic speech recognizers (ASRs) as an index. An PTN is almost the same to a sub-word based confusion network (CN), which is derived from an output of an ASR. The PTN-based index we proposed is made of the outputs of multiple ASRs, which is known to be robust to certain recognition errors and the out-of-vocabulary problem. Our PTN was very effective at detecting query terms. However, the PTN generates a lot of false detections especially for short query terms. Therefore, we applied two false detection control parameters to the Dynamic Time Warping-based term detection engine. In addition, we changed the search parameters depending on length of a query term. Finally, the STD performance was better (0.785 of F-measure) than without any parameters (0.717).

**Pipeline Decomposition of Speech Decoders and Their Implementation Based on Delayed Evaluation**
Takahiro Shinozaki *Chiba University*, Sadaoki Furui *Tokyo Institute of Technology*, Yasuo Horiuchi *Chiba University*, Shingo Kuroiwa *Chiba University*

For large vocabulary continuous speech recognition, speech decoders treat time sequence with context information using large probabilistic models. The software of such speech decoders tend to be large and complex since it has to handle both relationships of its component functions and timing of computation at the same time. In the traditional signal processing area such as measurement and system control, block diagram based implementations are common where systems are designed by connecting blocks of components. The connections describe flow of signals and this framework greatly helps to understand and design complex systems. In this research, we show that speech decoders can be effectively decomposed to diagrams or pipelines. Once they are decomposed to pipelines, they can be easily implemented in a highly abstracted manner using a pure functional programming language with delayed evaluation. Based on this perspective, we have re-designed our pure-functional decoder Husky proposing a new design paradigm for speech recognition systems. In the evaluation experiments, it is shown that it efficiently works for a large vocabulary continuous speech recognition task.

### Consonant Enhancement for Articulation Disorders Based on Non-negative Matrix Factorization

Ryo Aihara *Kobe University*, Ryoichi Takashima *Kobe University*, Tetsuya Takiguchi *Kobe University*, Yasuo Ariki *Kobe University*

We present consonant enhancement on a voice for a person with articulation disorders resulting from athetoid cerebral palsy. The movement of such speakers is limited by their athetoid symptoms, and their consonants are often unstable or unclear, which makes it difficult for them to communicate. Speech recognition for articulation disorders has been studied; however, its recognition rate is still lower than that of physically unimpaired persons. In this paper, an exemplar-based spectral conversion using non-negative Matrix Factorization (NMF) is applied to consonant enhancement of a voice with articulation disorders. The source speaker's spectrum is easily converted into a well-ordered speaker's spectrum. Its effectiveness is examined for voice quality and clarity of consonants for a person with articulation disorders.

## PS.1-SLA.3 Speech Recognition (I)

Session Chair: Masato Akagi Location: Solano

### Fast Spoken Term Detection Using Pre-retrieval Results of Syllable Bigrams

Hiroyuki Saito *Iwate Prefectural University*, Yoshiaki Itoh *Iwate Prefectural University*, Kazunori Kojima *Iwate Prefectural University*, Masaaki Ishigame *Iwate Prefectural University*, Kazuyo Tanaka *Tsukuba University*, Shi-wook Lee *National Institute of Advanced Industrial Science and Technology*

We propose a method of the Spoken Term Detection (STD) based on a priori retrieval results in which plural syllables are used as query terms. In the proposed method, all N-syllable combinations such as syllable bigrams are searched for in spoken documents. In the first step of the method, the retrieval results are prepared a priori, where pre-retrieval results include candidates with scores matching those of each N-syllable sequence. Given a query, the syllable sequence of the query is divided into plural syllable sequences whose lengths are the same as those of the pre-retrieval results. In the second step, the candidate sections are filtered by using the scores of query's syllable combinations. This reduction in the number of candidate sections for detailed matching leads to a large reduction of the retrieval time. In the third step, these candidates sections are re-scored by performing detailed matching. Experimental results show that the proposed method reduces the retrieval time by 93% with a performance degradation of less than 2 points.

### Simplifying Emotion Classification Through Emotion Distillation

Emily Provost *University of Michigan*, Shrikanth Narayanan *University of Southern California*

Many state-of-the-art emotion classification systems are computationally complex. In this paper we present an emotion distillation framework that decreases the need for computational complex algorithms while maintaining rich, and interpretable, emotional descriptors. These representations are important for emotionally-aware interfaces, which we will increasingly see in technologies such as mobile devices with personalized interaction paradigms and in behavioral informatics. In both cases these technologies require the rapid distillation of vast amounts of data to identify emotionally salient portions. We demonstrate that emotion distillation can produce rich emotional descriptors that serve as an input to simple classification techniques. This system obtains results that match state-of-the-art classification results on the USC IEMOCAP data.

### Speech Emotion Recognition System Based on a Dimensional Approach Using a Three-Layered Model

Reda Elbarougy *Japan Advanced Institute of Science and Technology*, Masato Akagi *Japan Advanced Institute of Science and Technology*

This paper proposes a three-layer model for estimating the expressed emotions in a speech signal based on a dimensional approach. Most of the previous studies using the dimensional approach mainly focused on the direct relationship between acoustic features and emotion dimensions (valence, activation, and dominance). However, the acoustic features that correlate to valence dimension are less numerous, less strong, and the valence dimension has being particularly difficult to be predicted. The ultimate goal of this study is to improve the dimensional approach in order to precisely predict the valence dimension. The proposed model consists of three layers: acoustic features, semantic primitives, and emotion dimensions. We aimed to construct a three-layer model in imitation of the process of how human perceive and recognize emotions. In this study, we first investigated the correlations between the elements of the two-layered model and elements of the three-layered model. In addition, we compared the two models by applying a fuzzy inference system (FIS) to estimate emotion dimensions. In our model FIS was used to estimate semantic primitives from acoustic features, then to estimate emotion dimensions from the estimated semantic primitives. The experimental results show that the proposed three-layered model outperforms the traditional two-layered model.

### A Spoken Dialogue System Using Virtual Conversational Agent with Augmented Reality

Shinji Miyake *Tohoku University*, Akinori Ito *Tohoku University*

We have developed a spoken dialogue system using virtual conversational agent with augmented reality. The proposed dialogue system has architecture based on question and answer database that contains many question and answer pairs. Additionally, we have developed two agents displayed using augmented reality, which behave as avatars of objects to be operated. We evaluated user's impression as well as response accuracy of our proposed system. As a result, the existence of an agent increased user's feeling of vividness of conversation and easiness to talk to the system. In addition, the system with an agent showed better response accuracy than the system without agents.

### Data-driven Rescaled Teager Energy Cepstral Coefficients for Noise-robust Speech Recognition

Chia-Ping Chen *National Sun Yat-sen University*, Miau-Luan Hsu *National Sun Yat-sen University*

We investigate data-driven rescaled Teager energy cepstral coefficients (DRTECC) features for noise-robust speech recognition. In the first stage, we apply a bank of auditory gammatone filters (GTF) and extract Teager-Kaiser energy (TE) estimates, which substitute the commonly used mel-spectrum. The output features of the first stage are called the Teager energy cepstral coefficients (TECC). In the second stage, we apply a piecewise rescaling operation of the cepstral coefficients of the zeroth order to bridge the difference between clean and noisy utterances. The segmentation point is determined by voice activity detection (VAD), and the proportional constants are data-driven. The resultant features are called DRTECC. The proposed features are evaluated on the Aurora 2.0 database. The relative improvements over the baseline MFCC features are significant.

## PS.2-SLA.4 Audio & Music Processing (I)

Session Chair: Ryo Takahashi Location: Solano

### Diffusion Noise Suppression by Crystal-Shape Subtraction Array

Akira Tanaka *Hokkaido University*, Ryo Takahashi *Hokkaido University*

Noise suppression of diffusion noise by microphone arrays is discussed in this paper. In our previous work, we proposed a method for jointly estimating signal and noise correlation matrices from observations with diffusion noise by using so-called crystal shape microphone arrays; and discussed the performance of the Wiener filter based on those correlation matrices. In this paper, we propose a novel method for noise suppression of diffusion noise based on the newly adopted spectral subtraction scheme with the estimated correlation matrices by our previous work. We also verify the efficacy of the proposed method by some computer simulations and show that the proposed method outperforms our previous method by the Wiener filter.

### Reproduction of Varied Sound Image Localization for Real Source in Stereo Audio System

Satoshi Okuro *Kansai University*, Yoshinobu Kajikawa *Kansai University*

In this paper, we propose a sound reproduction system which can realize varied sound image localization in stereo audio systems. The proposed system can suppress unnatural variations of sound image localization with listener's movement and maintain the absolute position of sound image so that a real source exists in the corresponding position. Generally, human being perceives the direction of sound image on horizontal plane according to Interaural Level Difference (ILD) and Interaural Time Difference (ITD) between signals arriving at both ears. Accordingly, unnatural variation of sound image localization accompanying listener's movement is due to the differences of ILD and ITD between the stereo audio system and the real source. The proposed system therefore compensates ILD and ITD using digital filters. Some subjective assessment tests with ten subjects demonstrate that fixed sound image can be realized in the proposed system when listener moves away by giving appropriate signal level ratios.

### A Japanese Lyrics Writing Support System for Amateur Songwriters

Chihiro Abe *Tohoku University*, Akinori Ito *Tohoku University*

In this paper, we propose a lyrics writing support system focused on the number of syllables, rhyme and word accent. The system generates candidate sentences that satisfy user-specified conditions based on Ngram, and presents them. Users can use the system like a dictionary, and write lyrics be choosing presented sentences. In our subjective evaluations, we have investigated how the system is utilized for writing lyrics actually.The log of using the system and the questionnaires showed that users want the system to present words suitable for their images, and they used the presented words as keywords of a lyrics rather than as they are.

### Comparative Study on Various Noise Reduction Methods with Decision-Directed a Priori SNR Estimator via Higher-Order Statistics

Suzumi Kanehara *Nara Institute of Science and Technology*, Hiroshi Saruwatari *Nara Institute of Science and Technology*, Ryoichi Miyazaki *Nara Institute of Science and Technology*, Kiyohiro Shikano *Nara Institute of Science and Technology*, Kazunobu Kondo *Yamaha Corporate Research & Development Center*

In this paper, we propose a new theoretical analysis of amount of musical noise generated in several noise reduction methods with a decision-directed a priori SNR estimator using higher-order statistics. In our previous study, a musical noise assessment based on kurtosis has been successfully applied to spectral subtraction and Wiener filter. However, this approach cannot be applied to some high-quality noise reduction methods, namely, the minimum mean-square error short-time spectral amplitude estimator, the minimum mean square error log-spectral amplitude estimator and the maximum a posteriori estimator, because such methods include the decision-directed a priori SNR estimator, which corresponds to a nonlinear recursive (infinite) process for noise power spectral sequences. Therefore, in this paper, we introduce a computationally efficient higher-order-moment calculation method based on generalized Gauss-Laguerre quadrature. We also mathematically clarify the justification of using a typical decision-directed parameter, namely, magic number 0.98,_in the three types of the decision-directed-based estimators from a viewpoint of amounts of musical noise and speech distortion. In addition, we perform comparison between these noise reduction methods based on the mathematical analysis and human perception test.

### Toward Polyphonic Musical Instrument Identification using Example-based Sparse Representation

Okamura Mari *Gifu University*, Masanori Takehara *Gifu University*, Tamura Satoshi *Gifu University*, Hayamizu Satoru *Gifu University*

Musical instrument identification is one of the major topics in music signal processing. In this paper, we propose a musical instrument identification method based on sparse representation for polyphonic sounds. Such the identification has been still categorized into challenging tasks, since it needs high-performance signal processing techniques. The proposed scheme can be applied without any signal processing such as source separation. Sample feature vectors for various musical instruments are used for the base matrix of sparse representation. We conducted two experiments to evaluate the proposed method. First, the musical instrument identification is tested for monophonic sounds using five musical instruments. The average accuracy of 91.9% was obtained and it shows the effectiveness of the proposed method. Second, musical instrument composition of polyphonic sounds is examined, which contain two instruments. It is found that the estimated weight vector by sparse representation indicates the mixture ratio of two instruments.

### Optimization Scheme of Joint Noise Suppression and Dereverberation Based on Higher-Order Statistics

Fine Aprilyanti *Nara Institute of Science and Technology*, Hiroshi Saruwatari *Nara Institute of Science and Technology*, Kiyohiro Shikano *Nara Institute of Science and Technology*, Tomoya Takatani *Toyota Motor Company*

In this paper, we apply the higher-order statistics parameter to automatically improve the performance of blind speech enhancement. Recently, a method to suppress both diffuse background noise and late reverberation part of speech has been proposed combining blind signal extraction and Wiener filtering. However, this method requires a good strategy for choosing the set of its parameters in order to achieve the optimum result and to control the amount of musical noise, which is a common problem in non-linear signal processing. We present an optimization scheme to control the value of Wiener filter coefficients used in this method, which depends on the amount of musical noise generated, measured by higher-order statistics. The noise reduction rate and cepstral distortion are also evaluated to confirm the effectiveness of this scheme.

## 14:30 - 16:10

### OS.9-IVM.4 Image and Video Coding (I)

Session Chair: Hsueh-Ming Hang Location: Doheny

**Depth Coding using Coded Boundary Patterns**

Kai-Hsiang Yang *National Chiao Tung University*, Hsueh-Ming Hang *National Chiao Tung University*

The depth information plays an essential role in the virtual-view (or free-viewpoint) 3D video systems. In this paper, we propose a new algorithm to code a depth map for the purpose of virtual view synthesis. The idea is to use H.264/AVC to represent the rough shape (including depth values) of a depth map and then additional information is transmitted to improve the depth values around the object boundaries. The complete encoding and decoding simulation system was built on the H.264/AVC JM 18.0 platform. In our experiments, three tools can be turned on individually and thus four coding modes are defined and tested. Our data show that these proposed tools offer advantages in either coding efficiency or image quality improvement and some tools work best on simple images while the others work best on complex images. With proper parameter setting, the overall quality of virtual view rendering is noticeably improved.

**Color Image Coding based on the Colorization**

Takashi Ueno *Keio University*, Yoshida Taichi *Keio University*, Masaaki Ikehara *Keio University*

Colorization is a method which adds color components to grayscale images using color assigned information provided by the user. Recently, a novel approach to image compression called colorization based coding has been proposed. It automatically extracts color assignations from original color images at an encoder and restores color components by colorization method at a decoder. In this paper, we propose the method which improves the conventional color image coding methods by regarding colorization as interpolation. At the encoder, the proposed method subsamples chrominance components considering colorization and subsampled chrominance components are compressed by conventional methods. At the decoder, subsampled chrominance components are interpolated by colorization. Simulations reveal that the proposed method improves quality of reconstructed images, objectively.

**Improved JPEG 2000 System Using LS Prediction and Grouping Context Coding Scheme**

Jian-Jiun Ding *National Taiwan University*, Hsin-Hui Chen *National Taiwan University*, Guan-Chen Pan *National Taiwan University*, Po-Hung Wu *National Taiwan University*

In this paper, two coding strategies are proposed to improve the coding efficiency of the JPEG2000 system. First, instead of using the embedded block coding with optimized truncation (EBCOT) scheme in all subbands, we apply the algorithm based on least square prediction in the LL part of the discrete wavelet transform domain. Moreover, in the LH, HL, and HH parts, instead of using the MQ coder and the fixed probability table, we group the 19 contexts into 7 classes. Since the characteristics of the contexts in EBCOT are different from one another, it is proper to use different probability tables for different classes. Simulation results demonstrate that the proposed methods significantly improve the coding efficiency of the JPEG2000 system.

**A New In-Loop Filter for Depth Map Coding in HEVC**

Hyunsuk Ko *University of Southern California*, C.-C. Jay Kuo *University of Southern California*

A depth-map is used to synthesize virtual texture views in the multi-view plus depth (MVD) format. In conventional video coding, a coded depth-map often suffers from compression artifacts along object boundaries, which have a negative effect on the quality of rendered images in the view synthesis process. To address this problem, we propose a depth-map boundary filtering technique to eliminate coding artifacts while preserving sharp edges. This can be mathematically formulated as a L0-norm minimization problem. This filtering process is cascaded with the de-blocking filter in the emerging HEVC video coding standard to result in a new in-loop filter. Experimental results are given to show that the subjective and objective quality of the synthesized views is enhanced by the introduction of the new in-loop fitler.

**Modification of Intra Angular Prediction in HEVC**

Shohei Matsuo *NTT Corporation*, Seishi Takamura *NTT Corporation*, Atsushi Shimizu *NTT Corporation*

Intra prediction of the emerging High Efficiency Video Coding (HEVC) standard has new features that the existing video coding standard H.264/AVC does not have. One example is that new angular prediction modes are added. Finer prediction directions enable to reduce the prediction error energy by making the predicted signals more flexibly. A method to make reference samples for the intra angular prediction plays an important role in terms of the coding efficiency. In the angular prediction of HEVC, a simple 2-tap linear filter is used to make reference samples. In this paper, the reference samples are generated by the conventional 2-tap linear filter or a DCT-based interpolation filter. The proposal improves the intra prediction performance especially for small prediction units such as 4x4 and 8x8. The average coding gains against the anchor of HEVC test model (HM6.0) were about 0.34% and 0.31%, when the tap length of DCT-IF is set to four and six, respectively. The maximum coding gains were about 2.2%, 3.3%, and 3.9% for each component (Y, Cb, and, Cr). In the case of four tap interpolation, the average run-times of encoding and decoding were about 102.84% and 100.96%, respectively.

### OS.10-IVM.5 3DTV and Free-viewpoint TV (I)

Session Chairs: Masayuki Tanimoto, Yo-Sung Ho Location: Beachwood

**3D Video Coding with Depth Modeling Modes and View Synthesis Optimization**

Karsten Mueller *Fraunhofer HHI*, Philipp Merkle *Fraunhofer HHI*, Gerhard Tech *Fraunhofer HHI*, Thomas Wiegand *Fraunhofer HHI*

This paper presents efficient coding tools for depth data in depth-enhanced video formats. The method is based on the high-efficiency video codec (HEVC). The developed tools include new depth modeling modes (DMMs), in particular using non-rectangular wedgelet and contour block partitions. As the depth data is used for synthesis of new video views, a specific 3D video encoder optimization is used. This view synthesis optimization (VSO) considers the exact local distortion in a synthesized intermediate video portion or image block for the depth map coding. In a fully optimized 3D-HEVC coder, VSO achieves average bit rate savings of 17%, while DMMs gain 6%in BD rate, even though the depth rate only contributes 10% to the overall MVD bit rate.

### Depth Map Up-sampling Based on Edge Layers

Danillo Graziosi *MERL*, Dong Tian *MERL*, Anthony Vetro *MERL*

Depth map images are characterized by large ho- mogeneous areas and strong edges. It has been observed that efficient compression of the depth map is achieved by applying a down-sampling operation prior to encoding. However, since high resolution depth maps are typically required for view synthesis, an up-sampling method that is able to recover the loss of information is needed within this framework. In this paper, an up-sampling algorithm that recovers the high frequency content of depth maps using a novel edge layer concept is proposed. This algorithm includes a method for extracting edge layers from the corresponding texture images, which are then used as part of a non-linear interpolation filter for depth map up- sampling. In the present work, the up-sampling is applied as a post-processing operation to generate multiview output for display. Views synthesized with our up-sampled depth maps show the efficiency of our proposed technique relative to conventional interpolation filters.

### Hybrid Plane Fitting for Depth Estimation

Lingfeng Xu *HKUST*, Oscar C. Au *HKUST*, Wenxiu Sun *HKUST*, Yujun Li *HKUST*, Jiali Li *HKUST*

In this paper, a novel plane Þtting algorithm with low complexity and high accuracy is proposed to reÞne the depth maps generated by stereo matching. We Þrst compute the conÞdence coefÞcient for each pixel in the depth map by cross checking and stable pixel calculation. According to the outlier pixel percentage for each segment, we choose one method, either proposed weighted least square error based or RANSAC based plane Þtting algorithm, to estimate the plane parameters. Experimental results show that our method outperforms other existing plane Þtting algorithms.

### Ray Capture Systems for FTV

Masayuki Tanimoto *Nagoya Industrial Science Research Institute*

FTV (Free-viewpoint Television) is an innovative visual media that allows users to view a 3D scene by freely changing their viewpoints. Thus, it enables realistic viewing and free navigation of 3D scenes. FTV is the ultimate 3DTV with infinite number of views and ranked as the top of visual media. FTV is not a conventional pixel-based system but a ray-based system. Ray capture, processing and display technologies have been developed for FTV. Here, three types of ray capture systems are presented. They are multi-camera ray capture with view interpolation, all-around dense ray capture without view interpolation and computational ray capture by reduced number of pixel data.

## OS.11-SLA.5 Multimodal Information Processing - Algorithms and Applications

Session Chairs: Lei Xie, Helen Meng Location: Runyon

### Dimensional Emotion Driven Facial Expression Synthesis Based on the Multi-Stream DBN Model

Hao Wu *Northwestern Polytechnical University*, Dongmei Jiang *Northwestern Polytechnical University*, Yong Zhao *Northwestern Polytechnical University*, Hichem Sahli *Vrije Universiteit Brussel*

This paper proposes a dynamic Bayesian network (DBN) based MPEG-4 compliant 3D facial animation synthesis method driven by the (Evaluation, Activation) values in the continuous emotion space. For each emotion, a state synchronous DBN model (SS_DBN) is firstly trained using the Cohn-Kanade (CK) database with two streams of inputs: (i) the annotated (Evaluation, Activation) values, and (ii) the extracted Facial Action Parameters (FAPs) of the face image sequences. Then given an input (Evaluation, Activation) sequence, the optimal FAP sequence is estimated via the maximum likelihood estimation (MLE) criterion, and then used to construct the MPEG-4 compliant 3D facial animation. Compared with the state-of-the-art approaches where the mapping between the emotional space and the FAPs has been made empirically, in our approach the mapping is learned and optimized using DBN to fit the input (Evaluation, Activation) sequence. Emotion recognition results on the constructed facial animations, as well as subjective evaluations, show that the proposed method obtains natural facial animations representing well the dynamic process of the emotions from neutral to exaggerate.

### Modeling the Correlation between Modality Semantics and Facial Expressions

Jia Jia *Tsinghua University*, Xiaohui Wang *Tsinghua University*, Zhiyong Wu *Tsinghua University*, Lianhong Cai *Tsinghua University*, Helen Meng *Chinese University of Hong Kong*

Facial expression plays an important role in face-to-face human-computer communication. Although considerable efforts have been made to enable computers to speak like human beings, how to express the rich semantic information through facial expression still remains a challenging problem. In this paper, we use the concept of “modality” to describe the semantic information which is related to the mood, attitude and intention. We propose a novel parametric mapping model to quantitatively characterize the non-verbal modality semantics for facial expression animation. In particular, seven-dimensional semantic parameters (SP) are first defined to describe the modality information. Then, a set of motion patterns represented with Key FAP (KFAP) is used to explore the correlations of MPEG-4 facial animation parameters (FAP). The SP-KFAP mapping model is trained with the linear regression algorithm (AMMSE) and an artificial neural network (ANN) respectively. Empirical analysis on a public facial image dataset verifies the strong correlation between the SP and KFAP. We further apply the mapping model to two different applications: facial expression synthesis and modality semantics detection from facial images. Both objective and subjective experimental results on the public datasets show the effectiveness of the proposed model. The results also indicate that the ANN method can significantly improve the prediction accuracies in both applications.

### Detection of Ball Hits in a Tennis Game Using Audio and Visual Information

Qiang Huang *University of East Anglia*, Stephen Cox *University of East Anglia*, Xiangzeng Zhou *Northwestern Polytechnical University*, Lei Xie *Northwestern Polytechnical University*

In this paper we describe a framework to improve the detection of ball hit events in tennis games by combining audio and visual information. Detection of the presence and timing of these events is crucial for the understanding of the game. However, neither modality on its own gives satisfactory results: audio information is often corrupted by noise and also suffers from acoustic mismatch between the training and test data, and visual information is corrupted by complex backgrounds, camera calibration, and the presence of multiple moving objects. Our approach is to first attempt to track the ball visually and hence estimate a sequence of candidate positions for the ball, and to then locate putative ball hits by analysing the ball’s position in this trajectory. To handle the severe interferences caused by false ball candidates, we smooth the trajectory by using locally weighted linear regression and removing the frames where there are no candidates. We use Gaussian mixture models to generate estimates of the times of hits using the audio information, and then integrate these two sources of information in a probabilistic framework. Testing our approach on three complete tennis games shows significant improvements in detection over a range of conditions when compared with using a single modality.

### Face Sketch-to-Photo Synthesis from Simple Line Drawing

Yang Liang *Zhejiang University*, Mingli Song *Zhejiang University*, Lei Xie *Northwestern Polytechnical University*, Jiajun Bu *Zhejiang University*, Chun Chen *Zhejiang University*

Face sketch-to-photo synthesis has attracted increasing attention in recent years for its useful applications on both digital entertainment and law enforcement. Although great progress has been made, previous methods only work on face sketches with rich textures which are not easily to obtain. In this paper, we propose a robust algorithm for synthesizing a face photo from a simple line drawing that contains only a few lines without any texture. In order to obtain a robust result, firstly, the input sketch is divided into several patches and edge descriptors are extracted from these local input patches. Afterwards, an MRF framework is built based on the divided local patches. Then a series of candidate photo patches are synthesized for each local sketch patch based on a coupled dictionary learned from a set of training data. Finally, the MRF is optimized to get the final estimated photo patches for each input sketch patch and a realistic face photo is synthesized. Experimental results on CUHK database have validated the effectiveness of the proposed method.

### High Quality Lips Animation with Speech and Captured Facial Action Unit as A/V Input

Lijuan Wang *Microsoft*, Frank Soong *Microsoft*

Rendering realistic lips movements in avatar with camera captured human's facial features is desirable in many applications, e.g. telepresence, video gaming, social networking, etc. We have proposed to use Gaussian Mixture Model (GMM) to generate lips trajectory and successfully tested in speech-to-lips conversion experiments, where only audio signal (speech) is used as input. In this paper real-time user's facial features called the Action Units (AUs) well tracked by Microsoft Kinect SDK with a consumer-grade RGB camera, are combined with speech to form joint A/V input for lips animation. We test the lips animation performance and show that the new combined A/V input can improve the conversion error rate by 22% in a speaker dependent test, compared with a baseline system.

## OS.12-SLA.6 Recent Advances in Audio and Acoustic Signal Processing (I)

Session Chairs: Shoji Makino, Hiroshi Saruwatari Location: Laurel

### Beamformer Design Using Measured Microphone Directivity Patterns: Robustness to Modelling Error

Mark Thomas *Microsoft Research*, Jens Ahrens *Microsoft Research*, Ivan Tashev *Microsoft Research*

The design process for time-invariant acoustic beamformers often assumes that the microphones have an omnidirectional directivity pattern, a flat frequency response in the range of interest, and a 2D environment in which wavefronts propagate as a function of azimuth angle only. In this paper we investigate those cases in which one or more of these assumptions do not hold, considering a Minimum Variance Distortionless Response (MVDR)-based solution that is optimized using measured directivity patterns as a function of azimuth, elevation and frequency. Robustness to modelling error is controlled by a regularization parameter that produces a suboptimal but more robust solution. A comparative study is made with the 4-element cardioid microphone array employed in Microsoft Kinect for Windows, whose beamformer weights are calculated with directivity patterns using (a) 2D cardioid models, (b) 3D cardioid models and (c) 3D measurements. Speech recognition and PESQ results are used as evaluation criteria with a noisy speech corpus, revealing empirically optimal regularization parameters for each case and up to a 70% relative improvement in word error rate comparing (a) and (c).

### Theoretical Analysis of Musical Noise in Nonlinear Noise Reduction Based on Higher-Order Statistics

Yu Takahashi *Yamaha Corporation*, Ryoichi Miyazaki *Nara Institute of Science and Technology*, Hiroshi Saruwatari *Nara Institute of Science and Technology*, Kazunobu Kondo *Yamaha Corporate Research & Development Center*

In this paper, we review a musical-noise-generation analysis of nonlinear noise reduction techniques with using higher-order statistics (HOS). Recently, an objective metric based on HOS to analyze nonlinear artifacts, i.e., musical noise, caused by nonlinear noise reduction techniques has been proposed. Such metric enables us to perform objective comparison of any nonlinear methods from the perspective of the amount of musical noise generated. Furthermore, such metric enables us to control the musical noise generated by nonlinear noise reduction techniques. In the paper, first, the mathematical principle of the analysis for the amount of musical noise based on HOS is described, and analyses and comparison examples of typical nonlinear noise reduction techniques are demonstrated. Next, it is clarified that to find a fixed point in HOS leads to no-musical noise property in noise reduction. Finally, several expansions on the theory are discussed.

### Auxiliary-function-based Independent Vector Analysis with Power of Vector-norm Type Weighting Functions

Nobutaka Ono *National Institute of Informatics*

This paper presents auxiliary-function-based independent vector analysis (AuxIVA) based on Generalized super Gaussian source model or Gaussian source model with time-varying variance. AuxIVA is a convergence-guaranteed iterative algorithm for independent vector analysis (IVA) with a spherical and super Gaussian source model, and the source model can be characterized by a weighting function. In this paper, as typical source models in AuxIVA, the generalized super Gaussian distribution or the Gaussian distribution with time-varying variance are considered. Both of them yield a power of vector-norm type weighting functions with an exponent parameter ß such that 0 ≤ ß ≤ 2. A scaling and a clipping technique for numerical stability are also discussed. The separation performance of AuxIVA with several ßs is compared.

### New Analytical Calculation and Estimation of TDOA for Underdetermined BSS in Noisy Environments

Takuro Maruyama *University of Tsukuba*, Shoko Araki *NTT Corporation Science Laboratories*, Tomohiro Nakatani *NTT Corporation*, Shigeki Miyabe *University of Tsukuba*, Takeshi Yamada *University of Tsukuba*, Shoji Makino *University of Tsukuba*, Atsushi Nakamura *NTT Corporation Science Laboratories*

We have proposed a new algorithm for sparseness-based underdetermined blind source separation (BSS) that can cope with diffused noise environments. This algorithm includes a technique for estimating time-difference-of-arrival (TDOA) parameter separately in individual frequency bins for each source. In this paper, we propose some methods that integrate the frequency-bin-wise TDOA parameter to estimate a TDOA of each source. The accuracy of the TDOA estimation by the proposed approach is shown by experiments in comparison with a conventional approach. The separation performance and calculation time of the proposed approach is also examined.

### A New Permutation Control Method for Frequency Domain BSS

Steven Grant *Missouri S&T*, Christopher Osterwise *Missouri S&T*

This paper introduces a new frequency domain blind source separation algorithm: Inter-frequency Correlation with Microphone Diversity (ICMD). Here, we consider using different sets of microphones where in each set the number of microphones and sources are equal. In the frequency domain, cascaded ICA initialization (CII) is used, where the separation matrix of one bin is used to initialize the ICA iterations of the next. CII greatly re-duces the number of permutation changes in successive bins. However, for a given microphone set, it is not uncommon that ICA will fail to separate some bins, thus defeating CII. This problem is addressed as follows. 1) In addition to CII the inter-frequency correlation matrix of the separated signals is used to align permutations in successive frequency bins. 2) The condition number of this matrix is monitored to determine if separation has failed for the current bin and microphone set. 3) If so, an alter-nate set with better separation is selected and again, inter-frequency correlation is used to align the permutations of the new set of microphones with the old. Results show a marked im-provement in separation when there are three or more sources.

## OS.13-BioSPS.2 Signal Processing Aspects of Brain Computer/Machine Interfaces

Session Chairs: Toshihisa Tanaka, Yodchanan Wongsawat Location: Trousdale Estates

### Toward Multi-Command Auditory Brain Computer Interfacing Using Speech Stimuli

Shuho Yoshimoto *University of Electro-Communications*, Yoshikazu Washizawa *University of Electro-Communications*, Toshihisa Tanaka *Tokyo University of Agriculture and Technology*, Hiroshi Higashi *Tokyo University of Agriculture and Technology*, Jun Tamura *Nara Institute of Science and Technology*

Brain-computer interfaces (BCIs) based on eventrelated potentials (ERP) are promising tools to communicate with patients suffering from some severe disabled diseases. ERP is evoked by various stimuli such as auditory, olfactory, and visual stimuli. Some auditory based BCIs using certain synthetic tone have been proposed, however, it is still challenging to increase the number of commands in auditory-based BCIs, since it is usually difficult for users to remember and distinguish multiple tones that corresponds to commands. We propose a new auditory BCI framework using speech stimuli. It is easier for users to distinguish different speech stimuli than different simple tones. We show experimental results of four-command BCI. The proposed speech-based BCI achieved a classification accuracy of more than 70 percents.

### EEG Energy Analysis Based On MEMD With ICA Pre-Processing

Yunchao Yin *Saitama Institute of Technology*, Cao Jianting *Saitama Institute of Technology*, Toshihisa Tanaka *Tokyo University of Agriculture and Technology*

Analysis of EEG energy is a useful technique in the brain signal processing. This paper presents a data analysis method based on multivariate empirical mode decomposition (MEMD) with ICA pre-processing to calculate and evaluate the energy of EEG recorded from the quasi brain deaths. The main advantage of introducing ICA pre-processing is that we can reduce the noise and other unexpected components. The simulation results illustrate the effectiveness and performance of the proposed method in brain death determination.

### Auditory Steady-State Response Stimuli Based BCI application - The Optimization of the Stimuli Types and Lengths

Yoshihiro Matsumoto *University of Tsukuba*, Tomek Rutkowski *University of Tsukuba*

We propose a method for an improvement of auditory BCI (aBCI) paradigm based on a combination of ASSR stimuli optimization by choosing the subjects' best responses to AM-, flutter-, AM/FM and click-envelope modulated sounds. As the ASSR response features we propose pairwise phaseÐlockingÐ values calculated from the EEG and next classified using binary classifier to detect attended and ignored stimuli. We also report on a possibility to use the stimuli as short as half a second, which is a step forward in ASSR based aBCI. The presented results are helpful for optimization of the aBCI stimuli for each subject.

### On the Classification of EEG/HEG-based Attention Levels Via Time-Frequency Selective Multilayer Perceptron For BCI-based Neurofeedback System

Supassorn Rodrak *Mahidol University*, Yodchanan Wongsawat *Mahidol University*

Attention Deficit/Hyperactivity Disorder (ADHD) is a neurobehavioral disorder which leads to the difficulty on focusing, paying attention and controlling normal behavior. Globally, the prevalence of ADHD is estimated to be 6.5%. Medicine has been widely used for the treatment of ADHD symptoms, but the patient may have a chance to suffer from the side effects of drug, such as vomit, rash, urticarial, cardiac arrthymia and insomnia. In this paper, we propose the alternative medicine system based on the brain-computer interface (BCI) technology called neurofeedback. The proposed neurofeedback system simultaneously employs two important signals, i.e. electroencephalogram (EEG) and hemoencephalogram (HEG), which can quickly reveal the brain functional network. The treatment criteria are that, for EEG signals, the patient needs to maintain the beta activities (13-30 Hz) while reducing the alpha activities (7-13 Hz). Simultaneously, HEG signals need to be maintained continuously increasing to some setting thresholds of the brain blood oxygenation levels. Time-frequency selective multilayer perceptron (MLP) is employed to capture the mentioned phenomena in real-time. The experimental results show that the proposed system yields the sensitivity of 98.16% and the specificity of 95.57%. Furthermore, from the resulting weights of the proposed MLP, we can also conclude that HEG signals yield the most impact to our neurofeedback treatment followed by the alpha, beta, and theta activities, respectively.

### Minimal-Assisted SSVEP-based Brain-Computer Interface Device

Yunyong Punsawad *Mahidol University*, Yodchanan Wongsawat *Mahidol University*

Steady-state visual evoked potential (SSVEP)-based brain computer interface (BCI) device is one of the most accurate assistive technologies for the persons with severe disabilities. However, for the existing systems, the persons with disabilities still need the assistance for the long period of time as well as the continuous time usages. In order to minimize this problem, we propose the SSVEP-based BCI system that the persons with disabilities can enable /disable the BCI device by alpha band EEG and control the electrical devices by SSVEP. A single-channel EEG (O1 or O2) is employed. Power spectral density via periodogram at the four stimulated frequencies (6, 7, 8, and 13 Hz) and their harmonics are used as the features of interest. Simple threshold-based decision rule is applied to the selected features. With the minimal need for assistance, the classification accuracy of the proposed system ranged from 75 to 100%.

### The Spatial Real and Virtual Sound Stimuli Optimization for the Auditory BCI

Nozomu Nishikawa *University of Tsukuba*, Tomek Rutkowski *University of Tsukuba*

The paper presents results from a project aiming to create horizontally distributed surround sound sources and virtual sound images as auditory BCI (aBCI) stimuli. The purpose is to create evoked brain wave response patterns depending on attended or ignored sound directions. We propose to use a modified version of the vector based amplitude panning (VBAP) approach to achieve the goal. The so created spatial sound stimulus system for the novel oddball aBCI paradigm allows us to create a multiÐcommand experimental environment with very encouraging results reported in this paper. We also present results showing that a modulation of the sound image depth changes also the subject responses. Finally, we also compare the proposed virtual sound approach with the traditional one based on real sound sources generated from the real loudspeaker directions. The so obtained results confirm the hypothesis of the possibility to modulate independently the brain responses to spatial types and depths of sound sources which allows for the development of the novel multi-command aBCI.

## OS.14-IVM.6 Visual Signal Compression

Session Chair: Gwo Giun Lee Location: Franklin Hills

### Image Set Modeling by Exploiting Temporal-Spatial Correlation and Photo Album Compression

Ruobing Zou *HKUST*, Oscar C. Au *HKUST*, Guyue Zhou *HKUST*, Sijin Li *HKUST*, Lin Sun *HKUST*

With the advance of digital photographing technology, large amount of personal photos are created and stored online or in personal computers. To save storage space and transmission bandwidth, we proposed a new photo album compression scheme by reducing both intra- and inter-image redundancy. Specifically, we first cluster a collection of images into groups each of which contains a set of similar images. Under a proposed graph framework, an optimal compression structure is derived from each cluster by finding the minimum spanning tree (MST) at a minimum prediction cost. The MST is trimmed for compression. Then by High Efficiency Video Coding (HEVC), the album is compressed as a whole and every cluster as a group of pictures (GOP), according to the predictive order of the optimal structures. Our experimental results show that there was around 60% improvement over only using JPEG compression.

### Largest Coding Unit Based Framework for Non-local Means Filter

Masaaki Matsumura *NTT Corporation*, Seishi Takamura *NTT Corporation*, Atsushi Shimizu *NTT Corporation*

One of the important factors for in-loop filter of video codec is low-delay capability for encoding and decoding. In this paper, we employ non-local means filter between sample adaptive offset and adaptive loop filter to the reference software of High Efficiency Video Coding HM7.0, and propose largest coding unit (LCU) based framework for non-local means filter that can reconstruct a decoded picture in LCU order at encoder and decoder. As the result, compared to HM7.0 anchor, in the case of picture-based RD-optimization, the average improvements of BD-rate for luma and chroma are 0.36 to 1.52% and 0.04 to 1.37%, respectively. Similarly, LCU-based one improves 0.20 to 1.27% and 0.67 to 1.91%, respectively. We confirm the maximum gain in the sequence of "Kimono" on low-delay P; the gains are 3.50% (Y), 2.89% (U) and 1.84% (V), respectively. Subjective quality improvements are also observed.

### A Five-stage Pipeline Design of Binary Arithmetic Encoder in H.264/AVC

Rui Song *Xidian University*, Hongfei Cui *Xidian University*, Yunsong Li *Xidian University*, Song Xiao *Xidian University*

Context-based Adaptive Binary Arithmetic Coding (CABAC) is a well known bottleneck in H.264/AVC encoder. Despite its high performance, the tight feedback loops make it difficult to parallelize. Most researchers are concerned about multi-bin processing regardless of the pipeline design. But without pipeline, the overall performance is greatly limited. In this paper, the critical path for hardware implementation of binary arithmetic encoder (BAE) was analyzed in detail. We break the computing steps to the best extent, and re-arrange it to the appropriate pipeline to get a balanced latency at each stage. Further, new binary arithmetic encoder architecture with five stage pipeline and 1 bin per cycle was proposed, the latency of critical path were cut off exceedingly, and the frequency and throughput rate was improved. An FPGA implementation of the proposed pipelined architecture in our H.264 encoder is capable of 190Mbps encoding rate. And a maximum 483MHz could be achieved on SMIC 0.13${\mu}m$ technology, which meets the requirements of QFHD encoding at 30fps. The proposed architecture could be utilized in other designs to get a better performance.

### Architecture of High-throughput Context Adaptive Variable Length Coding Decoder in AVC/H.264

Gwo Giun (Chris) Lee *National Cheng Kung University*, Shu-Ming Xu *National Cheng Kung University*, Chun-Fu Chen *National Cheng Kung University*, Ching-Jui Hsiao *National Cheng Kung University*

In this paper, a High-throughput Context Adaptive Variable Length Coding decoder which is capable for supporting AVC/H.264 HP@level 4.2 has been presented. To increase throughput, multi-symbol decoders for LEVEL and "RunBefore" and architecture of fast zero insertion are presented to reduce processing cycles to reach high-throughput rate. Finally, the experimental results show that the throughput of presented Context Adaptive Variable Length Coding decoder achieves the level limitation of level 4.2 in AVC/H.264 and the synthesis result shows that the gate count is about 17.2K gates at a clock constrain of 108MHz.

### An Inter-Frame/Inter-View Cache Architecture Design for Multi-View Video Decoders

Jui-Sheng Lee *National Chiao Tung University*, Sheng-Han Wang *National Chung-Cheng University*, Chih-Tai Chou *National Chung-Cheng University*, Cheng-An Chien *National Chung-Cheng University*, Hsiu-Cheng Chang *National Chung-Cheng University*, Jiun-In Guo *National Chiao Tung University*

In this paper we propose a low-bandwidth two-level inter-frame/inter-view cache architecture for a view scalable multi-view video decoder, which adopts two decoder cores to decode multi-view videos in parallel. The first level L1 cache is developed for the single video decoder core, which is able to reduce 60% bandwidth in doing inter-frame prediction in average. Moreover, we develop the second level L2 cache architecture to reuse the same reference data for doing inter-view prediction among different decoder cores, which can further reduce 35% bandwidth. By adopting the proposed two-level cache architecture for doing inter-frame/inter-view prediction, we can reduce 80% bandwidth through a view scalable multi-view video decoder implementation, which achieves real-time HD1080 dual-view video decoding.

## OS.15-SLA.7 Speech Recognition (II)

Session Chair: Seiichi Nakagawa Location: Whitley Heights

### Acoustic Model Training Using Committee-Based Active and Semi-Supervised Learning for Speech Recognition

Tsutaoka Takuya *Tokyo Institute of Technology*, Koichi Shinoda *Tokyo Institute of Technology*

We propose an acoustic model training method which combines committee-based active learning and semi-supervised learning for large vocabulary continuous speech recognition. In this method, each untranscribed training utterance is examined by a committee consists of multiple speech recognizers and the degree of disagreement in the committee on its transcription is used for selecting utterances. Those utterances the committee members disagree with each other are transcribed for active learning, while those they agree are used for semi-supervised learning. Our method was evaluated using the Corpus of Spontaneous Japanese. It was shown that it achieved higher recognition accuracy with lower transcription costs than random sampling, active learning alone, and semi-supervised learning alone. We also propose an alternative data selection method in semi-supervised learning.

### Distance Attenuation Control of Spherical Loudspeaker Array

Shigeki Miyabe *University of Tsukuba*, Takaya Hayashi *University of Tsukuba*, Takeshi Yamada *University of Tsukuba*, Shoji Makino *University of Tsukuba*

This paper describes control of distance attenuation using spherical loudspeaker array. Fisher et al. proposed radial filtering with spherical microphone to control the sensitivity to distance from a sound source by modeling the propagation of waves in spherical harmonic domain. Since transfer functions are not changed by swapping their inputs and outputs, we can use the same theory of radial filtering for microphone arrays to the filter design of distance attenuation control with loudspeaker arrays. Experimental results confirmed that the proposed method is effective in low frequencies.

### Recognition of Utterances with Grammatical Mistakes based on Optimization of Language Model towards Interactive CALL Systems

Takuya Anzai *Tohoku University*, Akinori Ito *Tohoku University*

To realize a voice-interactive CALL system, it is necessary to recognize the learner's utterance correctly including the grammatical mistakes. In this paper, we proposed methods for improving recognition accuracy of speech with grammatical mistakes. The proposed method is based on the method that uses n-gram model trained from sentences that are generated using grammatical error rules. We introduced two improvements to the previous method: one is the utterance discrimination to avoid introducing errors into correct utterances, and the other one is optimization of language model where probability of grammatical mistakes in the generated training text is optimized using the score of utterance discrimination. As a result, we obtained 0.92 point improvement, which is 12% error reduction.

### Fast NMF Based Approach and Improved VQ Based Approach for Speech Recognition From Mixed Sound

Shoichi Nakano *Toyohashi University of Technology*, Kazumasa Yamamoto *Toyohashi University of Technology*, Seiichi Nakagawa *Toyohashi University of Technology*

We have considered a speech recognition method for mixed sound, consisting of speech and music, that removes only the music based on vector quantization (VQ) and non-negative matrix factorization (NMF). This paper describe fast calculation technique of music removal based on NMF and improvement using a VQ method. For isolated word recognition using the clean speech model, an improvement of 46% word error reduction rate was obtained compared with the case of not removing music. Furthermore, a high recognition rate, close to clean speech recognition was obtained at 10 dB. For the case of the multi-conditions, our proposed method reduced the error rate of 50% compared with the multi-conditions model.

### Expansion of Training Texts to Generate a Topic-Dependent Language Model for Meeting Speech Recognition

Kazushige Egashira *Nagasaki University*, Kazuya Kojima *Nagasaki University*, Masaru Yamashita *Nagasaki University*, Katsuya Yamauchi *Nagasaki University*, Shoichi Matsunaga *Nagasaki University*

This paper proposes expansion methods for training texts (baseline) to generate a topic-dependent language model for more accurate recognition of meeting speech. To prepare a universal language model that can cope with the variety of topics discussed in meetings is very difficult. Our strategy is to generate topic-dependent training texts based on two methods. The first is text collection from web pages using queries that consist of topic-dependent confident terms; these terms were selected from preparatory recognition results based on the TF-IDF (TF; Term Frequency, IDF; Inversed Document Frequency) values of each term. The second technique is text generation using participants' names. Our topic-dependent language model was generated using these new texts and the baseline corpus. The language model generated by the proposed strategy reduced the perplexity by 16.4% and out-of-vocabulary rate by 37.5%, respectively, compared with the language model that used only the baseline corpus. This improvement was confirmed through meeting speech recognition as well.

## OS.16-WCN.2 Wireless Communications and Networking (I)

Session Chair: Ioannis Katsavounidis Location: Mt. Olympus

### Packet Loss Rate Estimation with Active and Passive Measurements

Atsushi Miyamoto *Nara Institute of Science and Technology*, Kazuho Watanabe *Nara Institute of Science and Technology*, Kazushi Ikeda *Nara Institute of Science and Technology*

Network tomography is a problem of estimating network properties such as the packet loss rates of links using available packets. There are two kinds of methods to measure packets: active and passive. An active measurement specifies link information (paths) of packets a priori while a passive measurement gets only the origins and destinations of packets. The conventional methods for estimating the packet loss rate of each link, one of the network tomography problems, utilize only active measurements because passive measurements have no link information. We propose a method to utilize passive measurements also. The method regards the link information in the passive measurements as latent variables and estimates the variables and the loss rates of links simultaneously in the framework of Bayesian inference. We show through numerical experiments that our method outperforms the conventional algorithm with only active measurements in the estimation accuracy.

### Investigating Wireless Sensor Network Lifetime under Static Routing with Unequal Energy Distribution

Apostolis Xenakis *University of Thessaly*, Ioannis Katsavounidis *University of Thessaly*, George Stamoulis *University of Thessaly*

In a Wireless Sensor Network (WSN) the sensed data must be gathered and transmitted to a base station where it it further processed by end users. Since that kind of network consists of low-power nodes with limited battery power, power efficient methods must be applied for nodes communication and data gathering in order to achieve long network lifetimes. In such networks where in a round of communication each of the sensor nodes has data to send to a base station, it is very important to minimize the total energy consumed by the system in a round so that the total network lifetime is maximized. The lifetime of such sensor network is the time until base station can receive data from all sensors in the network. In this paper, besides the conventional protocol of direct transmission or the use of dynamic routing protocols proposed in literature that aggregates data, we propose an algorithm based on static routing among sensor nodes with unequal energy distribution in order to extend network lifetime and find a near optimal node energy charge scheme that leads to both node and network lifetime prolong. Our simulation results show that our algorithm achieve longer network lifetimes mainly because each node is free from maintaining complex route information, less infrastructure communication is needed and the charge of nodes is not uniform.

### Efficient Algorithm with Lognormal Distributions for Overloaded MIMO Wireless Systems

Kazi Obaidullah *Hokkaido University*, Yoshikazu Miyanaga *Hokkaido University*

Due to outstanding search strength and well organized steps, genetic algorithm (GA) has gained high interest in the field of overloaded multiple-input/multiple-output (MIMO)wireless communications system. For overloaded MIMO system employing spatial multiplexing transmission we evaluate the performance and complexity of genetic algorithm (GA)-based detection, against the maximum likelihood (ML) approach. We consider transmit-correlated fading channels with realistic Laplacian power azimuth spectrum. The values of the azimuth spread (AS) and Rician K-factor are set by the means of the lognormal distributions obtained from WINNER II channel models. First, we confirm that for constant complexity, GA performance is same for different combinations of GA parameters. Then, we compare the GA performance with ML in several WINNER II scenarios and channel matrix means. Finally, we compare the complexity of GA with ML. We find that GA perform similarly with ML throughout the SNR points for different scenarios and different deterministic rank. We also find that for achieving performance, GA complexity is much less than ML and thus, is an advantage in field programmable gate array (FPGA) design.

**Location Based Relay Selection Optimization in Mobile Cooperative Environment**

Esam Obiedat *CommScope Inc.*, Chirag Warty *Stanford University*, Lei Cao *University of Mississippi*

This paper proposes an optimal relay selection criteria based on the location of the relay (0 $^2$ _ $^2$ 1) relative to source and destination in the cooperative coded system. The proposed optimization algorithm employs distributed turbo product coding technique with hard and soft decoding. It is shown that the link quality depends on the location of the relay which in turn affects overall system Bit Error Rate (BER) performance. The simulation model creates several scenarios for location of intermediate relays when the inter-user channel is experiencing distortion in presence of different Signal to Noise Ratio (SNR)s. It is observed that the link performance degrades as the relay proximity changes with respect to the source and the destination. The relay selection optimization algorithm provides the participating nodes necessary information to select neighboring nodes depending on link quality, thus lowering the BER and increasing the overall network capacity.

**A Real-time Streaming Media Transmission Protocol for Multi-hop Wireless Networks**

Jianchao Du *Xidian University*, Song Xiao *Xidian University*, Lei Quan *Xidian University*

Time delay as well as error accumulation makes transmission of high-quality streaming media over multi-hop wireless networks more challenging. Since conventional TCP/IP-based protocol drops and retransmits the whole packet once error(s) occur above Physical Layer, which leads to long time delay and low efficiency in transmission, a new protocol for real-time streaming media transmission is proposed in this paper. A packet control layer (PCL) is added between Data Link and Network Layer to enable error-polluted data be transmitted continuously while an error control layer (ECL) is inserted between Transport and Application Layer to further correct errors in data stream. Moreover, a robust header conversion method is applied to PCL to shorten time delay by reducing packet retransmission probability and decrease the redundancy of packet header. And an error-CRC-erasure coding scheme embedded with CRC error correcting algorithm is adopted in ECL. Simulations on DSP show that the proposed protocol can greatly reduce time delay and obtain better error correction performance compared with traditional protocol in BSC channel.

**Optimal Bit Allocation of Limited Rate Feedback for Cooperative Jamming**

Xinjie Yang *University of California, Irvine*, A. Lee Swindlehurst *University of California, Irvine*

In this paper, we investigate bit allocation schemes with limited rate feedback for cooperative jamming. In addition to the transmitter and receiver, we assume a passive eavesdropper and cooperative jammer are present. In order to achieve a secure communications link against the eavesdropper, the transmitter and jammer require channel state information (CSI) to be fed back to them from the receiver. Assuming feedback channels with a maximum sum feedback rate constraint, the receiver must allocate the total number of bits available to quantize the CSI between the transmitter and jammer. This requires the receiver to balance the need for a strong channel from the transmitter against the need for the jammer to accurately null the receiver and reduce the resulting interference. We propose an optimal bit allocation strategy for this problem using mean-squared error as the performance metric, and we use simulation examples to illustrate its advantage over a non-optimized feedback allocation.

## PS.3-IVM.7 Selected Topics in Computer Vision and Multimedia

Session Chair: Mark Liao Location: Solano

**An Algorithm for Radar Power Line Detection with Tracking**

Qirong Ma *University of Washington*, Darren Goshi *Honeywell Corporation*, Long Bui *Honeywell Corporation*, Ming-Ting Sun *University of Washington*

In this paper we deal with the problem of power line detection from millimeter-wave radar video. We propose an algorithm that is based on Hough Transform, Support Vector Machine, and particle filter tracking. We explore the defining characteristics of the power lines in the radar video, and present an approach to utilize these characteristics together with the temporal correlation property of the power line objects. The particle filter framework naturally captures the temporal correlation of the power line objects, and the power-line-specific feature is embedded into the conditional likelihood measurement process of the particle filter. Experimental result validates the effectiveness of the power line detection approach.

**Multiple Exposure Integration with Image Denoising**

Ryo Matsuoka *The University of Kitakyushu*, Masahiro Okuda *The University of Kitakyushu*, Takao Jinno *The University of Kitakyushu*

We propose a denoising technique for multiple exposure image integration. In our method, noise removal is achieved by the wavelet-shrinkage for multiple exposures, and a novel weighting scheme for the integration. A weighted image is converted to the low and the high frequency elements by the shift invariant wavelet transform, and the wavelet coefficient in the high frequencies are decreased by thresholding based on the wavelet-based hard shrinkage. The weight is designed to reduce sensor noise and quantization noise in the process of the multiple exposure integration. Our method works well especially for noise in shadow areas. We show the validity of the proposed algorithm by simulating the method with some actual noisy images.

**Palmprint Verification using Gradient Maps and Support Vector Machines**

Chun-Wei Lu *Academia Sinica*, Ivy Fan *Academia Sinica*, Chin-Chuan Han *National United University*, Jyh-Chian Chang *Chinese Culture University*, Kuo-Chin Fan *National Central University*, Hong-Yuan Liao *Academia Sinica*

With the urgent demand in information security, biometric feature-based verification systems have been extensively explored in many application domains. However, the efficacy of existing biometric-based systems is unsatisfactory and there are still a lot of difficult problems to be solved. Among many existing biometric features, palmprint has been regarded as a unique and useful biometric feature due to its stable principal lines. In this paper, we proposed a new method to perform palmprint recognition. We extract the gradient map of a palmprint and then verify it by a trained support vector machine (SVM). The procedure can be divided into three steps, including image preprocessing, feature extraction, and verification. We used the multi-spectral palmprint database prepared by Hong Kong PolyU [14] which included 6000 palm images collected from 250 individuals to test our method. The experimental results demonstrate our proposed method is reliable and efficient to verify whether the person is genuine or not.

**OpenQoS: An OpenFlow Controller Design for Multimedia Delivery with End-to-End Quality of Service over Software-Defined Networks**

Hilmi Egilmez *Koc University*, S. Tahsin Dane *Koc University*, K. Tolga Bagci *Koc University*, A. Murat Tekalp *Koc University*

OpenFlow is a Software Defined Networking (SDN) paradigm that decouples control and data forwarding layers of routing. In this paper, we propose OpenQoS, which is a novel OpenFlow controller design for multimedia delivery with end-to-end Quality of Service (QoS) support. Our approach is based on QoS routing where the routes of multimedia traffic are optimized dynamically to fulfill the required QoS. We measure performance of OpenQoS over a real test network and compare it with the performance of the current state-of-the-art, HTTP-based multi-bitrate adaptive streaming. Our experimental results show that OpenQoS can guarantee seamless video delivery with little or no video artifacts experienced by the end-users. Moreover, unlike current QoS architectures, in OpenQoS the guaranteed service is handled without having adverse effects on other types of traffic in the network.

**GIF-LR:GA-based Informative Feature for Lipreading**

Naoya Ukai *Gifu University*, Takumi Seko *Gifu University*, Tamura Satoshi *Gifu University*, Hayamizu Satoru *Gifu University*

In this paper, we propose a general and discriminative feature GIF (GA-based Informative Feature), and apply the feature to lipreading (visual speech recognition). The feature extraction method consists of two transforms, that convert an input vector to GIF for recognition. The transforms can be computed using training data and Genetic Algorithm (GA). For lipreading, we extract a fundamental feature as an input vector from an image; the vector consists of intensity values at all the pixels in an input lip image, which are enumerated from left-top to right-bottom. Recognition experiments of continuous digit utterances were conducted using an audio-visual corpus including more than 268,000 lip images. The recognition results show that the GIF-based method is better than the baseline method using eigenlip features.

**One-to-N Wireless Power Transmission System Based on Multiple Access One-Way In-Band Communication**

Dong-Zo Kim *Samsung Electronics*, Ki young Kim *Samsung Advanced Institute of Technology*, Nam Yoon Kim *Samsung Advanced Institute of Technology*, Yun-Kwon Park *Samsung Advanced Institute of Technology*, Wang-Sang Lee *KAIST*, Jong-Won Yu *KAIST*, Sangwook Kwon *Samsung Advanced Institute of Technology*

For an efficient wireless charging to multiple devices, transmitter (TX) system should be able to control the transmitting power-level by monitoring number of receiver (RX) devices via their identification information and charging status and capacity of each receivers. This work proposes a one-way in-band communication scheme, corresponding system structure, and charging algorithm for efficient one-to-N wireless charging system, which has been verified experimentally.

## PS.4-BioSPS.3 Biomedical Signal Processing and Systems

Session Chair: Bonnie Law Location: Solano

**Comparative Study of Interactive Seed Generation for Growcut-Based Fast 3D MRI Segmentation**

Toshihiko Yamasaki *The University of Tokyo*, Tsuhan Chen *Cornell University*, Masakazu Yagi *Oaska University*, Toshinori Hirai *Kumamoto University*, Ryuji Murakami *Kumamoto University*

This paper proposes a speed-enhanced growcut method and presents comparative study of seed setting methods for fast 3D medical image (MRI) segmentation. The processing time tends to be larger in 3D image segmentation because of the large number of neighboring voxels as well as the number of voxels themselves. In this paper, two seed setting methods are proposed for our fast growcut-based segmentation algorithm: sphere-based bounding box method and label transfer based method using SIFT flow. Experimental results demonstrate that the tumor segmentation for each patient can be done very quickly as compared to the previous works. The segmentation accuracy can also be made very high with only a few user interactions.

**Compressed Sensing with Super-resolution in Magnetic Resonance using Quadratic Phase Modulation**

Satoshi Ito *Utsunomiya University*, Yoshifumi Yamada *Utsunomiya University*

In recent years, compressed sensing (CS) has attracted considerable attention in areas of rapid MR imaging. Our group and Y. Wiaux have shown independently that the use of quadratic phase modulation prior to data acquisition can greatly improve the accelerating factor of CS. The use of quadratic phase modulation has distinctive features that the extrapolation of signal by post processing calculation is feasible. In this paper, we propose a novel image reconstruction method in which extrapolation of signal is executed in the CS reconstruction algorithm, resulting in the improvement of spatial resolution. Simulation and experimental studies have revealed that the spatial resolution is fairly improved compared to the images obtained in standard CS based on Fourier transform imaging

**Exploiting Biclustering for Missing Value Estimation in DNA Microarray Data**

Kin-On Cheng *The Hong Kong Polytechnic University*, Bonnie Law *The Hong Kong Polytechnic University*, Wan-Chi Siu *The Hong Kong Polytechnic University*

The missing values in gene expression data harden subsequent analysis such as biclustering which aims to find a set of coexpressed genes across a number of experimental conditions. Missing values are thus required to be estimated before biclusters detection. Existing estimation algorithms rely on finding coherence among expression values throughout the entire genes and/or across all the conditions. In view that both missing values estimation and biclusters detection aim at exploiting coherence inside the expression data, we propose to integrate them into a single framework. The benefits are twofold, the missing value estimation can improve bicluster analysis and the coherence in detected biclusters can be exploited for better missing value estimation. Experimental results show that the integrated framework outperforms existing missing values estimation algorithms. It reduces error in missing value estimation and facilitates the detection of biologically meaningful biclusters.

**A Breast Tumor Classification Method based on Ultrasound BI-RADS Data Mining**

Jin Man Park *Samsung Electronics*, Hyoungmin Park *Samsung Electronics*, Jong-Ha Lee *Samsung Electronics*, Yeong Kyeong Seong *Samsung Electronics*, Kyoung-Gu Woo *Samsung Electronics*, Kyuseok Shim *Seoul National University*

In this paper, we propose a data analysis method to select important characteristics of ultrasonic breast images which suggest the malignancy of breast tumor. Based on the analysis, we also present a method of creating a classifier that can quickly and precisely predict the malignancy of breast tumors from an ultrasonic breast image. The selection of important characteristics enables us to focus on the image processing algorithms that can effectively represent the selected characteristics. By applying the data analysis method to more than 5,000 clinical cases, we select a subset of image processing algorithms which have better representative power for important characteristics. Our classifier based on the subset of image processing algorithms shows a comparable accuracy to a naive classifier which uses a full set of the image processing algorithms. Thus, our classifier can reduce the prediction time on demand by minimizing the number of image processing algorithms. The experiments show that the malignancy of tumor could be successfully predicted by our approach.

**An AdaBoost-Based Weighting Method for Localizing Human Brain Magnetic Activity**

Tetsuya Takiguchi *Kobe University*, Ryoichi Takashima *Kobe University*, Yasuo Ariki *Kobe University*, Toshiaki Imada *University of Washington*, Lotus Lin *University of Washington*, Patricia Kuhl *University of Washington*, Masaki Kawakatsu *Tokyo Denki*

*University*

This paper shows that pattern classification based on machine learning is a powerful tool for analyzing human brain activity data obtained by magnetoencephalography (MEG). In our previous work, a weighting method using multiple kernel learning was proposed, but this method had a high computational cost. In this paper, we propose a novel and fast weighting method using an AdaBoost algorithm to find the sensor area contributing to the accurate discrimination of vowels. Our AdaBoost simultaneously estimates both the classification boundary and the weight to each MEG sensor, with MEG amplitude obtained from each pair of sensors being an element of the feature vector. The estimated weight indicates how the corresponding sensor is useful for classifying the MEG response patterns. Our results for vowel recognition show the large-weight MEG sensors mainly in a language area of the brain and the high classification accuracy (91.0%) in the latency range between 50 and 150 ms.

## 16:30 - 18:10

### OS.17-SLA.8 Speech Processing (II)

Session Chair: Waleed Abdulla Location: Doheny

**Optimizing the Parameters of Decoding Graphs Using New Log-based MCE**

Abdelaziz Abdelhamid *The University of Auckland*, Waleed Abdulla *The University of Auckland*

This paper proposes a new class loss function as an alternative to the standard sigmoid class loss function for optimizing the parameters of decoding graphs using discriminative training based on minimum classification error (MCE) criterion. The standard sigmoid based approach tends to ignore a significant number of training samples that have a large difference between the scores of the reference and their corresponding competing hypotheses and this affects the parameters optimization. The proposed function overcomes this limitation through considering almost all the training samples and thus improved the parameter optimization when tested on large decoding graphs. The decoding graph used in this research is an integrated network of weighted finite state transducers. The primary task examined is 64K words, continuous speech recognition task. The experimental results show that the proposed method outperformed the baseline system based on both the maximum likelihood estimation (MLE) and sigmoid-based MCE and achieved a reduction in the word error rate (WER) of 28.9% when tested on the TIMIT speech database.

**Voice Activity Detection Based on Augmented Statistical Noise Suppression**

Yasunari Obuchi *Hitachi Ltd.*, Ryu Takeda *Hitachi Ltd.*, Naoyuki Kanda *Hitachi Ltd.*

A new voice activity detection (VAD) algorithm using augmented statistical noise suppression is introduced. Statistical noise suppression is an effective tool for speech processing under noisy conditions. It achieves the best VAD performance when the noise suppression is augmented in various ways. The speech distortion, which is usually a severe side effect of strong noise suppression, does not affect the VAD performance, and the correctly estimated signal power provides accurate detection of speech. The performance of the proposed algorithm is evaluated using CENSREC-1-C public database, and it is confirmed that the proposed algorithm outperforms other algorithms such as the switching Kalman filter-based VAD.

**Language Modeling for Spoken Dialogue System based on Sentence Transformation and Filtering using Predicate-Argument Structures**

Koichiro Yoshino *Kyoto University*, Shinsuke Mori *Kyoto University*, Tatsuya Kawahara *Kyoto University*

We present a novel scheme of language modeling for a spoken dialogue system by effectively exploiting the back-end documents the system uses for information navigation. The proposed method first converts sentences in the document, which are written and plain style, into spoken question-style queries, which are expected in spoken dialogue. In this process, we conduct dependency analysis to extract verbs and relevant phrases to generate natural sentences by applying transformation rules. Then, we select sentences which have useful information relevant to the target domain and thus are more likely to be queried. For this purpose, we define predicate-argument (P-A) templates based on a statistical measure in the target document. An experimental evaluation shows that the proposed method outperforms the conventional method in ASR performance, and the sentence selection based on the P-A templates is effective.

**Hybrid Vector Space Model for Flexible Voice Search**

Cheongjae Lee *Kyoto University*, Tatsuya Kawahara *Kyoto University*

This paper addresses incorporation of semantic analysis into information retrieval (IR) based on the vector space model (VSM) for flexible matching of spontaneous queries in a voice search system. Information of semantic slots or concepts that correspond to database fields is expected to help enhancing IR, but the semantic analyzer often fails or needs a large amount of training data. We propose a hybrid model which combines dedicated VSMs for concept slots with a general VSM as a back-off. The model has been evaluated in a book search task and shown to be effective and robust against ASR and SLU errors.

**An Interference-Free Representation of Group Delay for Periodic Signals**

Hideki Kawahara *Wakayama University*, Masanori Morise *Ritsumeikan University*, Ryuichi Nisimura *Wakayama University*, Toshio Irino *Wakayama University*

This article introduces a new group delay representation for periodic signals. The proposed method yields a group delay representation that is free from interferences due to repetitive excitation. Power spectrum-weighted averaged group delay using shifted copies of the weighted group delay separated by a half fundamental frequency is proven to have the desired property.

### OS.18-SLA.9 Audio & Music Processing (II)

Session Chair: Chang-chun Bao Location: Beachwood

**A Blind Bandwidth Extension Method of Audio Signals based on Volterra Series**

Xing-tao Zhang *Beijing University of Technology*, Chang-chun Bao *Beijing University of Technology*, Xin Liu *Beijing University of Technology*, Li-yan Zhang *Beijing University of Technology*

In this paper, a blind bandwidth extension method of audio signals is proposed in which the fine structure of high-frequency information is recovered based on Volterra series. Combining with Gaussian mixture model and codebook mapping to adjust the spectrum envelope and energy gain of the extended high-frequency components separately, the bandwidth of audio signals is extended to super-wideband from wideband. Furthermore, the proposed method is applied into a real audio codec. The performance of the proposed method is evaluated through objective and subjective tests on the audio signals selected from MPEG items, and it is found that the proposed method outperforms the chaotic prediction method and nearest-neighbor matching method. When the proposed algorithm is applied into ITU-T G.722.1 wideband audio codec, the performance is comparable with that of G.722.1C super-wideband audio codec at 24 kbps.

### Personalized Music Emotion Recognition via Model Adaptation

Ju-Chiang Wang *Academia Sinica*, Yi-Hsuan Yang *Academia Sinica*, Hsin-Min Wang *Academia Sinica*, Shyh-Kang Jeng *National Taiwan University*

In the music information retrieval (MIR) research, developing a computational model that comprehends the affective content of music signal and utilizes such a model to organize music collections have been an essential topic. Emotion perception in music is in nature subjective. Consequently, building a general emotion recognition system that performs equally well for every user could be insufficient. In contrast, it would be more desirable for one's personal computer/device being able to understand his/her perception of music emotion. In our previous work, we have developed the acoustic emotion Gaussians (AEG) model, which can learn the broad emotion perception of music from general users. Such a general music emotion model, called the background AEG model in this paper, can recognize the perceived emotion of unseen music from a general point of view. In this paper, we go one step further to realize the personalized music emotion modeling by adapting the background AEG model with a limited number of emotion annotations provided by a target user in an online and dynamic fashion. A novel maximum a posteriori (MAP)-based algorithm is proposed to achieve this in a probabilistic framework. We carry out quantitative evaluations on a well-known emotion annotated corpus, MER60, to validate the effectiveness of the proposed method for personalized music emotion recognition.

### HRTF Magnitude Modeling Using a Non-Regularized Least-Squares Fit of Spherical Harmonics Coefficients on Incomplete Data

Jens Ahrens *Microsoft Research*, Mark Thomas *Microsoft Research*, Ivan Tashev *Microsoft Research*

Head-related transfer functions (HRTFs) represent the acoustic transfer function from a sound source at a given location to the ear drums of a human. They are typically measured from discrete source positions at a constant distance. Spherical harmonics decompositions have been shown to provide a flexible representation of HRTFs. Practical constraints often prevent the retrieval of measurement data from certain directions, a circumstance that complicates the decomposition of the measured data into spherical harmonics. A least-squares fit of coefficients is a potential approach to determining the coefficients of incomplete data. However, a straightforward non-regularized fit tends to give unrealistic estimates for the region were no measurement data is available. Recently, a regularized least-squares fit was proposed, which yields well-behaved results for the unknown region at the expense of reducing the accuracy of the data representation in the known region. In this paper, we propose using a lower-order non-regularized least-squares fit to achieve a well-behaved estimation of the unknown data. This data then allows for a high-order non-regularized least-squares fit over the entire sphere. We compare the properties of all three approaches applied to modeling the magnitudes of the HRTFs measured from a manikin. The proposed approach reduces the normalized mean-square error by approximately 7 dB in the known region and 11 dB in the unknown region compared to the regularized fit.

### Subjective Similarity of Music: Data Collection for Individuality Analysis

Shota Kawabuchi *Nagoya University*, Chiyomi Miyajima *Nagoya University*, Norihide Kitaoka *Nagoya University*, Kazuya Takeda *Nagoya University*

We describe a method of estimating subjective music similarity from acoustic music similarity. Recently, there have been many studies on the topic of music information retrieval, but there continues to be difficulty improving retrieval precision. For this reason, in this study we analyze the individuality of subjective music similarity. We collected subjective music similarity evaluation data for individuality analysis using songs in the RWC music database, a widely used database in the field of music information processing. A total of 27 subjects listened to pairs of music tracks, and evaluated each pair as similar or dissimilar. They also selected the components of the music (melody, tempo/rhythm, vocals, instruments) that were similar. Each subject evaluated the same 200 pairs of songs, thus the individuality of the evaluation can be easily analyzed. Using the collected data, we trained individualized distance functions between songs, in order to estimate subjective similarity and analyze individuality.

### Comparison of Superimposition and Sparse Models in Blind Source Separation By Multichannel Wiener Filter

Ryutaro Sakanashi *University of Tsukuba*, Shigeki Miyabe *University of Tsukuba*, Takeshi Yamada *University of Tsukuba*, Shoji Makino *University of Tsukuba*

Multichannel Wiener filter proposed by Duong et al. can conduct underdetermined blind source separation (BSS) with low distortion. This method assumes that the observed signal is the superimposition of the multichannel source images generated from multivariate normal distributions. The covariance matrix in each time-frequency slot is estimated by an EM algorithm which treats the source images as the hidden variables. Using the estimated parameters, the source images are separated as the maximum a posteriori estimate. It is worth nothing that this method does not assume the sparseness of sources, which is usually assumed in underdetermined BSS. In this paper we investigate the effectiveness of the three attributes of Duong's method, i.e., the source image model with multivariate normal distribution, the observation model without sparseness assumption, and the source separation by multichannel Wiener filter. We newly formulate three BSS methods with the similar source image model and the different observation model assuming sparseness, and we compare them with Duong's method and the conventional binary masking. Experimental results confirmed the effectiveness of all the three attributes of Duong's method.

## OS.19-IVM.8 Visual 3D Scene Reconstruction and its Applications

Session Chair: Kyoung Mu Lee Location: Runyon

### Violin Pedagogy for Finger and Bow Placement using Augmented Reality

Francois de Sorbier *Keio University*, Hiroyuki Shiino *Keio University*, Hideo Saito *Keio University*

Beginners need a long time before being able to play correctly violin. Learning the bowing technique appears to be a difficult task and retains most of the attention of beginners. Besides this point, the finger placement is also an important part of the learning but often under estimated. One difficulty is that the fingerboard of the violin does not have frets. In this on-going work, we present a marker-less augmented reality system that advises the novice players about their fingering and bowing. We display in real-time the virtual frets by tracking the violin with a depth camera. We also capture and recognize the note currently played to direct the placement of the bow on the strings.

### Robust View Synthesis Under Varying Illumination Conditions Using Segment-Based Disparity Estimation

Il-lyong Jung *Korea University*, Chang-Su Kim *Korea University*

An intermediate view synthesis scheme under varying illumination conditions is proposed in this work. First, we estimate the disparity map based on cumulative color histograms. Since the cumulative histogram of an image represents the brightness ranks of pixels, the disparity estimation is robust against varying illumination conditions. More specifically, we divide each image into segments, and compute the cumulative histogram of the representative values for these segments. Then, we estimate the disparity map based on the similarity of the cumulative histograms between stereo images. Second, we transform the colors of stereo images adaptively using the disparity map. Finally, we synthesize intermediate views using the transformed stereo images and the disparity map. Simulation results demonstrate that the proposed algorithm provides better disparity maps and intermediate views under varying illumination conditions than the conventional techniques.

### Memory-Efficient Belief Propagation in Stereo Matching on GPU

Young-kyu Choi *Inha University*, Williem *Inha University*, In Kyu Park *Inha University*

Belief propagation (BP) is a commonly used global energy minimization algorithm for solving stereo matching problem in 3D reconstruction. However, it requires large memory bandwidth and data size. In this paper, we propose a novel memory-efficient algorithm of BP in stereo matching on the Graphics Processing Units (GPU). The data size and transfer bandwidth are significantly reduced by storing only a part of the whole message. In order to maintain the accuracy of the matching result, the local messages are reconstructed using shared memory available in GPU. Experimental result shows that there is almost an order of reduction in the global memory consumption, and 21 to 46% saving in memory bandwidth when compared to the conventional algorithm. The implementation result on a recent GPU shows that we can obtain 22.8 times speedup in execution time compared to the execution on CPU.

### Real-Time Panorama Image Synthesis By Fast Camera Pose Estimation

Beom Su Kim *Seoul National University*, Sang Hwa Lee *Seoul National University*, Nam Ik Cho *Seoul National University*

This paper proposes a fast panorama synthesis algorithm that runs on a mobile devices real-time. Like most existing methods, the proposed method consists of following steps: feature tracking, rotation matrix estimation, and image warping on a targeting plane, where the feature tracking is usually a bottleneck for real-time implementation. Hence, we propose to track the features on a virtual sphere surface instead of projected surface or image domain as in the conventional methods. By performing the feature tracking on the sphere, the camera pose can be found by linear and non-iterative least squares method, which was usually obtained by nonlinear and iterative methods. The fast estimation of camera pose can make outlier rejection more robust since the camera pose can be inferred from the hypotheses by one iteration, which can't be done in real-time by iterative estimation. We also propose a two-step blending algorithm, i.e., celling-filling followed by linear blending along the cell boundary. The panorama canvas is partitioned into many cells where each cell contains pixels from the same shot. Hence there is no stitching seam within the cell and only the boundaries need to be blended, which reduces the stitching artifacts significantly.

### Combining Multi-view Stereo and Super Resolution in a Unified Framework

Haesol Park *Seoul National University*, Kyoung Mu Lee *Seoul National University*, Sang Uk Lee *Seoul National University*

In multi-view stereo setting, pixel correspondence problem and super resolution problem are inter-related in a sense that the result of each problem could help to solve the other. In this paper, we propose a novel method to solve two problems together by optimizing a unified energy functional. Main difference from the previous works is that the consistency between high resolution images is considered along with consideration to the consistency of high-resolution and low-resolution image pair with the same viewpoint. Experimental results show that our method outperforms the naive combination of single image super resolution and multi-view stereo method.

### Confidence-based Refinement of Corrupted Depth Maps

Satoshi Ikehata *University of Tokyo*, Kiyoharu Aizawa *University of Tokyo*

This paper present a practical depth-map refinement system designed for highly corrupted multiple depth maps. We define a pixel-wise confidence measurement of depth value and apply the three-steps depth-map refinement scheme (\ie confidence-based depth-map fusion, confidence-weighted bundle optimization and super-pixel-based planar propagation) to maximize the whole reliability of depth maps. Our experimental result shows that our refinement algorithm can dramatically improve highly corrupted depth maps acquired by previous approaches.

## OS.20-SIPTM.1 Control, Optimization and Information Processing for Smart Grid (I)

Session Chairs: Rongshan Yu, Binbin Chen Location: Laurel

### Distributed State Estimation in Smart Grid with Communication Constraints

Hang Ma *University of Maryland*, Yu-Han Yang *University of Maryland*, Yan Chen *University of Maryland*, K. J. Ray Liu *University of Maryland*

Distributed state estimation in smart grid highly relies on the availability of measurements. Transmitting a lot of measurements within a small time interval is costly and sometimes even impossible. This paper explores the problem of distributed state estimation in smart grid with constraint on the number of measurements that is able to be transmitted in one step. It is shown that there exists a lower bound which depends on the structure of the grid such that if the number of permissible measurements is beyond the bound, then the estimator achieves the same performance as its peer without the constraint. Further, if the number of permissible measurements is below the lower bound, a tradeoff between the performance of the estimator and the measurements transmitted is needed to meet the constraint. A method to attain the tradeoff is offered in this paper. The proposed conclusions and methods are illustrated in the simulation on the IEEE 14-bus system.

### Cyclostationary Noise Mitigation in Narrowband Powerline Communications

Jing Lin *University of Texas at Austin*, Brian Evans *University of Texas at Austin*

Future Smart Grid systems will intelligently monitor and control energy flows in order to improve the efficiency and reliability of power delivery. The monitoring and control require low-delay, highly reliable, two-way communications between customers, local utilities and regional utilities. Narrowband powerline communication (NB-PLC) systems operating in the 3--500 kHz band have been standardized to enable these two-way communication links. In NB-PLC systems, additive non-Gaussian noise/interference is primary limitation to the communication performance. From field trials, the dominant source of this non-Gaussian noise/interference is cyclostationary. In this paper, we address the problem of cyclostationary noise mitigation in NB-PLC systems and other orthogonal frequency division multiplexing

(OFDM) systems. The contributions of this paper include developing a parametric noise estimation algorithm based on switching linear autoregressive (AR) process, and a simple adaptive noise whitening approach that can be immediately integrated into the conventional OFDM transceiver structure to improve its performance. In our simulations, the proposed noise whitening method achieves up to 3dB SNR gain over conventional OFDM systems at SNRs higher than -3dB.

**Load Disaggregation Using Harmonic Analysis and Regularized Optimization**
Jerry Chiang *ADSC*, Tianzhu Zhang *ADSC*, Binbin Chen *ADSC*, Yih-chun Hu *University of Illinois at Urbana-Champaign*

In this paper, we present a load disaggregation technique that uses regularized optimization together with harmonic frequency signatures of appliances. The benefits of our technique are two fold: 1) The regularized optimization is faster than integer programming; and 2) The harmonic frequency signatures allow us to disaggregate the loads using as few as 10 cycles (equaling as little as 200~milliseconds) of samples, instead of having to wait for state changes from appliances or weekly usage pattern to emerge. We test our proposed technique in proof-of-concept experiments and show that our technique returns accurate disaggregation results.

**Dynamic Incentive Strategy for Voluntary Demand Response based on TDP Scheme**
Haiyan Shu *Institute for Infocomm Research*, Rongshan Yu *Institute for Infocomm Research*, Susanto Rahardja *Institute for Infocomm Research*

The enhanced real-time metering and communication capabilities from smart meters and their associated advanced metering infrastructure make it possible for utility company to extend demand response (DR) to small customers through time-dependent pricing (TDP). Considering the economic reason and infrastructure cost, the utility company has to design an incentive scheme to attract the traditional flat pricing (FP) users to be engaged in the TDP scheme. In this process, the utility company may share its revenue from the TDP scheme to those TDP users. It is found, with properly analyzing the energy procurement cost and user elasticity, a dynamic incentive strategy can be considered in dual-tariffs system when flat pricing (FP) and TDP pricing are co-existed. This dynamic incentive strategy gives appropriate stimulus to the users who are involved into the TDP program, and guarantee the utility company's profit at the same time.

**Toward Standards for Model-Based Control of Dynamic Interactions in Large Electric Power Grids**
Qixing Liu *Carnegie Mellon University*, Marija Ilic *Carnegie Mellon University*

This paper is motivated by the recent needs to manage possible instabilities between electrically-connected system components and/or sub-systems (layers) in future electric energy systems. It is shown that standard state-space models of general multi-layered energy systems have fundamentally the same structure which can be expressed in terms of: 1) state variables representing stand-alone layer (sub-system) dynamics; and, 2) an interaction variable between the layer and the rest of the system. Once this is recognized, three possible structure-based control designs are derived and analyzed for their performance using a small power system model. The three control designs considered are: 1) a decentralized component-level output controller; 2) decentralized sub-system (control area) layer output controller; and, 3) a full-state centralized system-level controller. Pros and cons of these three control architectures and their implications on three qualitatively different IT architectures and standards for dynamics in future electric energy systems are discussed.

## OS.21 -BioSPS.4 Biomedical Image Acquisition, Reconstruction, and Quantitation

Session Chairs: Richard Leahy, Justin Haldar, Krishna Nayak     Location: Trousdale Estates

**Magnetic Resonance Techniques for Fat Quantification in Obesity**
Houchun Hu *Children's Hospital Los Angeles*

As the prevalence of obesity and its comorbidities continue to rise in the United States and worldwide, robust imaging techniques and accurate post-processing strategies are critically needed to accurately quantify the distribution of fat in the human body. Magnetic resonance imaging and spectroscopy provide a wide array of sensitive methods to assess and characterize fat in storage locations such as white adipose tissue depots and "high-health-risk" ectopic sites such as organs and muscles. Quantitative fat measurements provide useful information to investigators in preventive medicine who monitor the efficacy of dietary, exercise, and surgical interventions to combat weight gain and obesity in longitudinal studies. They are also useful to clinicians who study the implications of steatosis and the pathophysiology of fat. The primary aim of this paper is to provide a technical review of state-of-the-art proton magnetic resonance methods in human body fat quantification. The paper will emphasize the fundamental principles with which several magnetic resonance techniques differentiate lean (water-dominant) and fatty (fat-dominant) tissues and illustrate with examples how each method can be appropriately used for fat quantification. The paper will also briefly summarize post-processing procedures that are currently in practice for extracting quantitative fat endpoints, such as adipose tissue depot volume and percent fat content in organs. Lastly, given its increased attention in recent literature, the paper will discuss progress in the imaging of human brown adipose tissue.

**Quantitatively Accurate Image Reconstruction for Clinical Whole-Body PET Imaging**
Evren Asma *GE Global Research*, Sangtae Ahn *GE Global Research*, Hua Qian *GE Global Research*, Girishankar Gopalakrishnan *GE Healthcare - Bangalore*, Kris Thielemans *King's College London*, Steven Ross *GE Healthcare*, Ravindra Manjeshwar *GE Global Research*, Alexander Ganin *GE Healthcare*

We present a PET image reconstruction approach that aims for accurate quantitation through model-based physical corrections and rigorous noise control with clinically acceptable image properties. We focus particularly on image generation chain components that are critical to quantitation such as physical system modeling, scatter correction, patient motion correction and regularized image reconstruction. Through realistic clinical datasets with inserted lesions, we demonstrate the quantitation improvements due to detector point spread function modeling, model-based single scatter estimation and the associated object-dependent multiple scatter estimation and non-rigid patient motion estimation and motion correction. We also describe a penalized-likelihood (PL) whole-body clinical PET image reconstruction approach using the relative difference penalty that achieves superior quantitation over the clinically-widespread ordered subsets expectation maximization (OSEM) algorithm while maintaining visual image properties similar to OSEM and therefore clinical acceptability. We discuss the axial and in-plane smoothing modulation profiles that are necessary to avoid large variations in noise and resolution levels. The overall approach of accurate models for data acquisition, corrections for patient related effects and rigorous noise control greatly improve quantitation and when combined with repeatable imaging protocols, limit quantitation variability only to factors related to patient physiology and scanner performance differences.

### Surface Fluid Registration and Multivariate Tensor-Based Morphometry in Newborns - The Effects of Prematurity on the Putamen

Jie Shi *Arizona State University*, Yalin Wang *Arizona State University*, Rafael Ceschin *Children's Hospital of Pittsburgh of UPMC*, Xing An *Arizona State University*, Marvin Nelson *Children's Hospital Los Angeles and University of Southern California*, Ashok Panigrahy *Children's Hospital of Pittsburgh of UPMC*, Natasha Lepore *Children's Hospital Los Angeles and University of Southern California*

Many disorders that affect the brain can cause shape changes in subcortical structures, and these may provide biomarkers for disease detection and progression. Automatic tools are needed to accurately identify and characterize these alterations. In recent work, we developed a surface multivariate tensor-based morphometry analysis (mTBM) to detect mor- phological group differences in subcortical structures, and we applied this method to study HIV/AIDS, William's syndrome, Alzheimer's disease and prematurity. Here we will focus more specifically on mTBM in neonates, which, in its current form, starts with manually segmented subcortical structures from MRI images of a two subject groups, places a conformal grid on each of their surfaces, registers them to a template through a constrained harmonic map and provides statistical comparisons between the two groups, at each vertex of the template grid. We improve this pipeline in two ways: first by replacing the constrained harmonic map with a new fluid registration algorithm that we recently developed. Secondly, by optimizing the pipeline to study the putamen in newborns. Our analysis is applied to the comparison of the putamen in premature and term born neonates. Recent whole-brain volumetric studies have detected differences in this structure in babies born preterm. Here we add to the literature on this topic by zooming in on this structure, and by generating the first surface-based maps of these changes. To do so, we use a dataset of manually segmented putamens from T1-weighted brain MR images from 17 preterm and 18 term-born neonates. Statistical comparisons between the two groups are performed via four methods: univariate and multivariate TBM, the commonly used medial axis distance, and a combination of the last two statistics. We detect widespread statistically significant differences in morphology between the two groups.

### High Spatio-Temporal Resolution Dynamic Contrast-Enhnaced MRI using Compressed Sensing

Kyunghyun Sung *University of California, Los Angeles*, Manoj Saranathan *Stanford University*, Bruce Daniel *Stanford University*, Brian Hargreaves *Stanford University*

Iterative thresholding methods have been extensively studied as faster alternatives to convex optimization methods for solving large-sized problems in compressed sensing MRI. A novel iterative thresholding method, called LCAMP (Location Constrained Approximate Message Passing), is presented for reducing computational complexity and improving reconstruction accuracy when a non-zero location (or sparse support) constraint can be obtained from view shared images in dynamic contrast-enhanced MRI (DCE-MRI). LCAMP modifies the existing approximate message passing algorithm by replacing the thresholding stage with a location constraint, which avoids adjusting regularization parameters or thresholding levels. This work is applied to breast DCE-MRI to demonstrate the excellent reconstruction accuracy and low computation time with highly undersampled data.

### Quantitative Analysis of Myocardial Perfusion Images

Piotr Slomka *Cedars-Sinai Medical Center*, Reza Arsanjani *Cedars-Sinai Medical Center*, Yuan Xu *Cedars-Sinai Medical Center*, Daniel Berman *Cedars-Sinai Medical Center*, Guido Germano *Cedars-Sinai Medical Center*

Myocardial perfusion imaging is a widely used test for the detection of coronary artery disease. Automated measurements of perfusion can be obtained from three-dimensional stress and rest images. The software segments the left ventricle of the heart and compares image intensities to normal subject database. In our research, we aim at reduction and ultimately elimination of human supervision in this process to improve overall reproducibility and accuracy for disease detection. We have developed several methods to this end such as automatic detection of potentially incorrect contours and direct measurement of stress-rest changes. Current state-of-the-art analysis methods demonstrate better reproducibility and similar accuracy when compared with experienced physicians. We aim to further improve the diagnostic accuracy by data mining techniques, combining several extracted image features with clinical information about the patients. Preliminary results show further improvements in accuracy, beyond that achieved by expert observers.

### Correcting Susceptibility-Induced Distortion in Diffusion-Weighted MRI using Constrained Nonrigid Registration

Chitresh Bhushan *University of Southern California*, Anand Joshi *University of Southern California*, Justin Haldar *University of Southern California*, Richard Leahy *University of Southern California*

Echo Planar Imaging (EPI) is the standard pulse sequence used in fast diffusion-weighted magnetic resonance imaging (MRI), but is sensitive to susceptibility-induced inhomogeneities in the main B0 magnetic field. In diffusion MRI of the human head, this leads to geometric distortion of the brain in reconstructed diffusion images and a resulting lack of correspondence with the high-resolution MRI scans that are used to define the subject anatomy. In this study, we propose and test an approach to estimate and correct this distortion using a non-linear registration framework based on mutual-information. We use an anatomical image as the registration-template and constrain the registration using spatial regularization and physics-based information about the characteristics of the distortion, without requiring any additional data collection. Results are shown for simulated and experimental data. The proposed method aligns diffusion images to the anatomical image with an error of 1-3 mm in most brain regions.

## OS.22-WCN.3 System Design, Architecture, Physical Layer Security in MIMO systems

Session Chair: Y.-W. Peter Hong Location: Franklin Hills

### Data Detection of Amplify-and-Forward User Cooperation in MIMO Broadcasting Systems without Channel State Information Feedback

Shih-Jung Lu *Academia Sinica*, Ronald Chang *Academia Sinica*, Wei-Ho Chung *Academia Sinica*

In this paper, we consider the broadcasting of data streams from a multiantenna source to several single-antenna or multiantenna users in a data broadcasting network. To improve the received data quality for all single-antenna users simultaneously while not compromising the data rates of multiantenna users, we propose a new cooperation scheme among single-antenna users. Users cooperate in the amplify-and-forward (AF) mode to jointly detect the spatially multiplexed data streams from the source with estimated channel state information. Simulation results verify the effectiveness of the proposed channel estimation scheme based on very few pilots in conjunction with the maximum likelihood (ML) detection scheme.

### On Secure Beamforming for Wiretap Channels with Partial Channel State Information at the Transmitter

Pin-Hsun Lin *National Taiwan University*, Shih-Chun Lin *NTUST*, Szu-Hsiang Lai *National Taiwan University*, Hsuan-Jung Su *National Taiwan University*

In this paper, we consider the secure transmission in ergodic fast fading multiple-input single-output single-antennaeavesdropper (MISOSE) wiretap channels with only the statistics of eavesdropper's channel state information at the transmitter (CSIT). Two kinds of the legitimate CSIT are assumed, that is, full and statistical legitimate CSIT. With full legitimate CSIT, we generalize and optimize the previously proposed artificial noise (AN) aided secure beamforming to improve its secrecy rate performance. The AN covariance matrix in our scheme is more flexible than previous scheme and the region of non-zero secrecy rate is enlarged significantly according to our simulations. For the case with statistical legitimate CSIT, we further prove that the secure beamforming is secrecy capacity achieving for the Rayleigh faded channels. In this case, the AN is not necessary. Extensions to cases where legitimate receiver and eavesdropper have multiple antennas will also be discussed.

### Secret Key Generation over Correlated Wireless Fading Channels using Vector Quantization

Hou-Tung Li *National Tsing Hua University*, Yao-Win Peter Hong *National Tsing Hua University*

Vector quantization schemes are proposed in this work to extract secret keys from correlated wireless fading channels. By assuming that the channel between two terminals are reciprocal, its estimates can be used as the common randomness for generating secret keys at the two terminals. Most schemes in the literature assume that channels are independent over time and utilize scalar quantization on each element of the estimated channel vector to generate secret key bits. These schemes are simple to implement but yield high key disagreement probability (KDP) at low SNR and low key entropy when channels are highly correlated. In this work, two vector quantization schemes, namely, the minimum key disagreement probability (MKDP) and the minimum quadratic distortion (MQD) secret key generation schemes, are proposed to effectively extract secret keys from correlated channel estimates. The vector quantizers are derived using KDP and QD as the respective distortion measures. To further reduce KDP, each channel vector is first pre-multiplied by an appropriately chosen unitary matrix to rotate the vector away from quantization cell boundaries. The MKDP scheme achieves the lowest KDP but requires high complexity whereas the MQD scheme yields lower complexity but at the cost of slightly increased KDP. Computer simulations are provided to demonstrate the effectiveness of the proposed vector quantization schemes.

### Two-stage Compensation for Non-ideal Effects in MIMO-OFDM Systems

Chih-Chi Wu *National Cheng Kung University*, Ping Ma *National Cheng Kung University*, Wun-De Hung *National Cheng Kung University*, Chih-Hung Kuo *National Cheng Kung University*

In this paper, we propose a two-stage estimation and compensation of the carrier frequency offset (CFO) and the IQ imbalance in MIMO-OFDM systems. In the first stage, the receiver IQ imbalance is compensated along with CFO estimation. In the second stage, the transmitter IQ imbalance is compensated by taking advantage of structure of Alamouti space time block codes. Robust least square estimations are applied for all compensation processes. Compared with the conventional system that only compensates receiver IQ imbalance, simulation results show that BER performance of the proposed system is significantly improved.

### On Blind Sequential Detection of Misbehaving Relay

Yang-Ming Yi *National Sun Yat-sen University*, Li-Chung Lo *National Sun Yat-sen University*, Wan-Jen Huang *National Sun Yat-sen University*

Consider a three-node cooperative system where the relay may misbehave for selfish or adversarial reasons. We propose a blind sequential detection to determine relay’s misbehavior with the least number of observations under requirement of detection performance. The likelihood function conditioning on the detected data symbols is derived here for three types of misbehaviors. The destination accumulates log-likelihood ratio (LLR) of current received symbols, and completes detection until the probabilities of false alarm and miss are both guaranteed below required thresholds. Simulation results show that the proposed scheme demands only small number of received symbols at SNR greater than 10dB.

## OS.23-IVM.9 Perception-based Multimedia Quality Assessment and Processing

Session Chairs: Xiaokang Yang, Weisi Lin, Zhou Wang, Jingliang Peng Location: Whitley Heights

### An Improved Full-Reference Image Quality Metric Based on Structure Compensation

Ke Gu *Shanghai Jiao Tong University*, Guangtao Zhai *Shanghai Jiao Tong University*, Xiaokang Yang *Shanghai Jiao Tong University*, Wenjun Zhang *Shanghai Jiao Tong University*

During the last two decades, image quality assessment has been a major research area, which considerably helps to promote the development of image processing. Following the tremendous success of Structural SIMilarity (SSIM) index in terms of the correlation between the quality predictions and the subjective scores, many improved algorithms have been further exploited, such as Multi-Scale SSIM (MS-SSIM) and Information content Weighted SSIM (IW-SSIM). However, a growing number of researchers have been devoted to the study of the effects of uneven responses to different image distortion categories on prediction accuracy of the quality metrics. Inspired by this, we propose an improved full-reference image quality assessment paradigm based on structure compensation. Experimental results on Laboratory for Image and Video Engineering (LIVE) database and Tampere Image Database 2008 (TID2008) are provided to confirm our introduced approach has superior prediction performance as compared to mainstream image quality metrics. Besides, it is worth emphasizing that our algorithm not introduces other operators but only applies the SSIM function to compensate itself, and furthermore, it also has an effective capability of image distortion classification.

### Video Quality Metric for Consistent Visual Quality Control in Video Coding

Long Xu *University of Science and Technology Beijing*, King Ngi Ngan *Chinese University of Hong Kong*, Song Nan Li *Chinese University of Hong Kong*, Lin Ma *Chinese University of Hong Kong*

The visual quality consistency is one of the most important issues in video quality assessment (VQA). When people view a sequential video, they may have an unpleasant perceptual experience if the visual quality of video frames is inconsistent even though the average visual quality of the video is not too bad. Thus, the consistent visual quality control is mostly expected in real-time video communication. Additionally, in conventional video communication, the channel bandwidth and buffer resources are limited. The unfair distribution of encoding resources among video frames would result in not only inconsistent visual quality but also other types of spatial distortions. In this paper, a new objective visual quality metric (VQM) is firstly proposed for measuring the video quality in video coding. It makes full use of the information of video coding without extra computational complexity. Secondly, a visual quality control algorithm is proposed to ensure the consistent visual quality of video coding under the given channel and buffer resources. Finally, the experimental results indicate that the proposed VQM is consistent well with the human visual system (HVS). In addition, the consistent visual quality, better rate-distortion efficiency, accurate bit control and compliant buffer can be achieved by the proposed visual quality control algorithm.

### A New No-reference Image Quality Assessment Model Based on DCT Coefficients Distribution and PSNR

Zhengyou Wang *Shijiazhuang Tiedao University*, Wan Wang *Shijiazhuang Tiedao University*, Zhenxing Li *Jiangxi University of Finance & Economics*, Jin Wang *Shijiazhuang Tiedao University*, Weisi Lin *Nanyang Technological University*

Based on the traditional quality metric PSNR, we propose a new no-reference image quality assessment model nPSNR (No-reference PSNR) for JPEG compressed images. The metric performs in DCT domain and the DCT coefficient distribution is used. This metric estimates the MSQE (mean-squared quantization error) of a decoded image with the distributions of AC coefficients and DC coefficients of the encoded image. Based on the MSQE, the overall nPSNR value of the image is calculated. We test the proposed metric on the selected images from the TID2008 database. Then, the computational scores (nPSNR) are compared with the ground truth values (MOS) based on three performance criteria. Experimental results demonstrate that the proposed metric is more consistent with the subjective perception than the state-of-the-art full-reference image quality assessment metrics.

### A Fusion Approach to Video Quality Assessment Based on Temporal Decomposition

Tsung-Jung Liu *University of Southern California*, Weisi Lin *Nanyang Technological University*, C.-C. Jay Kuo *University of Southern California*

In this work, we decompose an input video clip into multiple smaller intervals, measure the quality of each interval separately, and apply a fusion approach to integrating these scores into a final one. To give more details, an input video clip is first decomposed into smaller units along the temporal domain, called the temporal decomposition units (TDUs). Next, for each TDU that consists of a small number of frames, we adopt a proper video quality metric (specifically, the MOVIE index in this work) to compute the quality scores of all frames and, based on the sociological findings, choose the worst scores of TDUs for data fusion. Finally, a regression approach is used to fuse selected worst scores from all TDUs to get the ultimate quality score of the input video as a whole. We conduct extensive experiments on the LIVE video database, and show that the proposed approach indeed improves MOVIE and is also competitive with other state-ofÐthe-art video quality metrics.

### Performance Comparison of Decision Fusion Strategies in BMMF-based Image Quality Assessment

Lina Jin *Tampere University of Technology*, SeongHo Cho *University of Southern California*, Tsung-Jung Liu *University of Southern California*, Karen Egiazarian *Tampere University of Technology*, C.-C. Jay Kuo *University of Southern California*

The block-based multi-metric fusion (BMMF) is one of the state-of-the-art perceptual image quality assessment (IQA) schemes. With this scheme, image quality is analyzed in a block-by-block fashion according to the block content type (i.e. smooth, edge and texture blocks) and the distortion type. Then, a suitable IQA metric is adopted to evaluate the quality of each block. Various fusion strategies to combine the QA scores of all blocks are discussed in this work. Specifically, factors such as quality scores distribution and the spatial distribution of each block are examined using statistics methods. Finally, we compare the performance of various fusion strategies based on the popular TID database.

### Quality-of-Experience Perception for Video Streaming Services: Preliminary Subjective and Objective Results

Khalil ur Rehman Laghari *EMT-INRS*, Omneya Issa *Communications Research Centre Canada*, Filippo Speranza *Communications Research Centre Canada*, Tiago Falk *Institut National de la Recherche Scientifique*

Quality-of-Experience (QoE) is a human centric notion that produces the blue print of human perception, feelings, needs and intentions while Quality-of-Service (QoS) is a technology centric metric used to assess the performance of a multimedia application and/or network. To ensure superior video QoE, it is important to understand the relationship between QoE and QoS. To achieve this goal, we conducted a pilot subjective user study simulating a video streaming service over a broadband network with varying distortion scenarios, namely packet losses (0, 0.5, 1, 3,7, and 15%), packet reorder (0, 1, 5, 10, 20, and 30%), and coding bit rates (100, 400, 600, and 800 Kbps). Users were asked to rate their experience using a subjective quantitative metric (termed Perceived Video Quality, PVQ) and qualitative indicators of "experience." Simulation results suggest a) an exponential relationship between PVQ and packet loss and between PVQ and packet reorder, and b) a logarithmic relationship between PVQ and video bit rate. Similar trends were observed with the qualitative indicators. Exploratory analysis with two objective video quality metrics suggests that trends similar to those obtained with the subjective ratings were obtained, particularly with a full-reference metric.

## OS.24-IVM.10 Visual Media Data Representation, Retrieval and Recognition

Session Chairs: Jingliang Peng, Xin-Shun Xu Location: Mt. Olympus

### A User-driven Model for Content-based Image Retrieval

Yi Zhang *Tianjin University*, Zhipeng Mo *Tianjin University*, Wenbo Li *Tianjin University*, Tianhao Zhao *Tianjin University*

The intention of image retrieval systems is to provide retrieved results as close to users' expectations as possible. However, users' requirements vary from each other in various application scenarios for the same concept and keywords. In this paper, we introduce a personalized image retrieval model driven by users' operational history. In our simulated system, three types of data, which are browsing time, downloads and grades, are collected to generate a sort criterion for retrieved image sets. According to the criterion, the image collection is classified into a positive group, a negative group and a testing group. Then a SVM classifier is trained with image features extracted from three groups and used to refine retrieved results. We test the proposed method on several image sets. The experimental results show that our model is effective to represent users' demands and help improving retrieval accuracy.

### A Poselet Based Key Frame Searching Approach in Sports Training Videos

Wu Lifang *Beijing University of Technology*, Zhang Jingwen *Beijing University of Technology*, Yan Fenghui *Beijing University of Technology*

In some sport training application, it is necessary to search the key frames of training video for carefully analysis. In this paper, we take the key frame searching issue as a pose estimation problem. First, a set of various pose detectors are collected trough the twice SVM training process, each of which can be interpreted as a learned pose-specific HOG weight classifier. Then we run each linear SVM classifier over the image in a multi_scale scanning mode. In order to resolve the problem of extreme similarity between the adjacent frames, the detection hits at every scale in each frame is counted as the principle of optimal key frame selection. The frame with the most detection hits are chosen as the key frame for the pose detector. The experimental results using weight-lifting training videos show the efficiency of proposed approach

### A Novel Multi-instance Learning Algorithm with Application to Image Classification

Xiaocong Xi *Shandong University*, Xinshun Xu *Shandong University*, Xiaolin Wang *Shandong University*

Image classification is an important research topic due to its potential impact on both image processing and understanding. However, due to the inherent ambiguity of image-keyword mapping, this task becomes a challenge. From the perspective of machine learning, image classification task fits the multi-instance learning (MIL) framework very well owing to the fact that a specific keyword is often relevant to an object in an image rather than the entire image. In this paper, we propose a novel MIL algorithm to address image classification task. First, a new instance prototype extraction method is proposed to construct projection space for each keyword. Then, each training sample is mapped to this potential projection space as a point, which converts the MIL problem into standard supervised learning problem. Finally, an SVM is trained for each keyword. The experimental results on a benchmark data set Corel5k demonstrate that the new instance prototype extraction method can result in more reliable instance prototypes and faster running time, and the proposed MIL approach outperforms some state-of-the-art MIL algorithms.

### Hierarchical Bag-of-Words Model for Joint Multi-View Object Representation and Classification

Xiang Fu *University of Southern California*, Sanjay Purushotham *University of Southern California*, Daru Xu *University of Southern California*, Jian Li *University of Southern California*, C.-C. Jay Kuo *University of Southern California*

Multi-view object classification is a challenging problem in image retrieval. One common approach is to apply the visual bag-of-words (BoW) model to all view representations of each object class and compare them with the representation of the query image one by one so as to determine the closest view of the object class. This approach offers good matching performance, yet it demands a large amount of computation and storage space. To address these issues, we propose a novel hierarchical BoW model that provides a concise representation of each object class with multi-views. When the higher level BoW representation does not match with that of the query instance, further comparison can be saved. We can also incorporate similar views to reduce the storage space. We conduct experiments on a dataset of 3D object classes, and show that the proposed approach achieves higher efficiency in terms of lower computational complexity and storage space while preserving good matching performance.

### An Analysis of Eating Activities for Automatic Food Type Recognition

Hyun-Jun Kim *Samsung Electronics*, Mira Kim *Samsung Electronics*, Sun-Jae Lee *Samsung Electronics*, Young Sang Choi *Samsung Electronics*

Nowadays, chronic diseases such as type 2 diabetes or cardiovascular diseases are considered to be one of the most serious threats to healthy life. These kinds of diseases are primarily caused by an unhealthy lifestyle including lack of exercise, irregular meal patterns and abuse of addictive substances such as alcohol, caffeine and nicotine. Therefore, observing our daily lives is crucial in developing interventions to reduce the risk of lifestyle diseases. In order to manage and predict progression of diseases of a patient, objective measurement of lifestyle is essential. However, self-reporting questionnaires and interviews have limitation due to human errors and difficulty of conducting. In this paper, we analyzed users' eating activities and comprising sub-actions for developing eating activity recognition system based on a tri-axial accelerometer embedded wrist band. By analyzing actions in eating activities, we can improve the accuracy of the recognition of eating activities and also provide clues that indentifying the type of foods.

### A New Hybrid PCNN for Multi-Objects Image Segmentation

Zhenbo Li *China Agricultural University*

Many image based applications such as multi-object tracking were nagged by the problem of robust multi-objects image segmentation. In this paper, we propose a new hybrid Pulse Coupled Neural Network (PCNN) method for multi-object segmentation. Firstly, we use saliency detection methods, Graph-based visual saliency (GBVS) and Spectrum Residual (SR) to find more accurate object region (R1) and more number of object regions (R2) separately. Then an improved PCNN is used to work out the multi-objects with R1 and R2. The statistical result of R1 is selected as an adaptive generator threshold of PCNN and a selection standard of segmentation result. R2 determines the correct object number in the image. Experiments of images selected from BSD and VOC and two full image datasets (MSRC v2 and Weizmann) prove that our method can get more right object quantity and more accurate object region than GBVS-PCNN[1] and adaptive PCNN[2].

## Wednesday, December 5, 2012

## 10:50 - 12:30

## OS.25-IVM.11 Recent Topics in Image, Video, and Multimedia Processing (I)

Session Chair: Yi-Chong Zeng Location: Doheny

### 3D Shape Retrieval from a 2D Image as Query

Masaki Aono *Toyohashi University of Technology*, Hiroki Iwabuchi *Toyohashi University of Technology*

3D shape retrieval has gained popularity in recent years. Yet we have difficulty preparing a 3D shape by ourselves for query input. It is therefore much desired to have an easy way of doing 3D shape search in terms of query input. In this paper, we propose a new method for defining a feature vector for 3D shape retrieval from a single 3D photo image. Our feature vector is defined as a combination of Zernike moments and HOG (Histogram of Oriented Gradients), where these features can be extracted from both a 2D image and a 3D shape model. Comparative experiments demonstrate that our approach shows very promising and effective as an initial clue to searching for more relevant 3D shape model we have in mind.

### A Large-Scale Shape Benchmark for 3D Object Retrieval: Toyohashi Shape Benchmark

Atsushi Tatsuma *Toyohashi University of Technology*, Hitoshi Koyanagi *Toyohashi University of Technology*, Masaki Aono *Toyohashi University of Technology*

In this paper, we describe the Toyohashi Shape Benchmark (TSB), a publicly available new database of polygonal models collected form the World Wide Web, consisting of 10,000 models, as the largest 3D shape models to our knowledge used for benchmark testing. TSB includes 352 categories with labels. It can be used for both 3D shape retrieval and 3D shape classification. Formerly, the most well-known 3D shape benchmark has been the PSB, or the Princeton Shape Benchmark, consisting of 1,814 models, including the half as training data and the remaining half as testing. TSB is approximately 6 times larger than PSB. Unlike textual data such as TREC and NTCIR data collections, 3D shape repositories have been suffering from the shortage of data, and from the difficulty in testing the scalability of any algorithms that work on top of given benchmark data set. In addition to the TSB, we propose a new shape descriptor which we call DB-VLAT (Depth-Buffered Vector of Locally Aggregated Tensors). During the comparison with the TSB, we will demonstrate that our new shape descriptor exhibits the best search performance among those known programs to which we have had access on the Internet, including Spherical Harmonic

Descriptor and Light-Field Descriptor. We believe that the TSB can be a step toward the next generation 3D shape benchmark having massive 3D data collection, and hope it to be served for many purposes in both academia and industries.

### Quality Assessment of Finger-vein Image

Huafeng Qin *Chongqing University*, Sheng Li *Nanyang Technological University*, Alex Chichung Kot *Nanyang Technological University*, Lan Qin *Chongqing University*

In this paper, we propose a novel quality assessment of finger-vein images for quality control purpose. First of all, we divide a finger vein image into a set of non-overlapping blocks. In order to detect the local vein patterns, each block is projected into the Radon space using an average Radon transform. A local quality score is estimated for each block according to the curvature in the corresponding Radon space, based on which a global quality score of the finger-vein is computed and assessed. Experimental results show that our approach can effectively identify the low quality finger-vein images, which is also helpful in improving the performance of a finger-vein recognition system.

### Preserving Features in Multilevel Halftones

Lai-Yan Wong *Hong Kong Polytechnic University*, Yuk-Hee Chan *Hong Kong Polytechnic University*

Conventional threshold decomposition (TD) based multilevel halftoning algorithms decompose an input image into layers, halftone them sequentially with a binary halftoning algorithm, and combine their binary halftones to produce the final multilevel halftone. When these algorithms are exploited to produce multilevel halftones, bright spatial features are generally difficult to preserve as darker pixels in the final multilevel output are positioned first. We propose a solution to solve this problem in this paper. Simulation result shows that the proposed method can provide an output of better quality as compared with conventional TD-based algorithms.

### Automatic Recognition of Frame Quality Degradation For Inspection of Surveillance Camera

Yi-Chong Zeng *Institute for Information Industry*, Miao-Fen Chueh *Institute for Information Industry*, Chi-Hung Tsai *Institute for Information Industry*

When surveillance camera is broken down, it will degrade frame quality directly. Sometimes, quality degradation happens occasionally, it is difficult for people being aware it immediately. With the aim to automatically inspect surveillance camera, we propose an automatic method to recognize frame quality degradation. Seven features are extracted based on four kinds of measures, i.e. mean of structure similarity, variation of intensity difference, minimum of block correlation, and average color. Those measures have different reactions to different de-gradations. Subsequently, linear discriminant analysis (LDA) applied to the extracted features is able to train classifiers. Six classes of degradations are recognized in this work, including signal missing, color missing, local alternation, global alteration, periodic intensity change, and normal status. After implement-ing degradation recognition, we determine whether surveillance camera works normally or not. The experiment results demon-strate that the proposed method is capable of recognizing de-gradation as well as inspecting surveillance camera.

## OS.26-SLA.10 Immersive Audio and Cloud-assisted Audio Processing

Session Chairs: Woon-Seng Gan, Yu RongShan, Ee-Leng Tan Location: Beachwood

### Theories and Signal Processing Techniques for the Implementation of Sound Ball in Space Using Loudspeaker Array

Jung-Woo Choi *KAIST*, Yang-Hann Kim *KAIST*

It has been known that we can make a certain region of more acoustic energy than others. This can be achieved by utilizing loudspeaker arrays and designing a multichannel filter that effectively controls interferences of sound waves in space and time. The concept of a bright and dark zone[Choi and Kim, Generation of an acoustically bright zone with an illuminated region using multiple sources,J. Acoust. Soc. Am., Vol. 111(4), 1695-1700, Apr. 2002] showed that this idea can be realized in practice. Then we attempted to make a “sound ball” that utilizes the concept of bright and dark zone for generating a small spatial region of concentrated sound energy inside. The 32 speaker system which surrounds the zone of interest tried to implement the ball that can be positioned and also allowed to be moved. However, it was also found that the solution based on the bright and dark zones control does not, in strict sense, guarantee an effective radiation of sound from the ball. Understanding this inherent limitation motivated us to design a novel mean to have a sound ball that can radiate effectively. This has to solve the well-known Kirchhoff-Helmholtz equation for the case of which the source or sources are surrounded by an array of speakers. The sound ball is implemented by using a 50-channel spherical loudspeaker array, in which the loudspeakers are positioned on the Lebedev quadrature grid.

### Cloud-based Audio Fingerprinting Service

Wenyu Jiang *Institute for Infocomm Research*, Yongwei Zhu *Institute for Infocomm Research*, Xiaoming Bao *Institute for Infocomm Research*, Rongshan Yu *Institute for Infocomm Research*

Audio Fingerprinting allows the identification of a query audio clip by matching the query audio fingerprints against a reference database. Traditionally, the matching process, which is CPU and memory intensive, is implemented either on a single computer (which is confined by CPU and memory limits for large databases), or on a computer cluster in a proprietary manner (which has limited flexibility in scaling the database). We have implemented audio fingerprinting prototype software that can run in a Cloud environment, specifically Hadoop/MapReduce. Because the MapReduce framework is designed for stream data processing instead of database query, we discuss how we address this challenge as well as other challenges such as appropriate data input format and partitioning. A performance evaluation of the software on a real dataset of ~8500 songs and real Hadoop clusters is presented to illustrate its efficacy, where a batch query of 1000 60sec clips can be completed in ~50sec in addition to ~30sec of database loading time with a 12-node cluster configuration.

### Repeating Segment Detection in Songs using Audio Fingerprint Matching

Regunathan Radhakrishnan *Dolby Laboratories Inc.*, Wenyu Jiang *Institute for Infocomm Research*

We propose an efficient repeating segment detection approach that doesn't require computation of the distance matrix for the whole song. The proposed framework first extracts audio fingerprints for the whole song. Then,for each time step in the song we perform a query to match a sequence of M fingerprint codewords against the fingerprints of the rest of the song. In order to find a match for the first fingerprint query, a search tree data structure is built with the fingerprints of the rest of the song. For subsequent fingerprint queries for the rest of the song, the matching process dynamically updates the search tree data structure to exclude the M fingerprint codewords corresponding to each time step. For each matching segment, we record the time offset from the query segment. Following the matching process for the whole song, we compute the histogram of the number of matching segments for each offset. The peaks in this histogram correspond to offsets at which matches were found more often than others and can be used to pick out a set of repeating segments

### Streaming of Scalable Multimedia over Content Delivery Cloud

Xiaoming Bao *Institute for Infocomm Research*, Rongshan Yu *Institute for Infocomm Research*

Content Delivery Cloud (CDC) extends Content Delivery Network (CDN) to provide elastic, scalable and low cost services to the customers. For multimedia streaming over CDC, caching the media content onto the edge server from storage cloud is commonly used to minimize the latency of content delivery. It is very important for CDN to balance between the resources being used (storage space, bandwidth, etc) and the performance achieved. Commercial CDNs (such as Akamai, Limelight, Amazon CloudFront) have their proprietary caching algorithms to deal with this issue. In this paper, we propose a method to further improve the efficiency of the caching system for scalable multimedia contents. Specifically, we notice that a scalable multimedia content can be flexibly truncated to lower bit rates on-the-fly based on the available network bandwidth between the edge server to the end users. Therefore, it may not be necessary to cache such a content at its highest quality/rate. Based on this observation, we show that edge server can decide an optimized truncation ratio for the cached scalable multimedia contents to balance between the quality of the media and the resource usage. The proposed optimized truncation algorithm is analyzed and its efficacy in improving the efficiency of the caching system is justified with simulation result.

### Spatial Sound Reproduction Using Conventional and Parametric Loudspeakers

Ee-Leng Joseph Tan *Nanyang Technological University*, Woon-seng Gan *NTU*, Chiu-Hao Chen *Nanyang Technological University*

The auditory image of a movie or game scene can be decomposed into point-like sources and diffused sources for effective and accurate audio synthesis. By embedding appropriate visual and audio cues into objects in a 2D or 3D visual scene, an immersive and engaging experience can be created. While there are many breakthroughs in the display technology recently, such as the ultra high-definition (UHD) and 3D displays, conventional sound systems (stereo, 5.1, etc) are still being used. Such an audio-visual setup may degrade the overall experience. This degradation is directly linked to the dispersive nature of the conventional loudspeaker, and the rendered auditory image may be perceived to lack sharpness in the spatial imaging due to the reverberant nature of a room. This drawback tends to lead to comparably poor synthesis of point-like sources as compared to diffused sources in the rendered auditory image. On the other hand, the rendered auditory image from a directional loudspeaker, such as the parametric loudspeaker, may seem to lack spaciousness and sound envelopment due to very little influence of the acoustics of a room. Therefore, directional loudspeaker is suitable for rendering point-like sources, but not diffused sources. In this paper, we propose a unique sound system which comprises of conventional loudspeakers and parametric loudspeakers. This setup exploits the high directivity of the parametric loudspeakers to render sharp auditory images while producing the diffused sources of the auditory image using the conventional loudspeaker.

### Interactive 3D Audio Rendering in Flexible Playback Configurations

Jean-Marc Jot *DTS, Inc.*

Interactive object-based 3D audio spatialization technology has become commonplace in personal computers and game consoles. While its primary current application is 3-D game sound track rendering, it is ultimately necessary in the implementation of any personal or shared immersive virtual world (including multi-user communication and telepresence). The successful development and deployment of such applications in new mobile or online platforms involves maximizing the plausibility of the synthetic 3D audio scene while minimizing the computational and memory footprint of the audio rendering engine. It also requires a flexible, standardized scene description model to facilitate the development of applications targeting multiple platforms. This paper presents a general computationally efficient 3D positional audio and environmental reverberation rendering engine applicable to a wide range of loudspeaker or headphone playback configurations.

## OS.27-SPS.2 Advances in Circuits and Systems for Multimedia Processing and Analysis (I)

Session Chairs: Takeshi Ikenaga, Tse-Wei Chen Location: Runyon

### A Low Power ASIP for Precision Configurable FFT Processing

Yifan Bo *Fudan University*, Jun Han *Fudan University*, Yao Zou *Fudan University*, Xiaoyang Zeng *Fudan University*

Fast Fourier transformation (FFT) is a key operation in digital communication systems. Different communication standards require various FFT length and precision. In this paper, we present a low power Application-Specific Instruction-set Processor (ASIP) for variable length (16-point - 4096-point) and bit precision (8-bit - 16-bit) to meet different requirements. We use scalable multipliers to construct the butterfly unit, which support both 8-bit and 16-bit operation. The order of butterfly operation is adjusted to reduce twiddle-factor ROM accesses, so as to reduce overall power consumption efficiently. Clock Gating is implemented to shut down processor's pipeline during the FFT process in terms of special low power demands. Special Instructions are tailored to make full use of the flexible hardware.

### SIFT-Based Low Complexity Keypoint Extraction and Its Real-Time Hardware Implementation for Full-HD Video

Takahiro Suzuki *Waseda University*, Takeshi Ikenaga *Waseda University*

Scale-Invariant Feature Transform (SIFT) has lately attracted attention in computer vision as a robust keypoint detection algorithm which is invariant for scale, rotation and illumination change. However, its computational complexity is too high to apply practical real-time applications. This paper proposes a low complexity keypoint extraction algorithm based on SIFT descriptor and utilization of the database, and its real-time hardware implementation for Full-HD resolution video. The proposed algorithm computes SIFT descriptor on the keypoint obtained by corner detection and selects a scale from the database. It is possible to parallelize the keypoint detection and descriptor computation modules in the hardware. These modules do not depend on each other in the proposed algorithm in contrast with SIFT that computes a scale. The processing time of descriptor computation in this hardware is independent of the number of keypoints because its descriptor generation is pipelining structure of pixel. Evaluation results show that the proposed algorithm on software is 12 times faster than SIFT. Moreover, the proposed hardware on FPGA is 427 times faster than SIFT and 61 times faster than the proposed algorithm on software. The proposed hardware performs keypoint extraction and matching at 60 fps for Full-HD video.

### Halo Artifacts Reduction Method for Variational based Realtime Retinex Image Enhancement

Hiroshi Tsutsui *Kyoto University*, Satoshi Yoshikawa *Osaka University*, Hiroyuki Okuhata *Synthesis Corporation*, Takao Onoye *Osaka University*

In this paper, we propose a novel halo reduction method for variational based Retinex image enhancement. In variational based Retinex image enhancement, a cost function is designed based on the illumination characteristics. The enhanced image is obtained by extracting the illumination component, which gives minimum cost, from the given input image. Although this approach gives good enhancement quality with less computational cost, a problem that dark regions near edges remain dark after image enhancement, known as halo artifact, still exists. In order to suppress such artifacts effectively, the proposed method adaptively adjusts the parameter of the cost function, which influences the trade-off relation between reducing halo artifacts and preserving image contrast. The proposed method is applicable to an existing realtime Retinex image enhancement hardware implementation.

### Design and Analysis of a Many-Core Processor Architecture for Multimedia Applications

Jyu-Yuan Lai *National Tsing Hua University*, Po-Yu Chen *National Tsing Hua University*, Ting-Shuo Hsu *National Tsing Hua University*, Chih-Tsun Huang *National Tsing Hua University*, Jing-Jia Liou *National Tsing Hua University*

We present a design of many-core processor architecture with superior cost-effectiveness to fulfill the rapid increasing demand of high-speed embedded multimedia applications. The prototype platform consists of sixteen processor cores and a 4-by-4 mesh-based duplex network interconnection with external memory. The hardware and software interface in a bare-metal environment, i.e., without an Operating System (OS), has been emphasized in our architecture. An on-chip communication library is developed for practical parallel applications. In addition, we propose two memory-based file handling approaches to manipulate files with the lack of file-system support by OS. Our file handling approach can effectively reduce the minimum requirement of local memory without page swapping for each core from 4 MB to 64 KB in a case study of JPEG encoding. Furthermore, the analysis of instruction and data caches is addressed for the trade-off between area and speed. The experimental result indicates that our many-core platform with its application development infrastructure is efficient in delivering cost-effective multimedia applications in a bare-metal environment.

### An Edge-based Adaptive Image Interpolation and Its VLSI Architecture

Hongbin Sun *Xi'an Jiaotong University*, Fengwei Zhang *Xi'an Jiaotong University*, Nanning Zheng *Xi'an Jiaotong University*

The design of high quality yet real-time image interpolation has become increasingly important for digital TV SoC, as the size of flat-panel display has been steadily increased to 4K*2K definition in the very near future. This paper aims to develop a real-time image interpolation algorithm that can achieve the high-ratio image scaling with sharp and natural edges. Comparing with conventional image interpolation approaches that often suffer from either image blurring/jagged problem or high computational cost, this paper propose a high-efficient edge directional image interpolation approach that can support multiple interpolation directions and hence can well preserve the detail and edges. Experimental result shows that the proposed image interpolation algorithm is able to achieve the high quality at the high image scaling ratio while only incurring very low computational cost. And the VLSI architecture of proposed image interpolation is also presented.

## OS.28-SLA.11 Recent Advances in Audio and Acoustic Signal Processing (II)

Session Chairs: Yoshinobu Kajikawa, Woon-Seng Gan Location: Laurel

### Psychoacoustic Active Noise Control System

Tongwei Wang *NTU*, Woon-seng Gan *NTU*, Yong-Kim Chong *NTU*

In practical active noise control (ANC) applications, it is difficult to completely cancel out the undesired noise. In order to improve the user's comfort level in a noisy environment, a psychoacoustic ANC system that incorporates masking techniques is proposed in this paper. A two-stage approach of performing ANC, followed by masking the residual noise with carefully selected masking signal, is used to enhance the user's listening experience. In order to mask the residual noise effectively, an automatic gain controller (AGC) is used to give different gains to the masking signal according to the residual noise level. A new mechanism to control the gain of the AGC is proposed so that the system can produce a conclusive listening experience for the user. The proposed hybrid ANC-masking system also takes into account of any uncorrelated noise that is only captured by the error microphone. Computer simulations are conducted to show the superior performance of the proposed psychoacoustic ANC system. Two real-signal cases are also considered in this paper to test out the effectiveness in combining ANC with masking.

### Network-Based Multi-Channel Signal Processing Using the Precision Time Protocol

Yoshifumi Chisaki *Kumamoto University*, Dan Murakami *Kumamoto University*, Tsuyoshi Usagawa *Kumamoto University*

A conventional microphone array system uses a conductor to wire from a microphone to an input via an amplifier. While, a wireless transmission for an array system makes the configuration flexible, and it is expected to provide novel applications widely, such as measuring of impulse response in wide area. Not only in acoustic research field but also in research area of remote sensing of body motion and so on, data acquisition with a precise time synchronization is essential. When a signal received at distributed microphone's position is sent over a computer network, the data is packetized and can include some information at receiving point. When a time is included as information at data acquisition position, the time depends on each data acquisition client device. Since it is possible to use network time protocol or precision time protocol, multi-channel signal processing can be achieved with ease. This paper proposes multiple signals transmission system over a computer network with a time code embedding to synchronize those signals. The signal from a client is reconstructed at a server with the time code. Since the time differences between clients affects to performance of the multichannel signal processing, smaller error in time at a client is preferred. This paper discusses how the error in time between channels affects to performance of the distributed microphone array system.

### Content/Context-Adaptive Feature Selection for Environmental Sound Recognition

EnShuo Tsau *University of Southern California*, Sachin Chachada *University of Southern California*, C.-C. Jay Kuo *University of Southern California*

Environmental sound recognition (ESR) is a challenging problem that has gained a lot of attention in the recent years. A large number of audio features has been adopted for solving the ESR problem. In this work, we focus on the problem of automatic feature selection. Specifically, we propose two methods, called the content-adaptive and the context-adaptive feature selection schemes to achieve this goal. Finally, the superior performance of the proposed feature selection methods is demonstrated when they are applied to a medium-sized environmental database with a simple Bayesian network classifier.

### Blind Depth Estimation Based on Primary-To-Ambient Energy Ratio for 3-D Acoustic Depth Rendering

Se-Woon Jeon *Yonsei University*, Dae Hee Youn *Yonsei University*, Young-cheol Park *Yonsei University*

Since the advent of 3-D video, the acoustic depth rendering for the proximity effect has been an issue of great interest. In this study, we propose an algorithm for estimating acoustic depth cues from stereo audio signal, without a priori knowledge about the source-to-listener geometry and room environments. We employ the principal component analysis (PCA) to estimate the acoustic depth based on the primary-to-ambient energy ratio (PAR) which is related with the front-back movement of the sound source. And for the acoustic depth rendering, the distance variation of the sound source is parameterized through tracking the estimated depth cue. The proposed estimation algorithm was evaluated using stereo audio clips extracted from a real 3-D movie, and the results confirmed effectiveness of the proposed acoustic depth estimation algorithm.

### Theoretical Framework for Stochastic Modeling of FxLMS-based Active Noise Control Dynamics

Iman Tabatabaei Ardekani *University of Auckland*, Waleed Abdulla *University of Auckland*

There have been several contributions on theoretical modeling of FxLMS-based active noise control systems; however, when it is intended to derive elegant closed-form expressions for formulating dynamical behaviors of these systems, a number of simplifying assumptions regarding the acoustic noise, the actual secondary path and its model have to be used. This paper develops a dynamic model for FxLMS-based ANC systems, considering a general stochastic acoustic noise and a general secondary path. Also, an arbitrary secondary path model, which is not necessarily a perfect model, is considered. The main distinction of this model is that previously-derived dynamic models can be resulted in from it as special cases.

### Design of a Pitch Quantization and Pitch Correction System for Real-Time Music Effects Signal Processing

Corey Cheng *MIT*

This paper describes the design of a practical, real-time pitch quantization system intended for digital musical effects signal processing. Like most modern pitch quantizers, this system can be used to pitch correct and even reharmonize out-of-tune singing to alternative musical scales simultaneously (e.g. major, minor, diminished, etc.) Pitch Quantization can also be intentionally exaggerated to produce distinctive effects processing which results in an emotionally inflected and/or “robotic” sound. This system uses intentionally simple signal processing algorithms which make real-time processing possible on constrained devices. In particular, we employ tools such as an octave resolver and range limiter, grain boundary expansion and contraction, and transient detection to enhance the performance of our system.

## OS.29-SIPTM.2 Recent Topics on Signal Processing in Noisy Environments

Session Chairs: Kiyoshi Nishikawa, Arata Kawamura Location: Trousdale Estates

### Low-complexity Approximate LMMSE Channel Estimation for OFDM Systems

Shuichi Ohno *Hiroshima University*, Emmanuel Manasseh *Hiroshima University*

A low-complexity linear minimum mean square error (LMMSE) based channel estimator is proposed for orthogonal frequency-division multiplexing (OFDM) systems over frequency-selective channels. Using the law of a large number, we approximate the LMMSE estimator to reduce the numerical complexity of the channel estimation. Our estimator exhibits comparable performance with the LMMSE estimator at low SNR but suffers the performance floor due to the approximation, which are verified by numerical simulations.

### Mixture Structure of Kernel Adaptive Filters for Improving the Convergence Characteristics

Kiyoshi Nishikawa *Tokyo Metropolitan University*, Hiroya Nakazato *Tokyo Metropolitan University*

In this paper, we propose a mixture structure of the linear and kernel adaptive fiilters for improving the convergence characteristics of the kernel normalized least mean square (KNLMS) adaptive algorithm. The proposed method is based on the concept of the affine constrained mixture structure for the linear normalized LMS adaptive filters which uses the more than two adaptive filters concurrently. We derive the proposed structure, and its implementation method. We confirm the effectiveness of the proposed method through the computer simulations.

### Pink Noise Whitening Method for Pitch Synchronous LPC Analysis

Liu Liqing *Saitama University*, Shimamura Tetsuya *Saitama University*

We present a new noise whitening method for pitch synchronous LPC analysis under pink noise circumstances. First,we utilize a rectangular window to extract two frames whose shifting interval is a full pitch period. Then we perform a subtraction operation between the two frames to obtain a new noise signal which is considered to be not corrupted by the voiced speech signal. The obtained new noise signal can be used to design a new prediction whitening filter. The new whitening filter not only whitens the pink noise signal, but also can keep the vocal tract and formant natures of voiced speech signal. Utilizing the whitened signal, we can improve the pitch synchronous addition and subtraction (PSAS) method under pink noise circumstances. We discuss the properties of the whitened signal and PSAS method. Experimental results indicate the effectiveness of the proposed method.

### Self-Interference Canceller for Full-Duplex Radio Relay Station Using Virtual Coupling Wave Paths

Kazunori Hayashi *Kyoto University*, Yasuo Fujishima *Kyoto University*, Megumi Kaneko *Kyoto University*, Hideaki Sakai *Kyoto University*, Riichi Kudo *NTT Corporation*, Tomoki Murakami *NTT Corporation*

The paper considers a coupling wave canceller for full--duplex radio relay station using adaptive antenna array. Taking advantage of the fact that coupling waves to be cancelled at the relay station consist of its own past transmitted signals, we propose a beamforming method using not only received signals at actual antenna elements but also virtual received signals, which are generated in the relay station with artificial channel impulse responses, that is to say, virtual coupling wave paths. With the approach, the proposed method can eliminate coupling waves without increasing the number of actual antenna elements even when the number of coupling wave paths is large due to high--speed communications. Computer simulation results show that the proposed method achieves coupling wave cancellation with smaller number of antenna elements

### An Adaptive MAP Speech Spectral Amplitude Estimator Combined With a Zero Phase Noise Suppressor

Sayuri Kohmura *Osaka University*, Arata Kawamura *Osaka University*, Youji Iiguni *Osaka University*

We previously proposed an efficient MAP (Maximum {\it a posteriori}) speech spectral amplitude estimator for stationary noise suppression. Although this method can strongly reduce stationary noise signals, it cannot reduce impulsive noise signals, such that a thunder, clap, and other impact noise signals. On the other hand, we also previously proposed a zero phase noise suppression method to achieve impulsive noise reduction, where its effectiveness was confirmed through some simulations. In this paper, we combine these two effective noise reduction methods and achieve a noise suppressor which can remove both of stationary and impulsive noise signals. We evaluate its noise reduction capability for some types of noise. The simulation results show the effectiveness of the proposed noise suppression method.

## OS.30-SIPTM.3 Control, Optimization and Information Processing for Smart Grid (II)

Session Chairs: Anthony Kuh, Urbashi Mitra, Anna Scaglione Location: Franklin Hills

### An Optimal Dynamic Pricing and Schedule Approach in V2G

Yi Han *University of Maryland*, Yan Chen *University of Maryland*, Feng Han *University of Maryland*, K. J. Ray Liu *University of Maryland*

Smart Grid (SG) can greatly improve the efficiency and reliability of traditional grid. As a promising feature of future SG, the Vehicle-to-Grid (V2G) technique exhibits great potential to balance the supply and demand of electrical power as well as integrate renewable energy.

Recently, some V2G-based schemes have been proposed to leverage the energy-storage capability of electric vehicles (EVs) to effectively reduce energy loss caused by supply-demand mismatches. However, most of the existing schemes rely on the assumption that the charge station is profit-neutral, lacking of adequate incentive to the charge stations for wide deployment. In this paper, we investigate a scenario where the charge station is modelled as an entity driven by its own profit. We formulate the interactions between the charge station and multiple EVs as a game, in which two kinds of EVs, cooperative EVs and selfish EVs, are considered. Regarding the intelligence of selfish EVs, a dynamic pricing over multiple time slots is developed from the charge station's perspective to maximize its own profit. Both theoretical analysis and simulation results show that through our scheme of dynamic pricing for selfish EVs and charging scheduling for cooperative EVs, the charge station can maximize its profit while EVs maximize their utilities.

### Power Grid Vulnerability Measures to Cascading Overload Failures

Zhifang Wang *Virginia Commonwealth University*, Anna Scaglione *University of California, Davis*, Robert Thomas *Cornell University*

Cascading failure in power grids has long been recognized as a sever security threat to national economy and society, which happens infrequent but can cause severe consequences. The causes of cascading phenomena can be extremely complicated due to the many different and interactive mechanisms such as transmission overloads, protection equipment failures, transient instability, voltage collapse, etc. In the literature a number of vulnerability measures to cascading failures have been proposed to identify the most critical components in the grid and evaluate the damages caused by the removal of such recognized components from the grid. In this paper we propose a novel power grid vulnerability measure called the minimum safety time after 1 line trip, defined based on the stochastic cascading failure model [1]. We compare its performance with several other vulnerability measures through a set of statistical analysis.

### Quickest Detection of Unknown Power Quality Events for Smart Grids

Xingze He *University of Southern California*, Man-On Pun *Huawei Technologies*, C.-C. Jay Kuo *University of Southern California*

In this work, we study a change-point approach to provide the quickest detection of power quality (PQ) event occurrence for smart grids. Despite that both the occurrence time and the PQ event type are unknown beforehand, knowledge of the statistics of post-PQ event signals is required to implement the change-point approach. To circumvent this obstacle, we propose to model the unknown PQ events using different statistical distributions, namely the Gaussian, Gamma and inverse Gamma distributions. It is shown by computer simulation that all distributions under consideration can provide accurate PQ event detection. In particular, the inverse Gamma distribution demonstrates the most promising performance in our simulation.

### Some Problems in Demand Side Management

Lingwen Gan *California Institute of Technology*, Libin Jiang *California Institute of Technology*, Steven Low *California Institute of Technology*, Ufuk Topcu *California Institute of Technology*, Changhong Zhao *California Institute of Technology*

We present a sample of problems in demand side management in future power systems and illustrate how they can be solved in a distributed manner using local information. First, we consider a set of users served by a single load-serving entity (LSE). The LSE procures capacity a day ahead. When random renewable energy is realized at delivery time, it manages user load through real-time demand response and purchases balancing power on the spot market to meet the aggregate demand. Hence optimal supply procurement by the LSE and the consumption decisions by the users must be coordinated over two timescales, a day ahead and in real time, in the presence of supply uncertainty. Moreover, they must be computed jointly by the LSE and the users since the necessary information is distributed among them. We present distributed algorithms to maximize expected social welfare. Instead of social welfare, the second problem is to coordinate electric vehicle charging to fill the valleys in aggregate electric demand profile, or track a given desired profile. We present synchronous and asynchronous algorithms and prove their convergence. Finally, we show how loads can use locally measured frequency deviations to adapt in real time their demand in response to a shortfall in supply. We design decentralized demand response mechanism that, together with the swing equation of the generators, jointly maximize disutility of demand rationing, in a decentralized manner.

### Scale Invariance and Long-Range Dependence in Smart Energy Grids

Marco Levorato *University of Southern California*, Urbashi Mitra *University of Southern California*

The shift from the traditional energy grid to the SmartGrid makes the features of scale invariance and long-range dependence, traditionally examined in reference to communication networks, extremely relevant to the modeling, analysis and design of modern energy grids. The present paper reviews mathematical concepts and tools central for the understanding and analysis of these phenomena and contextualizes them to the energy scenario. The framework proposed herein enables, in addition to a more accurate modeling and design of smart energy grids, the definition of novel algorithms for the detection of events, e.g., anomalies, in SmartGrids.

## OS.31-IVM.12 Image/Video Retrieval and Multimedia Applications

Session Chair: Ming-Sui Lee Location: Whitley Heights

### Social Album: Linking and Merging Online Albums based on Social Relationship

Kai-Yin Cheng *National Taiwan University*, Tzu-Hao Kuo *National Taiwan University*, Yu-Ting Wong *National Taiwan University*, Ming-Sui Lee *National Taiwan University*, Bing-Yu Chen *National Taiwan University*

This work designs a novel prototype system, Social Album, by utilizing social relationship data to link and merge online albums of individuals together. Field study results indicate that co-event albums related to more than one participating individual are the majority of online albums. Two different views are designed based on feedbacks from the interviews: the indexing view provides a metro-map such as an overview of the linked albums, while the browsing view allows individuals to peruse photos without looking at mis-aligned and duplicate photos from merged albums. Hence, through our system, Social Album, to share and gather co-event photos becomes much easier than before, and to browse the photos in the co-event albums also becomes more efficient while still keeping the comprehensiveness of the whole event. Finally, a user study demonstrates the usefulness of the proposed system.

### An Efficient VLSI Architecture of Parallel Bit Plane Encoder Based on CCSDS IDC

Yi Lu *Xidian University*, Jie Lei *Xidian University*, Yunsong Li *Xidian University*

The Bit-Plane Encoder (BPE) is the key part of CCSDS-IDC that encodes the coefficients of 2-D Discrete Wavelet Transform (DWT). In common sense, it is considered as the bottleneck of throughput performance and hardware resource consumption. An efficient VLSI architecture of BPE implemented with parallel and pipeline technology is proposed in this paper. In this architecture, the whole bit planes of each DWT coefficient could be encoded simultaneously and pipeline is utilized in three functional parts of the bit plane coding. The proposed architecture has been implemented in a Xilinx FPGA, its throughput could be improved three times while its resource consumption is only about a quarter comparing with the published architectures.

### Video Instance Search for Embedded Marketing

Ting-Chu Lin *Academia Sinica*, Jau-Hong Kao *Industrial Technology Research Institute*, Chin-Te Liu *Industrial Technology Research Institute*, Chia-Yin Tsai *National Taiwan University*, Yu-Chiang Frank Wang *Academia Sinica*

With the rise of online sharing platforms such as YouTube, advertisers become more interested in providing relevant advertisements (ads) when the embedded products are presented in videos during broadcast, so that the number of hits and potential customers will be increased. Given the product image of interest, we present a framework which allows the advertisers or video deliverers to automatically detect the embedded products throughout the video, so that relevant ads or latest product information can be delivered to the viewers accordingly. We advance the boundary preserving dense local regions (BPLR) as the local descriptors for the query and each video frame, and utilize different types of features to describe the local region. To make our framework robust yet efficient, we reduce the search space by applying the technique of inverted index, and we propose a probabilistic framework to identify the video frames in which the product of interest is presented. Experiments on TRECVID, commercial, and movie datasets confirm the effectiveness of our proposed framework.

### Learning Sparse Dictionaries for Saliency Detection

Karen Guo *National Tsing Hua University*, Hwann-Tzong Chen *National Tsing Hua University*

We present a new method of predicting the visually salient locations in an image. The basic idea is to use the sparse coding coefficients as features and find a way to reconstruct the sparse features into a saliency map. In the training phase, we use the images and the corresponding fixation values to train a feature-based dictionary for sparse coding as well as a fixation-based dictionary for converting the sparse coefficients into a saliency map. In the test phase, given a new image, we can get its sparse coding from the feature-based dictionary and then estimate the saliency map using the fixation-based dictionary. We evaluate our results on two datasets with the shuffled AUC score and show that our method is effective in deriving the saliency map from sparse coding information.

### Voting-Based Depth Map Refinement and Propagation for 2D to 3D Conversion

Yu-Hsiang Chiu *National Taiwan University*, Ming-Sui Lee *National Taiwan University*, Wei-Kai Liao *The White Rabbit Entertainment*

In this paper, a voting-based filter which is capable of enhancing the quality of depth map sequence and interpolates missing frames is proposed. The main concept is that if the information in the filter window is not consistent, then only the multitude decides the output. The minority will be considered as outliers. Compare to other depth map refinement method using joint bilateral filter, the outlier detection of the proposed scheme ensures that only appropriate information is involved so that the halo effects can be avoided. Moreover, in order to refine a sequence of depth maps, a space-time filtering extension is proposed. This extension refines each depth map according to the information of several adjacent frames rather than only one frame. As a result, the proposed method is capable of interpolating missing depth maps of a sequence and the errors on the depth maps are successfully reduced. The experimental results demonstrate that the voting-based filter not only provides a depth map sequence with good quality and temporal consistency but also provides several post-processing for depth maps in 2D to 3D conversion due to its flexibility.

## OS.32-SLA.12 Recent Topics in Speech Processing (I)

Session Chair: Hiroshi Saruwatari Location: Mt. Olympus

### Hierarchical Prosodic Boundary Prediction For Uyghur TTS

Hamdulla Askar *Xinjiang University*, Guljamal Mamateli *Xinjiang University*, Askar Rozi *Xinjiang University*, Imam Sayyare *Xinjiang University*

Correct prosodic boundary prediction is crucial for the quality of synthesized speech. This paper presents the prosodic hierarchy of Uyghur-language which belongs to agglutinative language. A two-layer bottom-up hierarchical approach based on conditional random fields (CRF) is used for predicting prosodic word (PW) and prosodic phrase (PP) boundaries. In order to disambiguate the confusion between different prosodic boundaries at punctuation sites, CRF based prosodic boundary determination model is used and integrated with bottom-up hierarchical approach. Word suffix feature is considered useful for prosodic boundary prediction and added into the feature sets. The experimental results show that the proposed method successfully resolves the confusion between different prosodic boundaries. Consequently, further enhance the accuracy of prosodic boundary prediction.

### A Research of Dependencies Between Frequency Components and Speaker Characteristics

Hyon Songgun *Tianjin University*, Wang Hongcui *Tianjin University*, Jianguo Wei *Tianjin University*, Jianwu Dang *JAIST/Tianjin University*

This paper proposes a new speaker feature extraction method, which is based on the non-uniformly distributed speaker information in frequency bands. In order to discard the linguistic information effectively, in this study, we first examine the differences of the distribution of individual information in the frequency region when a speaker utters different phonemes. Then we adopt an improved F-ratio, a phoneme mean F-ratio, to measure the dependences between frequency components and individual characteristics. According to the result of the analysis, we adopt an adaptive frequency filter to extract more discriminative feature. The new feature was combined with GMM speaker models and applied to the speaker recognition database which includes 50 persons. The experiment shows that the error rate using the proposed feature is reduced by 28.5% compared with the F-ratio feature, and reduced by 68.02% compared with the MFCC feature.

### Development of Note-Taking Support System with Speech Interface

Kohei Ota *University of Yamanashi*, Hiromitsu Nishizaki *University of Yamanashi*, Yoshihiro Sekiguchi *University of Yamanashi*

This paper describes a note-taking support system with a speech interface. To solve problems with existing note-taking methods, we implemented a speech interface consisting of a combination of a touch panel and graphical user interface in a note-taking support system. As a system user listens to a speech, the content of the speech is recognized and displayed on the system's screen. Users can take notes by simply touching or tracing the words automatically displayed on the screen. In addition, the system can support keyboard and handwritten input to cope with speech recognition errors. The developed system was experimentally compared with another note-taking method, a text editor on a personal computer. Most of the subjects could take a note more quickly using the system than using the text editor. The effectiveness of the system was demonstrated in the experiment.

### An Online Evaluation System for English Pronunciation Intelligibility for Japanese English Learners

Hiroshi Kibishi *Toyohashi University of Technology*, Seiichi Nakagawa *Toyohashi University of Technology*

We have previously proposed a statistical method for estimating pronunciation proficiency and intelligibility of presentations delivered in English by Japanese speakers. In an offline test, we also evaluated possibly-confused pairs of phonemes that are often mispronounced by Japanese native speakers. In this study, we developed an online evaluation system for English spoken by Japanese speakers using offline

techniques and carried out an evaluation to obtain the effect thereof based on experimental results. The results showed that both the objective and subjective evaluations improved when using this system.

**Real-Time Semi-Blind Speech Extraction with Speaker Direction Tracking on Kinect**

Yuji Onuma *Nara Institute of Science and Technology*, Noriyoshi Kamado *Nara Institute of Science and Technology*, Hiroshi Saruwatari *Nara Institute of Science and Technology*, Kiyohiro Shikano *Nara Institute of Science and Technology*

In this paper, speech recognition accuracy improvement is addressed for ICA-based multichannel noise reduction in spoken-dialogue robot. First, to achieve high recognition accuracy for the early utterance of the target speaker, we introduce a new rapid ICA initialization method combining robot image information and a prestored initial separation filter bank. From this image information, an ICA initial filter fitted to the user's direction can be used to save the user's first utterance. Next, a new permutation solving method using a probability statistics model is proposed for realistic sound mixtures consisting of point-source speech and diffuse noise.We implement these methods using user tracking on Microsoft Kinect and evaluate it by speech recognition experiment in the real environment. The experimental results show that the proposed approaches can markedly improve the word recognition accuracy.

## 14:00 - 15:40

### OS.33-SLA.13 Fundamental Technologies in Modern Speech Processing (I)

Session Chairs: Sadaoki Furui, Li Deng Location: Doheny

**Survey on Approaches to Speech Recognition in Reverberant Environments**

Takuya Yoshioka *NTT Corporation*, Armin Sehr *University of Erlangen-Nuremberg*, Marc Delcroix *NTT Corporation*, Keisuke Kinoshita *NTT Corporation*, Roland Maas *University of Erlangen-Nuremberg*, Tomohiro Nakatani *NTT Corporation*, Walter Kellermann *University of Erlangen-Nuremberg*

This paper overviews the state of the art in reverberant speech processing from the speech recognition viewpoint. First, it points out that the key to successful reverberant speech recognition is to account for long-term dependencies between reverberant observations obtained from consecutive time frames. Then, a diversity of approaches that exploit the long-term dependencies in various ways is described, ranging from signal and feature dereverberation to acoustic model compensation tailored to reverberation. A framework for classifying those approaches is presented to highlight similarities and differences between them.

**Recent Developments in Large Vocabulary Continuous Speech Recognition**

George Saon *IBM T. J. Watson Research Center*, Jen-Tzung Chien *National Chiao Tung University*

This paper overviews a series of recent approaches to front-end processing, acoustic modeling, language modeling, and back-end search and system combination which have made contributions for large vocabulary continuous speech recognition (LVCSR) systems. These approaches include the feature transformations, speaker-adaptive features, and discriminative features in front-end processing, the feature-space and model-space discriminative training, deep neural networks, and speaker adaptation in acoustic modeling, the backoff smoothing, large-span modeling, and model regularization in language modeling, and the system combination, cross-adaptation, and boosting in search and system combination. Some future directions for LVCSR research are also addressed.

**Microphone Array Processing for Distant Speech Recognition: Towards Real-World Deployment**

Kenichi Kumatani *Disney Research*, Takayuki Arakawa *NEC*, Kazumasa Yamamoto *Toyohashi University of Technology*, John McDonough *Carnegie Mellon University/Voci Technologies, Inc.*, Bhiksha Raj *Carnegie Mellon Univerity*, Rita Singh *Carnegie Mellon Univerity*, Ivan Tashev *Microsoft Research*

Distant speech recognition (DSR) holds out the promise of providing a natural human computer interface in that it enables verbal interactions with computers without the necessity of donning intrusive body- or head-mounted devices. Recognizing distant speech robustly, however, remains a challenge. This paper provides a overview of DSR systems based on microphone arrays. In particular, we present recent work on acoustic beamforming for DSR, along with experimental results verifying the effectiveness of the various algorithms described here; beginning from a word error rate (WER) of 14.3% with a single microphone of a 64-channel linear array, our state-of-the art DSR system achieved a WER of 5.3%, which was comparable to that of 4.2% obtained with a lapel microphone. Furthermore, we report the results of speech recognition experiments on data captured with a popular device, the Kinect. Even for speakers at a distance of four meters from the Kinect, our DSR system achieved acceptable recognition performance on a large vocabulary task, a WER of 24.1%, beginning from a WER of 42.5% with a single array channel.

**Exploiting Speech Production Information for Automatic Speech and Speaker Modeling and Recognition -- Possibilities and New Opportunities**

Vikram Ramanarayanan *University of Southern California*, Prasanta Ghosh *IBM Research India*, Adam Lammert *University of Southern California*, Shrikanth Narayanan *University of Southern California*

We consider the potential for incorporating direct, or inferred, speech production knowledge in speech technology development. We first review the technologies that can be used to capture speech articulation information. We discuss how meaningful (speech and speaker) representations can be derived from articulatory data thus captured and further how they can be estimated from the acoustics in the absence of these direct measurements. We present some applications that have used speech production information to further the state-of-the-art in automatic speech and speaker recognition. We also offer an outlook on how such knowledge and applications can in turn inform scientific understanding of the human speech communication process.

### OS.34-IVM.13 Recent Topics in Image, Video, and Multimedia Processing (II)

Session Chair: Koichi Shinoda Location: Beachwood

**Color-Tone Similarity on Digital Images**

Hisakazu Kikuchi *Niigata University*, Heikki Huttunen *Tampere University of Technology*, Junghyeun Hwang *Niigata University*, Masahiro Yukawa *Niigata University*, Shogo Muramatsu *Niigata University*, Jaeho Shin *Dongguk University*

A color-tone similarity index (CSIM) between two color images is presented and another index, picture similarity index (PSIM), is also given for a comprehensive similarity com-parison between color images. CSIM is defined by a statistical analysis of cumulative histograms in a hue-oriented color space. It characterizes the color distributions, while the existing structural similarity index reflects the spatial structure involved with grayscale images. The behaviors of CSIM are checked by the comparisons of color code chips. Experimental results are given. The proposed indexes combined with SSIM are hopeful to provide a tool for color image quality analysis (IQA).

**Efficient Model Training for HMM-based Person Identification by Gait**
Muhammad Aqmar *Tokyo Institute of Technology*, Koichi Shinoda *Tokyo Institute of Technology*, Sadaoki Furui *Tokyo Institute of Technology*

In gait-based person identification, statistical methods such as hidden Markov models (HMMs) have been proved to be effective. Their performance often degrades, however, when the amount of training data for each walker is insufficient. In this paper, we propose walker adaptation and walker adaptive training, where the data from the other walkers are effectively utilized in the model training. In walker adaptation, maximum likelihood linear regression (MLLR) is used to transform the parameters of the walker-independent model to those of the target walker model. In walker adaptive training, we effectively exclude the inter-walker variability from the walker-independent model. In our evaluation, our methods improved the identification performance even when the amount of data was extremely small.

**Real-Time Both Hands Tracking Using CAMshift with Motion Mask and Probability Reduction by Motion Prediction**
Ryosuke Araki *Waseda University*, Takeshi Ikenaga *Waseda University*, Seiichi Gohshi *Kogakuin University*

Hand gesture interfaces are more intuitive and convenient than traditional interfaces. They are the most important parts in the relationship between users and devices. Hand tracking for hand gesture interfaces is an active area of research in image processing. However, previous works have limits such as requiring the use of multiple camera or sensor, working only with single color background, etc. This paper proposes a real-time both hands tracking algorithm based on "CAMshift (Continuous Adaptive Mean Shift Algorithm)" using only a single camera in multi-color backgrounds. In order to track hands robustly, the proposed algorithm uses "motion mask" to combine color and movement probability distributions and "probability reduction" for multi-hand tracking in nonlimiting environments. Experimental results demonstrate that this algorithm can precisely track both hands of an operator in multicolor backgrounds and process the VGA size input sequences from a web camera in real time (about 25 fps).

**High Contrast Tone-mapping and its Application for Two-layer High Dynamic Range Coding**
Takao Jinno *Toyohashi University of Technology*, Hiroya Watanabe *University of Kitakyushu*, Masahiro Okuda *University of Kitakyushu*

Many applications for High Dynamic Range (HDR) images require tone-mapping operations that preserve details in whole luminance range. This paper proposes a high contrast tone-mapping operator using a multi-scale contrast enhancement, and uses it for a high efficiency two-layer HDR coding. To visualize minute details, the high contrast tone-mapping operator often results in hard enhancement. In many conventional two-layer coding methods, it degrades compression efficiency. In contrast our method can achieve both of the high contrast and high compression efficiency. Moreover this paper can perform two types of tone-mapping which generates the images with strong enhancement and natural look. This paper shows the validity of our methods through some experimental results.

**Gradient-based Global Features and Its Application to Image Retargeting**
Izumi Ito *Tokyo Institute of Technology*

We propose gradient-based global features and its application to image retargeting. The proposed features are used for an importance map for image retargeting, which represents rough location of salient objects in an image. We focus on areas rather than points and lines to be assigned as an important part. The information about areas in multiple layers provides global features. Experimental results compared to the state-of- the-art salient features for image retargeting demonstrate the effectiveness of the proposed features.

**Finding Canonical Views by Measuring Features on the Viewing Plane**
Wencheng Wang *Institutue of Software, CAS*, Liming Yang *Institutue of Software, CAS*, Dongxu Wang *Institutue of Software, CAS*

Canonical views are referred to the classical three-quarter views of a 3D object, always preferred by human beings, because they are stable and able to produce more meaningful and understandable images for the viewer. Unlike existing methods to measure features in the 3D space for view selection, this paper proposes to measure features on the viewing plane, taking into account the influence of feature deformation due to perspective projection on view evaluation. Meanwhile, we try to have more features perceptible instead of having preferred features displayed more in a good view, which is aimed by existing methods. As a result, we can effectively obtain canonical views with only geometry computation, without troublesome semantic computation, which are always needed in existing techniques for obtaining good views.

## OS.35-SIPTM.4 Recent Advances in Sparse and Nonlinear Adaptive Signal Processing

Session Chairs: Mrityunjoy Chakraborty, Tokunbo Ogunfunmi Location: Runyon

**An Alternative Kernel Adaptive Filtering Algorithm for Quaternion-Valued Data**
Tokunbo Ogunfunmi *Santa Clara University*, Thomas Paul *Santa Clara University*

Nonlinear adaptive filters are getting more common and are useful especially where performance of linear adaptive filters may be unacceptable. Such areas include communications, image processing and biological systems. Quaternion valued data has also been drawing recent interest in various areas of statistical signal processing, including adaptive filtering, image pattern recognition, and modeling and tracking of motion. The benefit of quaternion valued processing includes performing data transformations in a 3 or 4-dimensional space in a more convenient fashion than using vector algebra. In this paper we describe an alternative kernel adaptive filter for quaternion valued data we refer to as the involution Quaternion Kernel Least Mean Square (iQuat-KLMS) algorithm. The approach is based on the Quaternion KLMS (Quat-KLMS) algorithm obtained previously, as well as the recently developed involution gradient (i-gradient). A modified HR Calculus for Hilbert Space is used for finding cost function gradients defined on a quaternion RKHS. Simulation tests with a synthetic quaternion channel are used to verify the benefit of iQuat-KLMS in convergence compared to Quat-KLMS.

**Subspace Based Blind Sparse Channel Estimation**
Kazunori Hayashi *Kyoto University*, Hiroki Matsushima *Kyoto University*, Hideaki Sakai *Kyoto University*, Elisabeth Carvalho *Aalborg University*, Petar Popovski *Aalborg University*

The paper proposes a subspace based blind sparse channel estimation method using L1-L2 optimization by replacing the L2-norm minimization in the conventional subspace based method by the L1-norm minimization problem. Numerical results confirm that the proposed method can significantly improve the estimation accuracy for the sparse channel, while achieving the same performance as the conventional subspace method when the channel is dense. Moreover, the proposed method enables us to estimate the channel response with unknown channel order if the channel is sparse enough.

### An Efficient Iterative Method for Basis Pursuit Adaptive Filters for Sparse Systems

Steven Grant *Missouri S&T*, Pratik Shah *Missouri S&T*, Jacob Benesty *University of Quebec*

The "proportionate" family of adaptive filters has been in use over the past decade. Their fast convergence for sparse systems makes them particularly useful in the network echo canceller application. Recently, an iterative form of the proportionate affine projection algorithm (PAPA), derived from the basic principles of basis pursuit, has been shown to have remarkably fast convergence for such sparse systems. The number of samples for convergence is proportional to the sparseness of the system which means that often full convergence occurs in fewer samples than the length of the system's impulse response. Here, we introduce a lower complexity implementation with the same performance that is an iterative version of proportionate normalized least mean squares (PNLMS).

### Sparse Recovery from Convolved Output in Underwater Acoustic Relay Networks

Sunav Choudhary *University of Southern California*, Urbashi Mitra *University of Southern California*

This paper explores criteria for unique recovery from blind deconvolution under sparsity priors. Additionally regularizing functions stemming from this problem framework are developed. For key cases, it is possible to ensure unique recoverability given the regularized problem statement. The uniqueness results are informed by a matrix completion-based viewpoint of blind deconvolution. Furthermore, this perspective enables characterization of why blind deconvolution with two sparse inputs is an inherently hard problem. Two blind deconvolution algorithms are proposed which do not rely on alternating between the estimation of one input signal, while holding the other constant. Evaluation of the algorithms is done via simulation and shown to significantly outperform a previously proposed method. Furthermore, numerical illustration of recovery failure considering sparsity of input signals that do not satisfy the recovery constraints is also provided.

### A Zero Attracting Proportionate Normalized Least Mean Square Algorithm

Rajib Lochan Das *Indian Institute of Technology*, Mrityunjoy Chakraborty *Indian Institute of Technology*

The proportionate normalized least mean square (PNLMS) algorithm, a popular tool for sparse system identification, achieves fast initial convergence by assigning independent step sizes to the different taps, each being proportional to the magnitude of the respective tap weight. However, once the active (i.e., non-zero) taps converge, the speed of convergence slows down as the effective step sizes for the inactive (i.e., zero or near zero) taps become progressively less. In this paper, we try to improve upon both the convergence speed and the steady state excess mean square error (EMSE) of the PNLMS algorithm, by introducing a l1 norm (of the coefficients) penalty in the cost function which introduces a so-called zero-attractor term in the PNLMS weight update recursion. The zero attractor induces further shrinkage of the coefficients, especially of those which correspond to the inactive taps and thus arrests the slowing down of the convergence of the PNLMS algorithm, apart from bringing down the steady state EMSE. We have also modified the cost function further generating a reweighted zero attractor which helps in confining the "Zero Attraction" to the inactive taps only.

## OS.36-SLA.14 Recent Advances in Signal Processing/Filter Applications

Session Chair: Xiangui Kang Location: Laurel

### A Comb Filter with Adaptive Notch Gain for Periodic Noise Reduction

Yosuke Sugiura *Osaka University*, Arata Kawamura *Osaka University*, Youji Iiguni *Osaka University*

A comb filter is used to eliminate a periodic noise signal from an observed signal. For extracting the desired signal, one of the most important factors of the comb filter is the notch gain which controls an elimination quantity of the observed signal at noise frequencies. Conventional comb filters employ a pre-designed notch gain under the assumption that the appropriate notch gain is known. Unfortunately, in many practical situations, the appropriate notch gain is unknown and often changes. In this paper, we propose a new comb filter with the adaptive notch gain to automatically achieve the appropriate notch gain. In the proposed method, we utilize an adaptive line enhancer (ALE) instead of the conventional notch gain multiplier. When the ALE completely estimates the periodic noise signal, the ALE's frequency response directly gives the appropriate notch gain. Simulation results show the effectiveness of the proposed adaptive comb filter.

### A Directional and Shift-Invariant Transform Based on M-channel Rational-Valued Cosine-Sine Modulated Filter Banks

Seisuke Kyochi *University of Kitakyushu*, Taizo Suzuki *Nihon University*, Yuichi Tanaka *Tokyo University of Agriculture and Technology*

This paper proposes a directional and shift-invariant transform based on M-channel rational-valued cosine-sine modulated filter banks (R-CSMFBs) for the practical implementation on hardware devices. M-channel CSMFBs can be easily designed by the modulation of a prototype filter and achieve a good stopband attenuation. In addition, in our previous work, the directionality and the shift-invariance of CSMFBs have been theoretically clarified. Thus, they can be an alternative choice of the dual-tree complex wavelet transform (DTCWT) which is one of the most popular directional and shift-invariant transforms. In this paper, it is shown that the proposed lifting-based structure of the R-CSMFB can also achieve rich directional selectivity and the shift-invariance even if the lifting coefficients are rounded to rational values. Finally, the R-CSMFB can provide better stopband attenuation and image denoising performance than that of the conventional M-channel rational-valued DTCWT in the simulation.

### Robust Median Filtering Forensics Based on the Autoregressive Model of Median Filtered Residual

Xiangui Kang *Sun Yat-Sen University*, Matthew Stamm *University of Maryland,* Anjie Peng *Sun Yat-Sen University*, K. J. Ray Liu *University of Maryland*

One important aspect of multimedia forensics is exposing an image's processing history. Median filtering is a popular noise removal and image enhancement tool. It is also an effective tool in anti-forensics recently. An image is usually saved in a compressed format such as the JPEG format. The forensic detection of median filtering from a JPEG compressed image remains challenging, because typical filter characteristics are suppressed by JPEG quantization and blocking artifacts. In this paper, we introduce a robust median filtering detection scheme based on the autoregressive model of median filter residual. Median filtering is first applied on a test image and the difference between the initial image and the filtered output image is called the median filter residual (MFR). The MFR is used as the forensic fingerprint. Thus, the interference from the image edge and texture, which is regarded as a limitation of the existing forensic methods, can be reduced. To capture the statistical properties of the MFR, we fit it to an autoregressive (AR) model. We then use the AR coefficients as features for median filter detection. Experimental results show that the proposed median filtering detection method is very robust to JPEG post-compression with a quality factor as low as 30. It distinguishes well between median filtering and other manipulations, such as Gaussian filtering, average filtering, and rescaling and performs well on low-resolution images of size 32 _ 32. The proposed method achieves not only much better performance than the existing state-of-the-art methods, but also has very small dimension of feature, i.e., 10-D.

### Nonlinear Signal Processing for Compensating Nonlinear Distortion of Louspeakers

Kenta Iwai *Kansai University*, Yoshinobu Kajikawa *Kansai University*

In this paper, we propose a 3rd-order nonlinear IIR filter for compensating nonlinear distortions of loudspeaker systems. The 2nd-order nonlinear IIR filter based on the Mirror filter is used for reducing nonlinear distortions of loudspeaker systems. However, the 2nd-order nonlinear IIR filter cannot reduce nonlinear distortions at high frequencies because it does not include the nonlinearity of the self-inductance of loudspeaker systems. On the other hand, the proposed filter includes the effect of such self-inductance and thus can reduce nonlinear distortions at high frequencies. Experimental results demonstrate that the proposed filter can realize a reduction by 3.2 dB more than the conventional filter on intermodulation distortions at high frequencies.

### Classifying NMF Components Based on Vector Similarity for Speech and Music Separation

Nengheng Zheng *Shenzhen University*, Xia Li *Shenzhen University*, Yi Cai *Shenzhen University*, Tan Lee *CUHK*

This paper presents a nonnegative matrix factorization (NMF) components classification algorithm for single-channel speech and music separation. Music only and music-speech mixture segments are firstly classified from the audio stream via audio segmentation technique. Then NMF is applied for signal decomposition. The basis matrix of the NMF output of music only segments provides the prior knowledge of music component in the mixture signal. NMF components, i.e. basis and gain vectors of the mixture signal are classified into speech and music based on the vector similarity between each basis vector and the priori music basis matrix. A set of SNR-dependent thresholding coefficients are empirically determined for the classification. The separated speech and music signals are reconstructed from the respectively classified NMF components. Experimental results show the effectiveness of the proposed method for speech and music separation, and its superior performance over the traditional NMF-based separation methods.

## OS.37-SPS.3 Sparse and Feature Representation for Image/Video Restoration

Session Chairs: Jiaying Liu, Weisheng Dong, Zongming Guo Location: Trousdale Estates

### Review of Image Interpolation and Super-resolution

Wan-Chi Siu *Hong Kong Polytechnic University*, Kwok Wai Hung *Hong Kong Polytechnic University*

Image/video interpolations and super-resolution are topics of great interest. Their applications include HDTV, image coding, image resizing, image manipulation, face recognition and surveillance. The objective is to increase the resolution of an image/video through upsampling, deblurring, denoising, etc. This paper reviews the development of various approaches on image interpolation and super-resolution theory for image/video enlargement in multimedia applications. Some basic formulations will be derived such that readers can make use of them to design their own, practical and efficient interpolation algorithms. New results, such as hole filling using non-local means for 3D video synthesis and fast interpolation using a simplified image model will be introduced. New directions and trends will also be discussed at the end of the paper.

### Super Resolution with Edge-Constrained Motion Estimation

Yue Zhuo *Peking University*, Jiaying Liu *Peking University*, Mading Li *Peking University*, Zongming Guo *Peking University*

Motion estimation is a critical step for most reconstruction-based super resolution methods. However, accurate motion estimation is difficult and the unavoidable error degrades performance of super resolution rapidly. In this paper, we present a robust way to perform super resolution by improving motion estimation. Beginning with feature points matching, we compute local motion parameter of feature point correspondences by weighted Lucas-Kanade algorithm. Then accurate motion field is estimated by support region search, which refers to edge information and considers discontinuity of motion boundary and consistency of motion field. Experimental results validate the efficacy of each step in the proposed algorithm and show that we produce super resolved images with higher quality.

### Super Resolution For Subpixel Based Downsampled Images

Ketan Tang *HKUST*, Oscar C. Au *HKUST*, Lu Fang *USTC*, Yuanfang Guo *HKUST*, Pengfei Wan *HKUST*, Lingfeng Xu *HKUST*

Subpixel-based downsampling is a new downsampling technique which utilizes the fact that each pixel in LCD is composed of three individually addressable subpixels. Subpixel-based downsampling can provide higher apparent resolution than pixel-based downsampling. In this paper we study the inverse problem of subpixel-based downsampling. We find that conventional pixel-based super resolution algorithms are not suitable for subpixel-based downsampled images due to the special downsampling pattern. In this paper we propose a super resolution algorithm specially for subpixel-based downsampled images, which uses piecewise autoregressive model to model spatial correlation of neighboring pixels, and incorporates the special data degradation term corresponding to the subpixel downsampling pattern. We formulate the super resolution problem as a constrained least square problem and solve it using Gauss-Seidel iteration. Experimental results demonstrate the effectiveness of the proposed algorithm.

### Image Deblurring with Low-rank Approximation Structured Sparse Representation

Weisheng Dong *Xidian University*, Guangming Shi *Xidian University*, Xin Li *West Virginia University*

In recent years sparse representation model (SRM) based image deblurring approaches have shown promising image deblurring results. However, since most of the current SRMs don't utilize the spatial correlations between the nonzero sparse coefficients, the SRM-based image deblurring methods often fail to faithfully recover sharp image edges. In this paper, a structured SRM is employed to exploit the local and nonlocal spatial correlation between the sparse codes. The connection between the structured SRM and the low-rank approximation model has also been presented, leading to an efficient low-rank based optimization algorithm. An effective image deblurring algorithm using the patch-based structured SRM is then proposed. Experimental results demonstrate the improvements of the proposed deblurring method over current state-of-the-art image deblurring methods.

### Face Super-Resolution Based on Singular Value Decomposition

Muwei Jian *Hong Kong Polytechnic University*, Kin Man Lam *Hong Kong Polytechnic University*

In this paper, a novel face image super-resolution approach based on singular value decomposition (SVD) is proposed. We prove that the singular values of an image at one resolution have approximately linear relationships with their counterparts at other resolutions. This makes the estimation of the singular values of the corresponding HR face images more reliable. From the signal-processing point of view, this can effectively preserve and reconstruct the dominant information in the HR face image. Interpolating the two other matrices obtained from the SVD of a LR face image does not change either the primary facial structure or the pattern of the face image. Furthermore, the mapping scheme for interpolating the matrices can be viewed as a “coarse-to-fine” estimation of HR face images, which uses the mapping matrices learned from the corresponding reference image pairs. Experimental results show that the proposed super-resolution scheme is effective and efficient.

**Robust Single Image Super-resolution Based on Gradient Enhancement**

Licheng Yu *Shanghai Jiao Tong University*, Hongteng Xu *Shanghai Jiao Tong University*, Yi Xu *Shanghai Jiao Tong University*, Xiaokang Yang *Shanghai Jiao Tong University*

In this paper, we propose an image super-resolution approach based on gradient enhancement. Local constraints are established to achieve enhanced gradient map, while the global sparsity constraints are imposed on the gradient field to reduce noise effects in super resolution results. We can then formulate the image reconstruction problem as optimizing an energy function composed of the proposed sharpness and sparsity regularization terms. The solution to this super-resolution image reconstruction is finally achieved using the well-known variable-splitting and penalty techniques. In comparison with the existing methods, the experimental results highlight our proposed method in computation efficiency and robustness to noisy scenes.

## OS.38-SIPTM.5 Signal and Information Processing of Energy Signals

Session Chairs: Anthony Kuh, Urbashi Mitra, Anna Scaglione Location: Franklin Hills

**Networked Loads in the Distribution Grid**

Zhifang Wang *Virginia Commonwealth University*, Xiao Li *University of California, Davis*, Vishak Muthukumar *University of California, Davis*, Anna Scaglione *University of California, Davis*, Sean Peisert *Lawrence Berkeley National Laboratory/ University of California, Davis*, Charles McParland *Lawrence Berkeley National Laboratory*

Central utility services are increasingly networked systems that use an interconnection of sensors and programmable logic controllers, and feed data to servers and human-machine interfaces. These systems are connected to the Internet so that they can be accessed remotely, and the network in these plants is structured according to the SCADA model. Although the physical systems themselves are generally designed with high degrees of safety in mind, and designers of computer systems are well advised to incorporate computer security principles, a combined framework for supervisory control of the physical and cyber architectures in these systems is still lacking. Often absent are provisions to defend against external and internal attacks, and even operator errors that might bypass currently standalone security measures to cause undesirable consequences. In this paper we examine a prototypical instance of SCADA network in the distribution network that handles central cooling and heating for a set of buildings. The electrical loads are networked through programmable logic controllers (PLCs), electrical meters, and networks that deliver data to and from servers that are part of a SCADA system, which has grown in size and complexity over many years.

**Instantaneous Frequency Estimation and Localization for ENF Signals**

Adi Hajj-Ahmad *University of Maryland*, Ravi Garg *University of Maryland*, Min Wu *University of Maryland*

Forensic analysis based on Electric Network Frequency (ENF) fluctuations is an emerging technology to authenticate multimedia recordings. This class of techniques requires extracting frequency fluctuations from multimedia recordings and comparing them with the ground truth frequencies, obtained from the power mains, at the corresponding time. Most current guidelines for frequency estimation from the ENF signal use non-parametric approaches. Such approaches have limited temporal-frequency resolution due to the tradeoffs of the time-frequency resolutions as well as computational power. To facilitate robust high-resolution matching, it is important to estimate instantaneous frequency using as few samples as possible. The use of subspace-based methods for high resolution frequency estimation is fairly new for ENF analysis. In this paper, a systematic study of several high resolution low-complexity frequency estimation algorithms is conducted, focusing on estimating the frequencies in short time-frames. After establishing the performance of several frequency estimation algorithms, a study towards using the ENF signal for estimating the location-of-recording is carried out. Experiments conducted on ENF data collected in several cities indicate the presence of location-specific signatures that can be exploited for future forensic applications.

**Real-Time Adaptive Distributed State Estimation in Smart Grids**

Soummya Kar *Carnegie Mellon University*, Jose Moura *Carnegie Mellon University*

The paper presents a fully distributed framework for sequential recursive state estimation in inter-connected electrical power systems. Specifically, the setup considered involves a grid partitioned into multiple control areas that communicate over a \emph{sparse} communication network. In the absence of a global sensor data fusion center (the conventional centralized SCADA) and with sensing model uncertainties, an adaptive distributed state estimation approach, the $\mathcal{DAE}$, is proposed in which the system control areas engage in a collaborative joint (model) learning and (state) estimation procedure through sequential information exchange over the pre-assigned communication network. The proposed distributed estimation methodology is recursive, in that, each system control area refines its state estimate at a given sampling instant by suitably combining its past estimate with the newly collected local measurement(s) and the information obtained from its communication neighbors. Under rather weak assumptions of global observability and connectivity of the control area communication network, the proposed distributed adaptive scheme is shown to yield consistent system state estimates (i.e., estimates that converge to the true system state in the large sample limit), the convergence rate being optimal in the Fisher information sense. As discussed, the proposed approach based on local communication and computation is suitable for real-time implementation as opposed to conventional centralized SCADA based estimation architectures with periodic data gathering and processing. This has the potential of being more responsive and adaptive to sensed data generated by advanced non-conventional sensing resources like the PMUs with significantly higher system sampling rates.

**Modeling Distributed PV Energy Using Stochastic Queuing Models**

Anthony Kuh *University of Hawaii*, Chuanyi Ji *Georgia Institute of Technology*, Yun Wei *Georgia Institute of Technology*

In the past few years there has been a tremendous growth in distributed PV generation on commercial and residential buildings. Increasing distributed PV generation has raised concerns about the stability of the distribution grid due to the intermittency of solar PV energy. Before smart grid optimization and control algorithms can be formulated we must obtain a better understanding of the behavior of the distributed PV energy contributions to the electrical grid. This paper develops stochastic models to model each distributed energy source using both spatial and temporal processing. A goal is to develop simple stochastic models that accurately model the distributed energy produced from the PV sources with possible storage so that key events (e.g. ramp downs due to cloud cover can be predicted). The production of energy from PV panels is modeled as a queue with inputs being the nonstationary solar irradiation, the energy produced modeled by a deterministic function, and a queue modeled by storage which can be sold to the grid or used by local loads. A second queue models solar irradiation with inputs being weather conditions (sunny, partly cloudy, cloudy).

## OS.39-SPS.4/BioSPS.5 Advances in Signal Processing Systems and Biomedical Signal Processing

Session Chair: Alex Kot Location: Whitley Heights

### EAG: Edge Adaptive Grid Data Hiding for Binary Image Authentication

Hong Cao *Institute for Infocomm Research*, Alex Chichung Kot *Nanyang Technological University*

This paper proposes a novel data hiding method for authenticating binary images through establishing dense edge adaptive grids (EAG) for invariantly selecting good data carrying pixel locations (DCPL). Our method employs dynamic system structure with carefully designed local content adaptive processes (CAP) to iteratively trace new contour segments and to search for new DCPLs. By maintaining and updating a location status map, we re-design the fundamental content adaptive switch and a protection mechanism is proposed to preserve the local CAPs' contexts as well as their corresponding outcomes. Different from existing contour-based methods, our method addresses a key interference issue and has unprecedentedly demonstrated to invariantly select a same sequence of DCPLs for an arbitrary binary host image and its marked versions for our contour-tracing based hiding method. Comparison also shows that our method achieves better trade-off between large capacity and good perceptional quality as compared with several prior works for representative binary text and cartoon images.

### Performance Improvement of Closed-Form Joint Diagonalizer of Non-Negative Hermitian Matrices

Akira Tanaka *Hokkaido University*, Miho Murota *Hokkaido University*

Joint diagonalization of a series of non-negative Hermitian matrices is one of important techniques in the fields of signal processing, such as blind source separation based on second order statistics. In our previous works, we introduced a closed-form solution of a joint diagonalizer of non-negative Hermitian matrices and also proposed a method for improving the performance of the solution for the cases where given series of Hermitian matrices are not jointly diagonalizable strictly. However, the performance of the method may degrade when the number of given Hermitian matrices are comparatively small. In this paper, we propose an improved version of the closed-form joint diagonalizer of given set of Hermitian matrices by increasing the number of Hermitian matrices virtually. Some numerical examples are also shown to verify the efficacy of the proposed method.

### Weighted-CS for Reconstruction of Highly Under-sampled Dynamic MRI Sequences

Dornoosh Zonoobi *NUS*, Ashraf A. Kassim *NUS*

This paper investigates the potential of the new Weighted-Compressive Sensing approach which overcomes the major limitations of other compressive sensing and current state-of-the-art methods for low-rate reconstruction of sequences of MRI images. The underlying idea of this approach is to use the image of the previous time instance to extract an estimated probability model for the image of interest, and then use this model to guide the reconstruction process. This is motivated by the observation that MRI images are hugely sparse in Wavelet domain and the sparsity changes slowly over time.

### An Anisotropic Diffusion Filter for Reducing Speckle Noise of Ultrasound Images Based on Separability

Shen Liu *Tianjin University*, Jianguo Wei *Tianjin University*, Bo Feng *Tianjin University*, Wenhuan Lu *Tianjin University*, Bruce Denby *ESPCI ParisTech/Université Pierre et Marie Curie*, Qiang Fang *Chinese Academy of Social Sciences*, Jianwu Dang *JAIST/Tianjin University*

Anisotropic diffusion is being widely used in reducing speckle noise of ultrasound images. However, the traditional anisotropic diffusion algorithms are poor at preserving edges and usually make the image edges blurred when denoising, which negatively affects the following image analysis. In this paper, we modify the standard speckle reducing anisotropic diffusion to increase its ability of detecting edges and suppress the smooth at edge by using the separability of images. We extract contours from the original images, denoised images by SRAD and the images denoised by our proposed method, respectively. We analyze and compare the accuracy of these three kinds of contours. The result shows the proposed method performs better in edge-preserve and gets better images of high quality than SRAD, which can contribute to get more accurate contours.

### Accelerating Householder Bidiagonalization with ARM NEON Technology

Wenjun Yang *Tsinghua University*, Zhenyu Liu *Tsinghua University*

Householder bidiagonalization is the first step of Singular Value Decomposition (SVD), an important algorithm in numerical linear algebra that is widely used in video processing. NEON is a general-purpose Single Instruction Multiple Data (SIMD) engine introduced in ARMv7 architecture, which is targeted to accelerate multimedia and signal processing on mobile platforms. In this paper, we propose a NEON-based implementation and optimization of Householder bidiagonalization, aiming at testifying the potential of NEON to handle with low-dimensional macroblocks if applied to future computing-intensive video codecs. Intrinsics and inline assembly, two most commonly used ways to utilize NEON, are compared in performance. Solutions to the problem of leftover elements in vectorization is also discussed. Our study finally shows that with hand-coded inline assembly and all kinds of optimization, our NEON implementation of Householder bidiagonlization will gain a speedup of 2.3 over the plain C version.

# Thursday, December 6, 2012

## 09:30 - 10:30

## OS.40-WCN.4 Wireless Communications and Networking (II)

Session Chair: Wan-Jen Huang Location: Doheny

### Filter-And-Forward Relay Design for OFDM Systems for Quality-of-Service Enhancement

Donggun Kim *KAIST*, Junyeong Seo *KAIST*, Youngchul Sung *KAIST*

In this paper, the filter-and-forward (FF) relay design for OFDM communication systems is considered to enhance the system performance over the conventional amplify-and-forward(AF) relaying scheme. The considered design criterion in this paper is to maximize the worst subcarrier channel signal-to-noise ratio (SNR) subject to the total relay transmit power constraint in order to improve the overall transmission quality-of-service. It is shown, by exploiting the eigen-property of circulant matrices and the structure of Toeplitz matrices, that the considered problem reduces to a semi-definite programming (SDP) problem. Numerical results are provided to validate our design method and the numerical results show that the proposed FF relay outperforms the AF scheme significantly.

### Preamble Design for Joint estimation of Channel and I/Q imbalance in MIMO-OFDM Systems

Emmanuel Manasseh *Hiroshima University*, Shuichi Ohno *Hiroshima University*, Masayoshi Nakamoto *Hiroshima University*

In this paper, preamble design for estimation of frequency selective channels and In-phase/Quadrature-phase (I/Q) imbalance in multiple input multiple output orthogonal frequency-division-multiplexing (MIMO-OFDM) systems is proposed. First we utilize convex optimization to optimize power of all active subcarriers, then we employ cross entropy (CE) optimization techniques to select optimal preamble sequence that minimizes the channel estimate mean squared error (MSE) while suppressing the effect of I/Q mismatch. To mitigate inter-antenna interferences, disjoint preamble sequences are utilized for each transmit antenna. An algorithm to guarantee that the preamble sequences are disjoint for each transmitter is proposed. Numerical simulations are provided to verify the advantages and effectiveness of the proposed preamble sequences over the conventional sequences.

## OS.41-SLA.15 Speech Recognition (III)

Session Chair: Atsuhiko Kai Location: Beachwood

### Soft-clustering Technique for Training Data in Age- and Gender-independent Speech Recognition

Daisuke Enami *Toyohashi University of Technology*, Faqiang Zhu *Toyohashi University of Technology*, Kazumasa Yamamoto *Toyohashi University of Technology*, Seiichi Nakagawa *Toyohashi University of Technology*

In this paper, we propose approaches for the Gaussian mixture model (GMM) based soft clustering of training data and the GMM- or/and hidden Markov model (HMM)-based cluster selection in age and gender-independent speech recognition. Typically, increasing the number of speaker classes leads to more specific models in speaker-class-dependent speech recognition, and thus better recognition performance. However, the amount of data for each class model is reduced by the increase in the number of classes, which leads to unreliable model parameters. To solve the problem of the reduction of training data, we propose a GMM-based soft clustering method that allows overlap, and a selecting method for selecting a speaker model using a GMM or/and HMM. In an experiment, we obtained a 5.0% absolute gain for word error rate (WER), and a 24.9% gain for the relative WER over an age- and gender-dependent baseline.

### Ensemble of SVM Trees for Multimodal Emotion Recognition

Viktor Rozgic *Raytheon BBN Technologies*, Rohit Prasad *Raytheon BBN Technologies*, Shirin Saleem *Raytheon BBN Technologies*, Sankaranarayanan Ananthakrishnan *Raytheon BBN Technologies*, Rohit Kumar *Raytheon BBN Technologies*

In this paper we address the sentence-level multi-modal emotion recognition problem. We formulate the emotion recognition task as a multi-category classification problem and propose an innovative solution based on the automatically generated ensemble of trees with binary support vector machines (SVM) classifiers in the tree nodes. We demonstrate the efficacy of our approach by performing four-way (anger, happiness, sadness, neutral) and five-way (including excitement) emotion recognition on the University of Southern California's Interactive Emotional Motion Capture (USC-IEMOCAP) corpus using combinations of acoustic features, lexical features extracted from automatic speech recognition (ASR) output and visual features extracted from facial markers traced by a motion capture system. The experiments show that the proposed ensemble of trees of binary SVM classifiers outperforms classical multi-way SVM classification with one-vs-one voting scheme and achieves state-of-the-art results for all feature combinations.

### Dereverberantion Based on Generalized Spectral Subtraction for Distant-talking Speaker Recognition

Zhaofeng Zhang *Shizuoka University*, Longbiao Wang *Nagaoka University of Technology*, Atsuhiko Kai *Shizuoka University*

A dereverberation method based on generalized spectral subtraction (GSS) using multi-channel least mean square (MCLMS) was proposed previously. The results on speech recognition experiments showed that this method achieved a significant improvement compare to the conventional methods. In this paper, we employ this method to distant-talking speaker recognition. However, the GSS-based dereverberation method using clean speech models degrades the speaker recognition performance while it is very effective for speech recognition. One of the reason may be that the GSS-based dereverberation method causes some distortions such as speaker characteristics distortion between clean speech and dereverberant speech. In this paper, we address this problem by training speaker models using dereverberant speech which is obtained by suppressing reverberation from arbitrary artificial reverberant speech. The speaker recognition experiment was performed on a large scale far-field speech with different reverberant environments to the training environments. The proposed method achieved a relative error reduction rate of 88.2% compared to conventional CMN with beamforming using clean speech models and 27.8% compared to reverberant speech models, respectively.

## OS.42-IVM.14 Image Enhancement & Restoration

Session Chair: Jian-Jiun Ding Location: Runyon

### Edge-Membership Based Blurred Image Reconstruction Algorithm

Wei-De Chang *National Taiwan University*, Jian-Jiun Ding *National Taiwan University*, Yu Chen *National Taiwan University*, Chir-Weei Chang *Industrial Technology Research Institute*, Chuan-Chung Chang *Industrial Technology Research Institute*

Enhancing the sharpness of edges and avoiding the ringing effect are two important issues in blurred image reconstruction. However, there is a tradeoff between the two goals. A reconstruction filter with long impulse response can reduce the ringing artifact, however, the sharpness of the edge is decreased. By contrast, a short impulse response reconstruction filter can perfectly retrieve the edge but is not robust to noise. In this paper, an edge-membership based blurred image reconstruction algorithm is proposed. In order to achieve the two goals simultaneously, we design two filters. One focuses on edge restoration and the other one focuses on noise removing. After performing linear combination of the outputs of the two reconstruction filters, the edges are preserved and the ringing artifacts are removed at the same time. Simulation results show that our approach can reconstruct the blurred image with sharp edge and less ringing effect.

### Flickers and Black-streak Artifacts Removal on Display Monitors taken by Video Cameras

Kotaro Nakazawa *Shinshu University*, Keiichiro Shirai *Shinshu University*, Masayuki Okamoto *Shinshu University*, Toshio Koga *Yamagata University*

In this paper, we focus on removing black streak artifacts from video images of LED and VFD display monitors taken by video cameras. These streak artifacts affect the image recognition of characters displayed on the monitor, and also affect the quality of video compression. To interpolate correct pixel intensities, we apply an image composition method using light blending of images selected at an appropriate frame interval. The frame interval is decided based on the spectrum of the luminance oscillations at each pixel. Our experimental results show some improvements in the appearance of images and in video encoding.

**Image Restoration with Union of Directional Orthonormal DWTs**

Shogo Muramatsu *Niigata University*, Natsuki Aizawa *Niigata University*, Masahiro Yukawa *Niigata University*

This work proposes to apply directional lapped orthogonal transforms to image restoration. A DirLOT is an orthonormal transform of which basis is allowed to be anisotropic with the symmetric, real-valued and compact-support property. In this work, DirLOTs are used to generate symmetric orthonormal discrete wavelet transforms and then a redundant dictionary as a union of unitary transforms. The multiple directional property is suitable for representing natural images which contain diagonal edges and textures. The performances of deblurring, super-resolution and inpainting are evaluated for several images with the iterative-shrinkage/thresholding algorithm. It is verified that the proposed dictionary yields comparable or superior restoration performance to the non-subsampled Haar transform.

## OS.43-SIPTM.6 Signal and Information Processing Theory and Methods

Session Chair: Jorge Silva Location: Laurel

**Low-Complexity Implementation of the Constrained Recursive Least-Squares Algorithm**

Reza Arablouei *University of South Australia*, Kutluyil Dogancay *University of South Australia*

A low-complexity implementation of the constrained recursive least squares (CRLS) adaptive filtering algorithm is developed based on the method of weighting and the dichotomous coordinate descent (DCD) iterations. The method of weighting is employed to incorporate the linear constraints into the least squares problem of interest. The DCD iterations are then used to solve the normal equations of the resultant unconstrained least squares problem. The new algorithm has a significantly smaller computational complexity than the CRLS algorithm while delivering convergence performance on par with it. Simulations demonstrate the effectiveness of the proposed algorithm.

**A Comparison of Two Algorithmic Recipes to Parametrize Rectangular Orthogonal Matrices**

Simone Fiori *DII - UNIVPM*, Tetsuya Kaneko *SIPLab - TUAT*, Toshihisa Tanaka *Tokyo University of Agriculture and Technology*

The present contribution focuses on the parametrization of rectangular ('tall-skinny') orthogonal matrices, which play a fundamental role in signal processing and machine learning. Such matrices form a smooth curved space termed 'compact Stiefel manifold'. The present contribution aims at illustrating a numerical comparison of two algorithmic recipes to parameterize Stiefel matrices in signal processing.

**Compressibility of Infinite Sequences and its Interplay with Compressed Sensing Recovery**

Jorge Silva *University of Chile*, Eduardo Pavez *University of Chile*

This work elaborates connections between notions of compressibility of infinite sequences, recently addressed by Amini et al., and the performance of the compressed sensing (CS) type of recovery algorithms from linear measurements in the under-sample scenario. In particular, in the asymptotic regime when the signal dimension goes to infinity, we established a new set of compressibility definitions over infinite sequences that guarantees arbitrary good performance in an $\ell_1$-noise to signal ratio ($\ell_1$-NSR) sense with an arbitrary close to zero number of measurements per signal dimension.

## OS.44-SLA.16 Behavioral Informatics: Enabling Technologies and Applications (I)

Session Chair: Shrikanth Narayanan Location: Trousdale Estates

**A Behaviorist Manifesto for the 21st Century**

Brian Baucom *University of Utah*, Esti Iturralde *University of Southern California*

Observational assessment of behavior is a core measurement tool in modern psychological research and practice. Despite the importance of observational assessment and the tremendous amount of research devoted to refining and enhancing the methodological foundations of these tools, current best practices still bear a strong resemblance to those from three decades prior. The emergent field of behavioral signal processing, though relatively little known within the field of psychology, has the potential to revolutionize observational practice and resolve long-standing limitations of current methods. In this paper, we illustrate the need for and potential value of behavioral signal processing methods for observational practice by examining three current issues in the observational assessment of clinically distressed, married couples that are representative of challenges faced in numerous areas of clinical psychology.

**Using Measures of Vocal Entrainment to Inform Outcome-Related Behaviors in Marital Conflicts**

Chi-Chun Lee *University of Southern California*, Athanasios Katsamanis *University of Southern California*, Brian Baucom *University of Southern California*, Panayiotis Georgiou *University of Southern California*, Shrikanth Narayanan *University of Southern California*

Behavioral entrainment is an important, naturally-occurring dynamic phenomenon in human interactions. In this paper, we carry out two quantitative analyses of the vocal entrainment phenomenon in the context of studying conflictual marital interactions. We investigate the role of vocal entrainment in reflecting different dimensions of couple-specific behaviors, such as withdrawal, that are commonly-used in assessing the effectiveness on the outcome of couple therapy. The results indicate a statistically-significant relation between these behaviors and vocal entrainment, as quantified using our proposed unsupervised signal-derived computational framework. We further demonstrate the potential of the signal-based vocal entrainment framework in characterizing influential factors in distressed couples relationship satisfaction outcomes.

**Automatic Detection of Psychological Distress Indicators in Online Forum Posts**

Shirin Saleem *Raytheon BBN Technologies*, Maciej Pacula *Raytheon BBN Technologies*, Rachel Chasin *Massachusetts Institute of Technology*, Rohit Kumar *Raytheon BBN Technologies*, Rohit Prasad *Raytheon BBN Technologies*, Michael Crystal *Raytheon BBN Technologies*,Brian Marx *National Center for PTSD at VA Boston Healthcare System*, Denise Sloan *National Center for PTSD at VA Boston Healthcare System*, Jennifer Vasterling *National Center for PTSD at VA Boston Healthcare System*, Theodore Speroff *Vanderbilt University*

The stigma associated with mental health issues makes face-to-face discussions with family members, friends, or medical professionals difficult for many people. In contrast, the Internet, due to its ubiquity and global outreach, is increasingly becoming a popular medium for distressed individuals to anonymously relate experiences. In this paper, we present a system for automatically detecting psychological distress indicators in informal text interactions on Internet discussion forums. We compare a suite of innovative features and classifiers on data downloaded from an online forum discussing psychological health issues. Psychologists annotated individual messages with a comprehensive set of distress labels derived from the Diagnostic and Statistical Manual of Mental Disorders (DSM) IV. The noisy nature of the forum posts and the large set of distress labels for multi-label text classification (many of which cannot be detected by a mere surface form analysis of the text), make the task extremely challenging. A late fusion technique combines outputs from different classifiers resulting in promising accuracy on this challenging multi-label classification problem.

## OS.45-SLA.17 Recent Topics in Speech Processing (II)

Session Chair: Longbiao Wang Location: Franklin Hills

### A Study of Emotional Information Present in Articulatory Movements Estimated Using Acoustic-To-Articulatory Inversion

Jangwon Kim *University of Southern California*, Prasanta Ghosh *IBM Research India*, Sungbok Lee *University of Southern California*, Shrikanth Narayanan *University of Southern California*

This study examines emotion-specific information (ESI) in the articulatory movements estimated using acoustic-to-articulatory inversion on emotional speech. We study two main aspects: (1) the degree of similarity between the pair of estimated and original articulatory trajectories for the same and different emotions and (2) the amount of ESI present in the estimated trajectory. They are evaluated using mean squared error between the articulatory pair and by automated emotion classification. This study uses parallel acoustic and articulatory data in 5 elicited emotions spoken by 3 native American English speakers. We also test emotion classification performance using articulatory trajectories estimated from different acoustic feature sets and they turn out subject-dependent. Experimental results suggest that the ESI in the estimated trajectory, although smaller than that in the direct articulatory measurements, is found to be complementary to that in the prosodic features and hence, suggesting the usefulness of estimated articulatory data for emotions research.

### Robust Feature Extraction to Utterance Fluctuations Due to Articulation Disorders Based on Sparse Expression

Toshiya Yoshioka *Kobe University*, Ryoichi Takashima *Kobe University*, Tetsuya Takiguchi *Kobe University*, Yasuo Ariki *Kobe University*

We investigated the speech recognition of a person with articulation disorders resulting from athetoid cerebral palsy. Recently, the accuracy of speaker-independent speech recognition has been remarkably improved by the use of stochastic modeling of speech. However, the use of those acoustic models causes degradation of speech recognition for a person with different speech styles (e.g., articulation disorders). In this paper, we discuss our efforts to build an acoustic model for a person with articulation disorders. The articulation of the first utterance tends to become more unstable than other utterances due to strain on speech-related muscles, and that causes degradation of speech recognition. Therefore, we propose a robust feature extraction method based on exemplar-based sparse representation using NMF (Non-negative Matrix Factorization). In our method, the unstable first utterance is expressed as a linear and non-negative combination of a small number of bases created using the more stable utterances of a person with articulation disorders. Then, we use the coefficient of combination as an acoustic feature. Its effectiveness has been confirmed by word-recognition experiments.

### Distant-talking Speaker Identification Using a Reverberation Model With Various Artificial Room Impulse Responses

Longbiao Wang *Nagaoka University of Technology*, Zhaofeng Zhang *Shizuoka University*, Atsuhiko Kai *Shizuoka University*, Yoshiki Kishi *NS Solutions Kansai Corp.*

In this paper, we propose a distant-talking speaker recognition method using a reverberation model with various artificial room impulse responses. These artificial room impulse responses with different speaker and microphone positions, room sizes, and reflection coefficients of walls and convoluted with clean speech are used to train an artificial reverberation speaker model. This artificial reverberation model is also combined with a reverberation speaker model trained with room impulse responses measured in real environments. Speaker identification performance using a combination of the two reverberation speaker models achieved a relative error reduction rate of 50.0% and 78.4% compared with that using a reverberation model trained with real-world room impulse responses and a clean speech model, respectively.

## OS.46-IVM.15 Intelligent Object Detection and Identification for Visual Surveillance and Security

Session Chair: Chia-Hung Yeh Location: Whitley Heights

### Background Subtraction by Modeling Pixel and Neighborhood Information

Shu-Jhen Fan Jiang *National Sun Yat-sen University*, Kahlil Muchtar *Asia University*, Chih-Yang Lin *Asia University*, Li-Wei Kang *Academia Sinica*, Chia-Hung Yeh *National Sun Yat-sen University*

In applications of the computer vision field, a vision system is usually composed of several low level and high level components, stacked on top of each other. A better design of the lower level components usually results in better accuracy of higher level functions, such as object tracking, face recognition, and surveillance. In this paper, we focus on the low level component design, background construction, which is one of the most basic elements for a surveillance system. The proposed method eases the problems that usually occur in background construction, including aperture problem, vacillating background, and shadow removal. In conventional background construction methods, only the history information (vertical direction) of pixels is usually considered. In contrast, the proposed scheme not only uses the vertical direction but also the neighborhood information (horizontal direction). Experimental results show that the proposed scheme can detect objects more delicate, alleviate the aperture problem, and identify shadow and discard it from detected objects.

### A Block Based (t, n) Visual Cryptography Scheme for Unbounded n and t=2, 3

Sian-Jheng Lin *Academia Sinica*, Wei-Ho Chung *Academia Sinica*

The (t, n) visual cryptography (VC) is a secret sharing scheme of decomposing a secret image into n transparencies, and the stacking of any t out of n transparencies reveals the secret content. The perfect security condition of VC scheme requires the strict requirement where any t_1 or fewer transparencies cannot extract any information about the secret. For n approaching infinity, previous studies consider the scenario where the probabilistic model is a pixel-to-pixel scheme that encodes each secret pixel to a corresponding pixel in each transparency. In this paper, we extend the pixel-to-pixel scheme to pixel-to-block scheme for the cases t=2 and 3. Given a secret image, the proposed VC scheme generates a transparency through coding each secret image pixel to a m-pixels shadow block on the corresponding position of the transparency. Experiments show that the stacking results reveal better visual quality than the probabilistic model scheme.

### A Real-Time Rear Obstacle Detection System Based On a Fish-Eye Camera

Che-Tsung Lin *ITRI*, Yu-Chen Lin *ITRI*, Wei-Cheng Liu *ITRI*, Chi-Wei Lin *ITRI*

This paper proposes a rear vision camera-based vehicle detection system which could detect if any rear vehicle exists in ego lane and if any vehicles in adjacent lanes are overtaking. The source image is firstly applied with distortion calibration which helps the following Hough transform to detect the existence of lane lines. The rear vehicle in ego lane is detected by a combination of feature-based approach and appearance-based approach. When a vehicle in adjacent lane is overtaking, the vanishing of its symmetry makes itself very difficult to be detected. Therefore, we propose a new detection algorithm applying corner detection and motion vector whose calculation are based on Local Binary Pattern (LBP) to find if any vehicles in adjacent lanes are overtaking. Our proposed algorithm achieves high detecting rate and low computing power and is successfully implemented in ADI-BF561 600MHz dual core DSP.

## OS.47-IVM.16 Image Analysis and Recognition

Session Chair: Toshihiko Yamasaki Location: Mt. Olympus

### Spatial Statistics for Spatial Pyramid Matching Based Image Recognition

Toshihiko Yamasaki *University of Tokyo*, Tsuhan Chen *Cornell University*

This paper presents an image feature extraction algorithm that enhances the object classification accuracy in the spatial pyramid matching (SPM) framework. The proposed method considers the spatial statistics of the feature vectors by calculating the moment vectors. While the original SPM algorithm captures the spatial distribution of the image feature descriptors, the proposed algorithm describes how such spatial distribution is variant. The experiments are conducted using two state-of-the-art SPM-based methods for two commonly used datasets. The results demonstrates the validity of our proposed algorithm. The cases where the proposed algorithm works well are also investigated. In addition, it is demonstrated that the proposed feature and adding more layers improve the classification accuracy in different situations.

### Using the Visual Words based on Affine-SIFT Descriptors for Face Recognition

Yu-Shan Wu *Chunghwa Telecommunication Laboratories*, Heng-Sung Liu *Chunghwa Telecommunication Laboratories*, Gwo-Hwa Ju *Chunghwa Telecommunication Laboratories*, Ting Wei Lee *Chunghwa Telecommunication Laboratories*, Yen-Lin Chiu *Chunghwa Telecommunication Laboratories*

Video-based face recognition has drawn a lot of attention in recent years. On the other hand, Bag-of-visual Words (BoWs) representation has been successfully applied in image retrieval and object recognition recently. In this paper, a video-based face recognition approach which uses visual words is proposed. In classic visual words, Scale Invariant Feature Transform (SIFT) descriptors of an image are firstly extracted on interest points detected by difference of Gaussian (DoG), then k-means-based visual vocabulary generation is applied to replace these descriptors with the indexes of the closet visual words. However, in facial images, SIFT descriptors are not good enough due to facial pose distortion, facial expression and lighting condition variation. In this paper, we use Affine-SIFT (ASIFT) descriptors as facial image representation. Experimental results on UCSD/Honda Video Database and VidTIMIT Video Database suggest that visual words based on Affine-SIFT descriptors can achieve lower error rates in face recognition task.

### An Open Framework for Video Content Analysis

Chia-Wei Liao *Institute for Information Industry*, Kai-Hsuan Chan *Institute for Information Industry*, Bin-Yi Cheng *Institute for Information Industry*, Chi-Hung Tsai *Institute for Information Industry*, Wen-Tsung Chang *Institute for Information Industry*, Yu-Ling Chuang *Institute for Information Industry*

In the past few years, the amount of the internet video has grown rapidly, and it has become a major market. Efficient video indexing and retrieval, therefore, is now an important research and system-design issue. Reliable extraction of metadata from video as indexes is one major step toward efficient video management. There are numerous video types, and theoretically, everybody can define his/her own video types. The nature of video can be so different that we may end up, for each video type, having a dedicated video analysis module, which is in itself nontrivial to implement. We believe an open video analysis framework should help when one needs to process various types of videos. In the paper, we propose an open video analysis framework where the video analysis modules are developed and deployed as plug-ins. In addition to plug-in management, it provides a runtime environment with standard libraries and proprietary rule-based automaton modules to facilitate the plug-in development. A prototype has been implemented and proved with some experimental plug-ins.

## PS.5-SLA.18 Speech Coding and Processing and Recognition

Session Chair: Hideki Kawahara Location: Solano

### Speaking Rate Dependent Multiple Acoustic Models Using Continuous Frame Rate Normalization

Ban Sung Min *Pusan National University*, Kim Hyung Soon *Pusan National University*

This paper proposes a method using speaking rate dependent multiple acoustic models for speech recognition. In this method, multiple acoustic models with various speaking rates are generated. Among them, the optimal acoustic model relevant to the speaking rate of test data is selected and used in recognition. To simulate the various speaking rates for the multiple acoustic models, we use the variable frame shift size considering the speaking rate of each utterance instead of applying a flat frame shift size to all training utterances. The Continuous Frame Rate Normalization (CFRN) is applied to each of training utterances to control the frame shift size. Experimental results show that the proposed method outperforms both the baseline and the conventional CFRN on test utterances.

### Emotion Classification of Infant Cries with Consideration for Local and Global Features

Kazuki Honda *Nagasaki University*, Kazuki Kitahara *Nagasaki University*, Shoichi Matsunaga *Nagasaki University*, Masaru Yamashita *Nagasaki University*

In this paper, we propose an approach to the classification of emotion clusters in infant cries with consideration for frame-wise/local acoustic features and global prosodic features. Our proposed approach has two main characteristics as follows. The emotion cluster detection procedure is based on the most likely segment sequence, which delivers the emotion cluster as a classification result. This is obtained based on a maximum likelihood approach using the frame-wise likelihood and the global prosodic likelihood. We exploit the duration ratios of resonant cry segments and silent segments as prosodic features, while the duration ratios are calculated using the derived segment sequence. The second characteristic is the use of pitch information, in addition to conventional power and spectral information, during the modeling of frame-wise acoustic features with hidden Markov models. The classification performance (74.7%) of our proposed approach with added pitch information was better than (71.5%) the classification method using only power and spectral features. The proposed method based on a maximum likelihood approach using both frame-wise and global features also achieved better performance (75.5%).

### On the Use of Phase Information-based Joint Factor Analysis for Speaker Verification under Channel Mismatch Condition

Ikuya Hirano *Shizuoka University*, Longbiao Wang *Nagaoka University of Technology*, Atsuhiko Kai *Shizuoka University*, Seiichi Nakagawa *Toyohashi University of Technology*

Recent studies have shown that phase information contains speaker characteristics. A new extraction method to extract pitch synchronous phase information has been proposed and shown that it was very effective under channel matched condition. However, phase changes between different channels. Therefore, the speaker recognition performance is drastically degraded under channel mismatch condition. On the other hand, Joint Factor Analysis (JFA) is an approach that is robust for channel variability. In this paper, we propose phase information-based JFA for speaker verification under channel mismatch condition. Speaker verification experiments were performed using the NIST 2003 SRE database. Phase information-based JFA achieved a relative equal error rate reduction of 20.9% for male and 17.4% for female compared to the traditional system based on Gaussian Mixture Model and Universal Background Model (GMM-UBM) that influenced by channel variability. Furthermore, by combining phase information-based method with the MFCC-based method, we obtained the better result than that of the only MFCC-based method.

### Modulation Transfer Function Design for a Flexible Cross Synthesis Vocoder Based on $F_O$ Adaptive Spectral Envelope Recovery

Taiki Nishi *Wakayama University*, Ryuichi Nisimura *Wakayama University*, Toshio Irino *Wakayama University*, Hideki Kawahara *Wakayama University*

A new design procedure for flexible cross synthesis VOCODER is proposed based on TANDEM-STRAIGHT framework, a $F_O$ adaptive spectral envelope estimator, and modulation transfer function design. The proposed design procedure enables control of speech intelligibility and timber identity of musical instruments or animal voices. Removal of the averaged and smoothed logarithmic spectrum of speech from the filter reduced the timbre modification effect of filtered sounds and manipulation of cut-off frequencies of modulation transfer function for designing the filter enabled control of trade-offs between intelligibility and timbre preservation.

### Detecting Child Speaker Based on Auditory Feature Vectors for VTL Estimation

Ryuichi Nisimura *Wakayama University*, Shoko Miyamori *Wakayama University*, Erika Okamoto *Wakayama University*, Hideki Kawahara *Wakayama University*, Toshio Irino *Wakayama University*

We introduce novel auditory features in the Hidden Markov Model (HMM) system for detecting child speakers. The features derived by the gammachirp auditory filterbank (GCFB) have been demonstrated to be suitable for vocal tract length (VTL) estimation, both theoretically and experimentally. We performed numerical experiments to distinguish between child and adult speakers using HMMs trained on 2,360 speech samples collected through a web-based query interface, and we compared the performance of the common mel-frequency cepstral coefficients (MFCC) and the GCFB-based feature vectors. We also introduced the modulation features as the substitution of delta parameters. It has been clearly demonstrated that the error rate distinguishing a child from an adult is reduced by GCFB. To enhance our method for use as a web application, we applied our original voice-enabled web framework to the front-end interface of the proposed system.

### Acoustic Model Training Using Feature Vectors Generated by Manipulating Speech Parameters of Real Speakers

Tetsuto Kawai *Nagoya University,* Norihide Kitaoka *Nagoya University*, Kazuya Takeda *Nagoya University*

In this paper, we propose a robust speaker-independent acoustic model training method using generative training to generate many pseudo-speakers from a small number of real speakers. We focus on the difference between each speaker's vocal tract length, and manipulate it in order to create many different pseudo-speakers with a range of vocal tract lengths. This method employs frequency warping based on the inverted use Vocal Tract Length Normalization (VTLN). Another method for creating pseudo-speakers is to vary the speaking rate of the speakers. This can be achieved by a method called PICOLA (Pointer Interval Controlled OverLap and Add). In experiments, we train acoustic models using these generated pseudo-speakers in addition to the original speakers. Evaluation results show that generating pseudo-speakers by manipulating speaking rates did not result in a sufficient increase in performance, however, vocal tract length warping was effective.

### A Concatenative Speech Synthesis for Monosyllabic Languages With Limited Data

Trung-Nghia Phung *Japan Advanced Institute of Science and Technology*, Luong Chi Mai *Vietnamese Academy of Science and Technology*, Masato Akagi *Japan Advanced Institute of Science and Technology*

Quality of unit-based concatenative speech synthesis is low while that of corpus-based concatenative speech synthesis with unit selection is great natural. However, unit selection requires a huge data for concatenation that reduces the range of its applications. In this paper, by using temporal decomposition for modeling contextual effects intra-syllable and inter-syllables, we propose a context-fitting unit modification method and a context-matching unit selection method. The two proposed context-specific methods are used in our proposed syllablebased concatenative speech synthesis applied for monosyllabic languages. The experimental results with a Vietnamese speech synthesis using a small corpus support that the proposed methods are efficient. As a consequence, the naturalness and intelligibility of the proposed speech synthesis is high even when we have only limited data for concatenation.

### Feature Reconstruction using Sparse Imputation for Noise Robust Audio-Visual Speech Recognition

Peng Shen *Gifu University*, Tamura Satoshi *Gifu University*, Hayamizu Satoru *Gifu University*

In this paper, we propose to use noise reduction technology on both speech signal and visual signal by using exemplar-based sparse representation features for audio-visual speech recognition. First, we introduce sparse representation classification technology and describe how to utilize the sparse imputation to reduce noise not only for audio signal but also for visual signal. We utilize a normalization method to improve the accuracy of the sparse representation classification, and propose a method to reduce the error rate of visual signal when using the normalization method. We show the effectiveness of our proposed noise reduction method and that the audio features achieved up to 88.63% accuracy at -5dB, a 6.24% absolute improvement is achieved over the additive noise reduction method, and the visual features achieved 27.24% absolute improvement at gamma noise.

### Statistical Voice Conversion using GA-based Informative Feature

Kohei Sawada *Gifu University*, Yoji Tagami *Gifu University*, Tamura Satoshi *Gifu University*, Masanori Takehara *Gifu University*, Hayamizu Satoru *Gifu University*

In order to make voice conversion (VC) robust to noise, we propose VC using GA-based informative feature (GIF), by adding an extraction process of GIF to a conventional VC. GIF is proposed as a feature that can be applied not only in pattern recognition but also in relative tasks. In speech recognition, furthermore, GIF could improve recognition accuracy in noise environment. We evaluated the performances of VC using spectral segmental features (conventional method) and GIF, respectively. Objective experimental result indicates that in noise environments, the proposed method was better than the conventional method. Subjective experiment was also conducted to compare the performances. These results show that application of GIF to VC was effective.

### GIF-SP: GA-based Informative Feature for Noisy Speech Recognition

Tamura Satoshi *Gifu University*, Yoji Tagami *Gifu University*, Hayamizu Satoru *Gifu University*

This paper proposes a novel discriminative feature extraction method. The method consists of two stages; in the first stage, a classifier is built for each class, which categorizes an input vector into a certain class or not. From all the parameters of the classifiers, a first transformation can be formed. In the second stage, another transformation that generates a feature vector is subsequently obtained to reduce the dimension and enhance recognition ability. These transformations are computed applying genetic algorithm. In order to evaluate the performance of the proposed feature, speech recognition experiments were conducted. Results in clean training condition shows that GIF greatly improves recognition accuracy compared to conventional MFCC in noisy environments. Multi-condition results also clarifies that out proposed scheme is robust against differences of conditions.

**A Packet Loss Recovery of G.729 Speech Under Severe Packet Loss Condition**
Takeshi Nagano *Tohoku University*, Akinori Ito *Tohoku University*

In a VoIP application, packet losses degrade speech quality. Especially, IP network under a large-scale disaster should cause severe packet losses. We investigate influence of parameter loss to speech quality using G.729. And we investigated an effect of packet loss concealment method using redundant G.729 parameters. As compared with ``repetition'' method, the proposed method could improve speech quality. We also propose a bitrate reduction method by sending bit flip position instead of codebook index.

**Microphone Array Processing for Distant Speech Recognition: Spherical Arrays**
John McDonough *Carnegie Mellon University/Voci Technologies, Inc.*, Kenichi Kumatani *Disney Research, Pittsburgh*, Bhiksha Raj *Carnegie Mellon Univerity*

Distant speech recognition (DSR) holds out the promise of the most natural human computer interface because it enables man-machine interactions through speech, without the necessity of donning intrusive body- or head-mounted microphones. With the advent of the Microsoft Kinect, the application of non-uniform linear arrays to the DSR problem has become commonplace. Performance analysis of such arrays is well-represented in the literature. Recently, spherical arrays have become the subject of intense research interest in the acoustic array processing community. Such arrays have heretofore been analyzed solely with theoretical metrics under idealized conditions. In this work, we analyze such arrays under realistic conditions. Moreover, we compare a linear array with 64-channel arrays and a total length of 128cm to a spherical array with 32 channels and a radius of 4.2cm; we found that these provided word error rates of 9.3% and 9.9%, respectively, on a DSR task. For a speaker positioned at an oblique angle with respect to the linear array, we recorded error rates of 12.8% and 10.7%, respectively, for the linear and spherical arrays. The compact size and outstanding performance of the spherical array recommends itself well to space-limited and mobile applications such as home-gaming consoles and humanoid robots.

## 10:50 - 12:30

### OS.48-SLA.19 Fundamental Technologies in Modern Speech Processing (II)

Session Chair: Chia-Ping Chen Location: Doheny

**Sub-word Modeling for Automatic Speech Recognition**
Karen Livescu *Toyota Technological Institute at Chicago*, Eric Foster Lussier *Ohio State University*, Florian Metze *Carnegie Mellon University*

**Discriminative Training for ASR: Modeling, Criteria, Optimization, Implementation, and Performance**
Georg Heigold *Google*, Hermann Ney *RWTH Aachen University*, Ralph Schlueter *RWTH Aachen University*, Simon Wiesler *RWTH Aachen University*

**Deep Neural Networks for Acoustic Modeling in Speech Recognition**
Geoffrey Hinton *University of Toronto*, Li Deng *Microsoft Research*, Dong Yu *Microsoft Research*, George Dahl *University of Toronto*, Abdel-rahman Mohamed *University of Toronto*, Navdeep Jaitly *University of Toronto*, Vincent Vanhoucke *Google*, Patrick Nguyen *Google*, Tara N. Sainath *IBM T. J. Watson Research Center*, Brian Kingsbury *IBM T. J. Watson Research Center*

**Exemplar-Based Processing for Speech Recognition**
Tara N. Sainath *IBM T. J. Watson Research Center*, Bhuvana Ramabhadran *IBM T. J. Watson Research Center*, David Nahamoo *IBM T. J. Watson Research Center*, Dimitri Kanevsky *IBM T. J. Watson Research Center*, Dirk Van Compernolle *KU Leuven*, Kris Demuynck *KU Leuven*, Jort F. Gemmeke *KU Leuven*, Jerome R. Bellegarda *Apple*, Shiva Sundaram *Deutsche Telekom Laboratories*

### OS.49-IVM.17 Recent Advances in High Fidelity Digital Imaging

Session Chairs: Masahiro Okuda, Yuichi Tanaka Location: Beachwood

**Directional Image Decomposition Using Retargeting Pyramid**
Yuichi Tanaka *Tokyo University of Agriculture and Technology*, Keiichiro Shirai *Shinshu University*

Retargeting pyramid (RP) is an alternative method for multiscale image decomposition to the well-known Laplacian pyramid (LP). RP can be obtained by replacing the low-pass filtering process in LP with content-aware image resizing (a.k.a. retargeting), which is a developing technique for computer vision researches. Furthermore, we use RP for contourlet-based directional image decomposition. In experimental results, the proposed decomposition outperforms the LP-based contourlet transform for image denoising.

**A Hierarchical Convex Optimization Approach for High Fidelity Solution Selection in Image Recovery**
Shunsuke Ono *Tokyo Institute of Technology*, Isao Yamada *Tokyo Institute of Technology*

The aim of this paper is to propose a hierarchical convex optimization for selecting a high fidelity image from possible solutions of a convex optimization problem associated with existing image recovery methods. Image recovery problems have been cast in certain convex optimization problems which have infinitely many solutions in general. However, existing convex optimization algorithms are designed to reach one solution randomly, and hence can not select a solution corresponding to a high fidelity image from the possible solutions. In this paper, we propose to select a high fidelity image by solving a newly-formulated hierarchical convex optimization problem. This problem is a constrained minimization of a convex criteria over the solution set of all images which are optimal in the sense of any existing image recovery method. The hierarchical convex optimization problem is efficiently solved by a proposed iterative scheme based on the hybrid steepest descent method with the help of a nonexpansive mapping related to the Douglas-Rachford splitting type algorithms. Numerical results indicate that our method appropriately selects a recovered image of high fidelity in the case of inpainting and compressed sensing recovery.

### Head Pose Estimation using Motion Subspace Matching on GPU

Nawarat Auttanugune *Chulalongkorn University*, Thanarat Chalidabhongse *Chulalongkorn University*, Supavadee Aramvith *Chulalongkorn University*

The head pose estimation is a process of recovering 3D head position in term of yaw, pitch and roll from 2D images. However, the reduction of information from 3D to 2D leads to an ill-posed problem. In this paper, we propose a novel algorithm of head pose estimation that includes facial features tracking for Thai sign language recognition. In order to estimate head pose correctly, feature points tracking requires high precision. Nevertheless it is difficult for low cost cameras where input image quality may be generally poor. To overcome this problem, we introduce an automatic camera signal calibration such that the features can be tracked correctly despite the quality of the input image sequences. Finally, as our approach bases on the state space searching, the local minima problem is common. Hence, we divide the search space into sub spaces and perform parallel computation on GPU.

### High Dynamic Range Image Compression using Base Map Coding

Takuya Fujiki *University of Kitakyushu*, Nicola Adami *Brescia University*, Takao Jinno *Toyohashi University Of Technology*, Masahiro Okuda *University of Kitakyushu*

As the High dynamic range (HDR) images generally have more than 10 bit/1024 colors per channel, its enormous data size often needs to be reduced when transmitting or storing the images. Thus development for functional compression is one of important research topics. Recently a lot of techniques for the HDR image compression are being suggested, and several two-layer coding algorithms which separately encode a low dynamic range (LDR) image and a residual image have been studied. However, those methods are inefficient in coding performance. In this article, we suggest a new two-layer coding algorithm for the HDR images, which realizes two-layered dynamic range. Our method encodes a base map, which is a blurred version of the HDR, and LDR image produced by the base map. Our algorithm significantly improves a compression performance.

### Efficient Lossless Bit Depth Scalable Coding for HDR Images

Masahiro Iwahashi *Nagaoka University of Technology*, Hitoshi Kiya *Tokyo Metropolitan University*

This report proposes two layered bit depth scalable coding methods for high dynamic range (HDR) images expressed in floating point data format. From the base layer bit stream, low dynamic range (LDR) images are decoded. They are tone mapped appropriately for human eye sensitivity, and shortened to a standard bit depth, e.g. 8 bit. From the enhance layer bit stream, HDR images are decoded. However the bit depth of this layer has been huge in the existing method. To reduce it, we divide the tone mapping into a reversible logarithmic mapping and its compensation. It was confirmed that the proposed methods significantly reduce the bit depth of the enhance layer, even though the compensation slightly increases coding noise.

## OS.50-WCN.5 Video/3D Data Transmission over Mobile Network

Session Chairs: Hwangjun Song, Jongwon Kim Location: Runyon

### Real-Time Acquisition and Representation of 3D Environmental Data

Se-Ho Lee *Korea University*, Seong-Gyun Jeong *Korea University*, Tae-Young Chung *Korea University*, Chang-Su Kim *Korea University*

We develop a mobile system to acquire and represent 3D environmental data for modeling indoor spaces. The system is composed of a laser range finder (LRF) and an omni-directional camera. Multiple 3D point clouds from different viewpoints are acquired as the geometric information by scanning a scene with the LRF, while an omni-directional texture image is acquired with the omni-directional camera. We merge those multiple 3D point clouds into a single point cloud. We then combine the point cloud and the texture image into a complete 3D mesh model in three steps. First, we downsample the point cloud based on a voxel grid and estimate the normal vector of each point. Second, using the normal vectors, we reconstruct a 3D mesh based on the Poisson surface reconstruction. Third, to assign texture information to the mesh surface, we estimate the matching region in the omnidirectional image that corresponds to each face of the mesh. Simulation results demonstrate that the proposed system can reconstruct indoor spaces effectively.

### Reliable Scalable Video Multi-cast with Source Diversity and Inter-source Network Decoding in Lossy Networks

Saran Tarnoi *National Institute of Informatics*, Yusheng Ji *National Institute of Informatics*, Wuttipong Kumwilaisak *King Mongkut´s University of Technology Thonburi*

This paper presents a reliable scalable video multi-cast with source diversity and inter-source network decoding in lossy networks. The source diversity technique gives path diversity providing a better quality of layered video transmission under hostile environments. For each source, an optimization formulation is set up to find the best transmission route of each transmitting video layer. The objectives of the formulation are to maximize a number of transmitting video layers and transmission reliability. The source providing the best overall throughput is selected to be the primary source, while the rest will be secondary sources. When the Quality-of-Service (QoS) guarantee of some transmitting video layers cannot be fulfilled by the primary source, the secondary source with the best QoS parameters is selected to transmit the layers to destinations. The number of secondary sources used for transmissions is increased until the QoS guarantees of all transmitting video layers are satisfied or all network resources are utilized. Network coding is deployed to multi-cast video layers from the same source for efficient resource usage. Network coded data from different sources can be used together to decode the transmitting video data. In other words, at each destination, it needs only a sufficient number of video packets from different sources to recover all transmitting video data. Simulations with different network topologies show the improvement in both objective and subjective qualities of layered video multi-cast under lossy environments.

### Advanced Logical Superposition Modulation based Video Streaming Multicast System over Wireless Networks

Kyuhwi Choi *POSTECH*, Hwangjun Song *POSTECH*

In this paper, we present an advanced logical superposition modulation based video streaming multicast system over a wireless networks. The traditional logical superposition modulation scheme generates symbols using a pre-fixed modulation scheme to overcome wireless channel diversity. In contrast, our proposed logical superposition modulation determines logical superposition modulation scheme according to time-varying wireless channel conditions in order to maximize overall network throughput, similar to state-of-art adaptive modulation schemes. In addition, two-layer SVC video streams are effectively mapped to the determined superposition modulation scheme to support better quality video streaming. Experimental results show the performance of our proposed system.

### Constant Frame Quality Control for H.264/AVC

Po-Chyi Su *National Central University*, Ching-Yu Wu *National Central University*, Long-Wang Huang *National Central University*, Chia-Yang Chiou *National Central University*

A frame quality control mechanism for H.264/AVC is proposed in this research. The research objective is to ensure that a suitable Quantization Parameter (QP) can be assigned to each frame so that the target frame quality can be achieved. One application is consistently maintaining the frame quality during the encoding process to facilitate such applications of video archiving or surveillance. A single-parameter Distortion to Quantization (D-Q) model is derived by training a large number of frame blocks and this model parameter can be determined by the frame content. Given the target quality for a video frame, we can then select an appropriate QP according to the proposed D-Q model. The model refinement and QP adjustment of subsequent frames can be applied according to the coding results of previous data. Structural Similarity (SSIM) is chosen as the quality measurement to demonstrate the feasibility of the proposed framework.

### Peer-assisted Video-on-Demand in Multi-channel Switching WiFi-based Mobile Networks

Jongwon Kim *GIST*, Hayoung Yoon *GIST*

Network convergence paradigm will substantially increase the pervasive use of WiFi-enabled smart mobile devices. Although various on-demand streaming services are already available over mobile WiFi-enabled devices, it remains a challenging problem due to WiFi's limited communication range, mobility and user population density issues. In this paper, we enhance our previous work on MOVi (Mobile Opportunistic Video-on-demand) by exploiting the use of multi-channel switching capability at mobile terminals. We reinvestigate the previous version of MOVi in this environment and propose an improved scheduling algorithm which incorporates collaborative congestion sensing and efficient channel allocation mechanisms. The scalability of the extended MOVi system is verified by extensive simulations. In terms of the number of supported user, an average of 25% improvement can be achieved. In addition, the proposed extension of MOVi provides more tolerance against increased volume of non-MOVi traffic.

## OS.51-WCN.6 Green Wireless Communicaions

Session Chair: Sumei Sun Location: Laurel

### An Improved LLR Approximation Algorithm for Low-Complexity MIMO Detection Towards Green Communications

Ruijuan Ma *Xi'an Jiaotong University*, Pinyi Ren *Xi'an Jiaotong University*, Chao Zhang *Xi'an Jiaotong University*, Qinghe Du *Xi'an Jiaotong University*

As the number of transmit/receive antennas gets large in wireless communication systems, the drastically-increasing complexity in MIMO detection imposes significant challenges in implementing green communications while achieving high spectral efficiency. The winner-path-extension (WPE) K-best algorithm is an efficient detection algorithm in uncoded MIMO systems, known for its stable throughput and excellent symbol-error-rate (SER) and bit-error-rate (BER) performances under relatively low complexity. However, when applying the WPE K-best algorithm into coded MIMO systems, where soft-output information such as log-likelihood ratio (LLR) is required, missing counter-hypotheses issue in LLR calculation often degrades the error performance. To solve this problem, in this paper we propose an improved LLR approximation algorithm, such that WPE K-best algorithm can be well suited to coded MIMO systems. Specifically, when a counter-hypothesis misses, we set a metric threshold for the missing counter-hypothesis by calculating the metric of the bit flipping vector, and then randomly choose a value below the threshold as the approximation. We conduct simulation evaluations for our proposed algorithm in an 8×8 MIMO multiplexing system employing 16QAM modulation and Turbo coding. Simulation results show that compared with other existing LLR approximation schemes, our proposed approach can effectively improve the block-error-rate (BLER) performance as well as reducing the complexity in the tree search of WPE K-best algorithm. Moreover, we use a look-up table method to determine the Schnorr-Euchner (SE) enumeration order, which can further decrease the computational complexity of WPE K-best algorithms.

### Green Wireless Communications: A Power Amplifier Perspective

Jingon Joung *Institute for Infocomm Research*, Chin Keong Ho *Institute for Infocomm Research*, Sumei Sun *Institute for Infocomm Research*

In this paper, we survey two essential and practical characteristics of radio-frequency power amplifier (PA), namely, linearity and efficiency. Nonlinear amplification yields significant distortion of the transmit signals and strong interference for cochannel users. Imperfect efficiency of the PA causes an overhead of the systems resulting in energy efficiency (EE) degradation. Therefore, the linearity and efficiency of the PA should be precisely characterized in the system design. We first survey the linearity and efficiency models of PA, and then introduce commonly used technologies for improving the EE according to three approaches: i) transmitter architecture, ii) signal processing, and iii) network protocols. We then introduce our recent work on a multiple PA switching (PAS) architecture, in which one or more PAs are switched on at any time to maximize the EE while satisfying the required spectral efficiency (SE). We consider the case where either full or partial channel state information is available at the transmitter (CSIT). Since the transmitter selects the most efficient PA that satisfies a target rate with the least power consumption, EE is improved, and a Pareto-optimal SE-EE tradeoff region can be enlarged as verified in the numerical results with real-life device parameters. For example, we observe around 323% and 50% EE improvements for a single antenna system with a full CSIT and for a transmit antenna selection and maximum ration combining system with a partial CSIT, respectively; as a result, we can surmise that the PAS is one promising technology for green, i.e., energy efficient, wireless communication systems.

### Area Spectral Efficiency of Cooperative Network with DF and AF Relaying

Lei Zhang *University of Victoria*, Hong-Chuan Yang *University of Victoria*, Mazen Hasna *Qatar University*

Most performance metrics for cooperative networks focus on the qualification of either spectrum efficiency or link reliability, without considering the spatial effect of radio transmission. Area spectral efficiency (ASE) was first introduced to qualify the spatial spectral utilization efficiency of cellular systems. In this paper, we generalize the definition of ASE and investigate this performance metric in a three-node cooperative network with decode-and-forward (DF) and amplify-and-forward (AF) relaying. We derive the mathematical expression of ASE with the consideration of path-loss and fading effects. We show through selected numerical examples that ASE provides a new perspective on the spectrum utilization efficiency and transmission power design.

### Validation of a Green Wireless Communication System with ICA based Semi-Blind Equalization

Teng Ma *University of Liverpool*, Xu Zhu *University of Liverpool*, Yufei Jiang *University of Liverpool*, Yi Huang *University of Liverpool*

In this paper, we validate a green wireless communication system with the independent component analysis (ICA) based and precoding aided semi-blind equalization, on a testbed which consists of a pair of Keithley signal generator and signal analyzer connected to antennas. The implemented system requires only a small amount of side information to be transmitted to the receiver and therefore achieves energy

and spectrum-efficient green communications. The system performance is measured in different wireless channels and compared with simulation results. The impact of precoding weight constant on the BER performance is showed and optimal constant value is found. The impact of frame length on the performance of ICA based equalization is also evaluated.

**Energy Efficient Cooperation for Two-Hop Relay Networks**

Ernest Kurniawan *Stanford University*, Stefano Rini *Technische Universitat Munchen*, Andrea Goldsmith *Stanford University*

We analyze the impact of cooperation on the energy efficiency of two-hop relay transmissions. While cooperation has been demonstrated to improve spectral efficiency, the benefits in terms of energy consumption are not well characterized. We show that cooperation is not always beneficial in this context, since the energy required to facilitate the cooperation can sometimes outweigh its benefit. In this work, we first characterize the optimal energy efficient strategy for a single source-destination transmission aided by multiple relay nodes as a function of network parameters and the transmission rate. We then extend this result to a relay-assisted cellular broadcast network and determine an optimal solution which provides guidelines on the cooperative strategy that improves energy efficiency in such networks.

## OS.52-SLA.20 Behavioral Informatics: Enabling Technologies and Applications (II)

Session Chairs: Shri Narayanan, Panayiotis Georgiou Location: Trousdale Estates

**Analyzing the Language of Therapist Empathy in Motivational Interview based Psychotherapy**

Bo Xiao *University of Southern California*, Dogan Can *University of Southern California*, Panayiotis Georgiou *University of Southern California*, David Atkins *University of Washington*, Shrikanth Narayanan *University of Southern California*

Empathy is an important aspect of social communication, especially in medical and psychotherapy applications. Measures of empathy can offer insights into the quality of therapy. We use an N-gram language model based maximum likelihood strategy to classify empathic versus non-empathic utterances and report the precision and recall of classification for various parameters. High recall is obtained with unigram while bigram features achieved the highest F1-score. Based on the utterance level models, a group of lexical features are extracted at the therapy session level. The effectiveness of these features in modeling session level annotator perceptions of empathy is evaluated through correlation with expert-coded session level empathy scores. Our combined feature set achieved a correlation of 0.56 between predicted and expert-coded empathy scores. Results also suggest that the longer term empathy perception process may be more related to isolated empathic salient events.

**Using Interval Type-2 Fuzzy Logic to Analyze Turkish Emotion Words**

Ozan Cakmak *Mustafa Kemal University*, Serdar Yildirim *Mustafa Kemal University*, Abe Kazemzadeh *University of Southern California*, Shrikanth Narayanan *University of Southern California*

This paper describes a methodology that shows the feasibility of a fuzzy logic (FL) representation of Turkish emotionrelated words. We analyzed 197 Turkish emotion words set through a web-based survey that prompted users with emotional words and asked them to enter an interval valence, activation, and dominance emotion attributes using a double slider. Our previous experimental results indicated that there was a strong correlation between the emotions attributed to Turkish word roots and the Turkish sentences. In this paper, we extend our previous work and analyze Turkish emotion words by using an interval type-2 fuzzy logic.

**Dialogue Support for Memory Impaired People**

Luca Bellodi *ASML*, Radu Jasinschi *Philips*, Gerard De Haan *Philips*, Murtaza Bulut *Philips Research*

People affected by the loss of short term memory and cognitive impairment have serious difficulties in communication. This may lead to social isolation and lack of community access, a fundamental key barrier to independence for people suffering from Alzheimer's Disease, the most common form of memory and cognitive impairment. We propose Automated Memory Support for Social Interaction (AMSSI), a system that helps memory impaired people with their social interaction. The system provides active support that may help reducing stress level of patients. AMSSI recognizes visitors, determines the purpose of the visit, monitors the dialogue, determines whether the patient needs support, and provides feedback. AMSSI is tailored to patient needs, it has fast computation, full automation, and can be handled by the patient without supervision. The proposed assistive system can be beneficial for improving the quality of life of patients with mild to moderate cognitive impairments. This paper describes the implementation of the first working prototype of the AMSSI system. Validation user tests are still to be conducted.

**Composite-DBN for Recognition of Environmental Contexts**

Selina Chu *Oregon State University*, Shrikanth Narayanan *University of Southern California*, C.-C. Jay Kuo *University of Southern California*

People's behaviors are usually dictated by their surroundings. The local environment often affects the character and disposition of the people within it. The goal of our work is to automatically recognize the type of environments a person might be in. We introduce a hierarchical structure to recognize environmental contexts using the surrounding audio. We can use this structure to discover high-level representations for different acoustic environments in a data-driven fashion. Being able to perform such function allow us to better understand how we could utilize such information to assist in predicting a person's emotion or behavior. To accurately make an informative decision about behaviors or emotions, it is important to have the ability to differentiate between different types of environments. The nature of environmental sound contains large variances even within a single environment type and is constantly changing. These changes and events are dynamic and inconsistent. The goal is to come up with models that is robust enough to generalize to different situations. Learning a hierarchy of sound types might improve and clarify problems caused by the confusion between multiple acoustic environments with similar characteristics. We propose a framework for a composite of deep belief networks (composite-DBNs) as a way to represent various levels of representations and to recognize twelve different types of common everyday environments. Experimental results demonstrate promising performance in improving the state of art recognition for acoustic environments.

## OS.53-IVM.18 3DTV and Free-viewpoint TV (II)

Session Chairs: Masayuki Tanimoto, Yo-Sung Ho Location: Franklin Hills

**Global Optimization for Spatio-Temporally Consistent View Synthesis**

Hsiao-An Hsu *National Tsing Hua University*, Chen-Kuo Chiang *National Tsing Hua University*, Shang-Hong Lai *National Tsing Hua University*

We propose a novel algorithm to generate a virtualview video from a video-plus-depth sequence. The proposed method enforces the spatial and temporal consistency in the disocclusion regions by formulating the problem as an energy minimization problem in a Markov random field (MRF) framework. In the system level, we first recover the depth images and the motion vector maps after the image warping with the preprocessed depth map. Then we formulate the energy function for the MRF with additional shift variables for each node. To reduce the high computational complexity of applying BP to this problem, we present a multi-level BPs by using BP with smaller numbers of label candidates for each level. Finally, the Poisson image reconstruction is applied to improve the color consistency along the boundary of the disocclusion region in the synthesized image. Experimental results demonstrate the performance of the proposed method on several publicly available datasets.

**Depth Intra Skip Prediction for 3D Video Coding**

Kwan-Jung Oh *Samsung Electronics*, Jaejoon Lee *Samsung Electronics*, Du-Sik Park *Samsung Electronics*

Depth image compression plays a key role in the 3D video system. It can be used as supplementary data for rendering. In this paper, we present a depth intra skip prediction for 3D video coding. The depth intra skip prediction is designed based on the intra 16x16 mode. It exploits the estimated prediction direction which is derived from the adjacent neighboring pixels and does not encode any residual data. The proposed depth intra skip prediction is allowed for I slice as well as P and B slices. The usage of the depth intra skip prediction is signaled by a newly defined macroblock level flag. From the experiments, we confirm that the proposed depth intra skip prediction reduces the depth bit rate by up to 18.69% while preserving same synthesis quality for the virtual views.

**Depth Boundary Filtering for View Synthesis in 3D Video**

Yunseok Song *Gwangju Institute of Science and Technology*, Cheon Lee *Gwangju Institute of Science and Technology*, Woo-Seok Jang *Gwangju Institute of Science and Technology*, Yo-Sung Ho *Gwangju Institute of Science and Technology*

This paper presents a boundary sharpening method for depth maps to improve synthesis view quality. In general, coded depth maps exhibit noise and artifacts around object boundaries, leading to ineffective view synthesis. In our approach, gradient information is used to extract depth boundary regions. Afterwards, filtering based on distance, similarity, and direction is performed on such regions to replace depth values. The proposed algorithm was implemented on 3DV-ATM v0.3 as post-processing to coded depth maps. Experimental results showed 5.23% compared to the anchor results of 3DV-ATM v0.3. Subjective quality was improved as well.

**Single-Image Depth Inference based on Blur Cues**

Jingwei Wang *University of Southern California*, Hao Xu *University of Southern California*, C.-C. Jay Kuo *University of Southern California*

With the rapid advancement of 3D visual technology, the technique of depth inference from a single image has received new attention. In this work, we present several single-image depth inference algorithms based on the blur degree in different regions in one image. We identify two major sources of image blur: camera defocus and atmospheric reflectance. The latter is also known as the haziness. We build models for these two scenarios with the depth information as a model parameter. Thus, we are able to infer the depth information from the observed image. Experimental results are conducted on a large variety of images to demonstrate that the robustness of the proposed depth inference method.

## OS.54-SLA.21 Recent Advances in Speaker Characterization and Recognition

Session Chairs: Changchun Bao, Jia Min Karen Kua Location: Whitley Heights

**An Investigation into Better Frequency Warping for Time-Varying Speaker Recognition**

Linlin Wang *Tsinghua University*, Xiaojun Wu *Tsinghua University*, Thomas Fang Zheng *Tsinghua National Laboratory for Information Science and Technology*, Chenhao Zhang *Tsinghua University*

Performance degradation has been observed in presence of time intervals in practical speaker recognition systems. Researchers usually resort to enrollment data augmentation, speaker model adaptation, and variable verification threshold to alleviate the time-varying impact. However, in this paper, efforts have been made in the feature domain and an investigation into better frequency warping for the target task has been done. Two methods to determine the discrimination sensitivity of frequency bands are explored: an energy-based F-ratio measure and a performance-driven one. Frequency warping is performed according to the discrimination sensitivity curves of the whole frequency range. Experimental results show that the proposed methods outperform both MFCC and LFCC, and to some extent, alleviate the time-varying impact on speaker recognition.

**A K-Phoneme-Class based Multi-Model Method for Short Utterance Speaker Recognition**

Chenhao Zhang *Tsinghua University*, Thomas Fang Zheng *Tsinghua National Laboratory for Information Science and Technology*, Linlin Wang *Tsinghua University*, Xiaojun Wu *Tsinghua University*, Cong Yin *Taiyuan University of Technology*

For GMM-UBM based text-independent speaker recognition, the performance decreases significantly when the test speech is too short. Considering that the use of text information is helpful, a K-phoneme-class scoring based multiple phoneme class speaker model method (shortened as K-phoneme-class based multi-model method, abbreviated as KPCMMM) is proposed including a phoneme class speech recognition stage and a phoneme class dependent multi-model speaker recognition stage, where K means the number of most likely phoneme classes to be used in the second stage. Two different phoneme class definitions, expert-knowledge based and data-driven, are compared, and the performance as a function of K is also studied. Experimental results show that the data-driven phoneme class definition outperforms the expert-knowledge based one, and that an appropriate K value can lead to much better performance. Compared with the baseline GMM-UBM system, the proposed KPCMMM can achieve a relative equal error rate (EER) reduction of 38.60% for text-independent speaker recognition with a length of less than two seconds of test speech.

**A Study on Spoofing Attack in State-Of-The-Art Speaker Verification: The Telephone Speech Case**

Zhizheng Wu *Nanyang Technological University*, Tomi Kinnunen *University of Eastern Finland*, Eng Siong Chng *Nanyang Technological University*, Haizhou Li *Nanyang Technological University/Institute for Infocomm Research*, Eliathamby Ambikairajah *University of New South Wales*

Voice conversion technique, which modifies one speaker's (source) voice to sound like another speaker (target), presents a threat to automatic speaker verification. In this paper, we first present new results of evaluating the vulnerability of current state-of-the-art speaker verification systems: Gaussian mixture model with joint factor analysis (GMM-JFA) and probabilistic linear discriminant analysis (PLDA) systems, against spoofing attacks. The spoofing attacks are simulated by two voice conversion techniques: Gaussian mixture model based conversion and unit selection based conversion. To reduce false acceptance rate caused by spoofing attack, we propose a general anti-spoofing attack framework for the speaker verification systems, where a converted speech detector is adopted as a post-processing module for the speaker verification system's acceptance decision. The detector decides whether the accepted claim is human speech or converted speech. A subset of the core task in the NIST SRE 2006 corpus is used to evaluate the vulnerability of speaker verification system and the performance of converted speech detector. The results indicate that both conversion techniques can increase the false acceptance rate of GMM-JFA and PLDA system, while the converted speech detector can reduce the false acceptance rate from 31.54% and 41.25% to 1.64% and 1.71% for GMM-JFA and PLDA system on unit-selection based converted speech, respectively.

### PNCC-ivector-SRC based Speaker Verification

Eliathamby Ambikairajah *University of New South Wales*, Jia Min Karen Kua *University of New South Wales*, Vidhyasaharan Sethu *University of New South Wales*, Haizhou Li *Nanyang Technological University/Institute for Infocomm Research*

Most conventional features used in speaker recognition are based on Mel Frequency Cepstral Coefficients (MFCC) or Perceptual Linear Prediction (PLP) coefficients. Recently, the Power Normalised Cepstral Coefficients (PNCC) which are computed based on auditory processing, have been proposed as an alternative feature to MFCC for robust speech recognition. The objective of this paper is to investigate the speaker verification performance of PNCC features with a Sparse Representation Classifier (SRC), using a mixture of l_1 and l_2 norms. The paper also explores the score level fusion of both MFCC and PNCC i-vector based speaker verification systems. Evaluations on the NIST 2010 SRE extended database show that the fusion of MFCC-SRC and PNCC-SRC gave the best performance with a DCF of 0.4977. Further, cosine distance scoring (CDS) based systems were also investigated and the fusion of MFCC-CDS and PNCC-CDS presented an improvement in terms of EER, from a 3.99% EER baseline to 3.55%.

### Speaker Verification using Lasso based Sparse Total Variability Supervector with PLDA modeling

Ming Li *University of Southern California*, Charley Lu *3M Cogent*, Anne Wang *3M Cogent*, Shrikanth Narayanan *University of Southern California*

In this paper, we propose a Lasso based framework to generate the sparse total variability supervectors (s-vectors). Rather than the factor analysis framework, which uses a low dimensional Eigenvoice subspace to represent the mean supervector, the proposed Lasso approach utilizes the l1 norm regularized least square estimation to project the mean supervector on a pre-defined dictionary. The number of samples in this dictionary is appreciably larger than the typical Eigenvoice rank but the l1 norm of the Lasso solution vector is constrained. Only a small number of samples in the dictionary are selected for representing the mean supervector, and most of the dictionary coefficients in the Lasso solution are 0. We denote these sparse dictionary coefficient vectors in the Lasso solutions as the svectors and model them using probabilistic linear discriminant analysis (PLDA) for speaker verification. The proposed approach generates comparable results to the conventional cosine distance scoring based i-vector system and improvement is achieved by fusing the proposed method with either the i-vector system or the joint factor analysis (JFA) system. Experiments results are reported on the female part of the NIST SRE 2010 task with common conditions using equal error rate (EER), norm old minDCF and norm new minDCF values. The norm new minDCF cost was reduced by 7.5% and 9.6% relative when fusing the proposed approach with the baseline JFA and i-vector systems, respectively. Similarly, 8.3% and 10.7% relative norm old minDCF cost reduction was observed in the fusion.

## PS.6-IVM.19 Selected Topics in Image Processing and Computer Vision

Session Chair: Jin-Jang Leou Location: Solano

### Image Inpainting by Block-Based Linear Regression with Optimal Block Selection

Akira Tanaka *Hokkaido University*, Takahiro Ogawa *Hokkaido University*, Miki Haseyama *Hokkaido University*

Estimation of missing entries in a multivariate data is one of classical problems in the field of statistical science. One of the most popular approaches for this problem is linear regression based on the EM algorithm. When we consider to apply this approach to block-based image inpainting problems, we have additional information, that is, a target lost pixel could be included in multiple blocks, which implies that we have multiple candidates of estimates for the pixel. In such cases, we have to choose a good estimate among the multiple candidates. In this paper, we propose a novel image inpainting method incorporating optimal block selection in terms of the expected squared errors among multiple candidates of the estimate for the target pixel. Results of numerical examples are also shown to verify the efficacy of the proposed method.

### Generalized Histogram Shifting-Based Reversible Data Hiding with an Adaptive Binary-to-*q*-ary Converter

Masaaki Fujiyoshi *Tokyo Metropolitan University*, Hitoshi Kiya *Tokyo Metropolitan University*

This paper increases the flexibility of generalized histogram shifting-based reversible data hiding (HS-RDH). An RDH method modifies an original image to hide data to the image, and the method not only extracts the hidden data but also restores the original image from the distorted image which conveys the hidden data. A generalized HS-RDH method increases (or decreases) particular pixel values in an original image by $(q - 1)$, based on its tonal distribution, to hide $q$-ary data symbols, whereas an ordinary HS-RDH method shifts a part of the histogram by one to embed binary symbols. This paper introduces an adaptive binary-to-$q$-ary watermark converter and a tonal distribution analysis to increase conveyable hidden data size, whereas a conventional generalized HS-RDH method with an arithmetic decoder-based converter cannot always convert the extracted $q$-ary strings to original binary strings correctly and the other method embeds $\hat{q}$-symbols instead of $q$-ary symbols where $\hat{q}$ is a power of two equal to or less than $q$. In addition, histogram packing technique is introduced in this paper to further increase $q$. Experimental results show the effectiveness of the proposed method.

### An Enhanced Seam Carving Approach for Video Retargeting

Tzu-Hua Chao *National Chung Cheng University*, Jin-Jang Leou *National Chung Cheng University*, Han-Hui Hsiao *National Chung Cheng University*

Video retargeting (resizing) is an important task for displaying videos on various display devices. In this study, an enhanced seam carving approach for video retargeting is proposed, in which a seam may be a non-8-connected one. Both the search window size and the temporal weight can be adaptively adjusted according to video contents (motion information). Additionally, to preserve temporal coherence, the appearance-based method is employed. The spatial and temporal costs of a pixel are linearly combined to compute the cumulative cost with an adaptive temporal weight. Finally, dynamic programming is used to determine the optimal non-8-connected seam (with the minimum cumulative cost) for carving out. Based on the experimental results obtained in this study, the performance of the proposed approach is better than those of two comparison approaches.

### Adaptive Reversible Data Hiding in Frequency Domain via Integer-to-Integer Transform

Yoshida Taichi *Keio University*, Yusuke Okamura *Keio University*, Taizo Suzuki *Nihon University*, Masaaki Ikehara *Keio University*

This paper proposes an adaptive reversible data hiding algorithm for images, which embeds significant information in frequency domain based on integer-to-integer transform. Its embedding method is realized to modify state-of-the-art one according to transformed coefficients. It overcomes a problem of the conventional method that particular visual degradations generated by embedding often appear.

Our proposed algorithm outperforms the conventional one about the visual quality of embedded images, objectively and perceptually, while keeping embedding capacity.

### 3D Editing System for Captured Real Scenes

Inwoo Ha *Samsung Advanced Institute of Technology*, Yong Beom Lee *Samsung Advanced Institute of Technology*, D. K. James Kim *Samsung Electronics*

This paper presents a complete 3D editing system for real scenes, captured from conventional color-depth camera. From captured source and target images, desired source objects and target locations are selected. The source objects are copied and pasted to the desired location of the target image in color layer not considering their mutual illumination between the source objects and target image. To seamlessly composite the source objects to the target image based on their mutual illuminations, 3D surface meshes of source and target real scenes are reconstructed interactively, and differential rendering framework based on the instant radiosity is applied. The final result is a seamlessly mixed image considered with correct occlusions and mutual illumination.

### A Novel Criterion for Quality Improvement of JPEG Images Based on Image Database and Re-application of JPEG

Katsuya Kohno *Hokkaido University*, Akira Tanaka *Hokkaido University*, Hideyuki Imai *Hokkaido University*

Image compression is one of the most important technologies in the fields of image processing. JPEG has been commonly used for image compression. Since JPEG is a lossy compression method, decoded images exhibit visually unwanted noises. A need for techniques for improving the quality of JPEG images remains because there still exist many images compressed by JPEG today. Many methods for improving the quality of JPEG images have been proposed. Among them, a method based on re-application of JPEG, which means compression and decoding, is recognized as one of efficient methods. In our previous study, we improved this method by incorporating an image database and novel distance measures between two images. In this paper, we propose a new distance measure between two images to improve the performance of our previous method. We also show some results of numerical experiments to verify the efficacy of the proposed criterion.

### Recovery Method based Particle Filter for Object Tracking in Complex Environment

Yuhi Shiina *Waseda University*, Takeshi Ikenaga *Waseda University*

Object tracking is a key process for various image recognition applications, and many algorithms have been proposed in this field. Especially, particle filter has possibility for tracking objects steadily thanks to prediction using many particles. However, other objects that are of similar color or shape with a tracking object hijack a tracking region if there were such objects nearby the tracking object. It is a critical problem. This paper proposes a recovery method based particle filter by focusing a feature regions attached to an object. This proposal tracks both a feature region and an object including the region at once. This proposal utilizes a recovery method that pulls a tracking region back to an appropriate position using the prior frame's distance and angle between the two tracking regions when the tracking region is hijacked by other objects. Some video sequences including complex environment have been tested for evaluating this proposal. The experimental results show that this proposal can track a specified person in the sequences, while conventional method cannot track the person. This result represents that recovery method of proposal effectively works when other objects hijack the tracking region.

### Detection of Salient Object Using Pixel Blurriness

Yi-Chong Zeng *Institute for Information Industry*, Chi-Hung Tsai *Institute for Information Industry*

In this paper we propose a method to detect salient object in still image and non-slow motion background video. The key technique is measuring pixel blurriness. Generally speaking, when the salient object was taken in focus, pixels within the salient object should be sharper than those within the background. In the first step, image intensity is extracted and then four different-size average filters are applied to intensity. Subsequently, variation of intensity differences (VID) is computed among the original intensity and four blurred versions. The VID is employed to represent degree of pixel blurriness. Finally, a thresholding method is applied to pixel blurriness in order to distinguish salient object from background, and salient object is composed of low-blurriness pixels. The experiment results demonstrate that the proposed method is efficient in detection of salient object in still image and non-slow motion background video. Moreover, our method has better detection performance than the two compared methods.