

Advances in Multimedia Content Analysis and Signal Processing

Dong Xu · Weisi Lin · Anthony T. S. Ho

Published online: 19 December 2013
© Springer Science+Business Media New York 2013

The papers in this special issue cover a wide range of topics in multimedia content analysis and signal processing including multimedia content analysis, image and audio processing, and video coding and compression.

A) Multimedia Content Analysis

Enriching the abstract tags for social images is important for keyword-based social image search and retrieval. The paper entitled “Automatic Abstract Tag Detection for Social Image Tag Refinement and Enrichment” by Xia et al. constructs the concept ontology with three-level semantics to detect the candidates of abstract tags. Based on the concept ontology, new tags can also be added to enrich the tags of social images. The proposed methods were compared with other existing approaches to demonstrate its effectiveness.

Multimedia event detection is an important research topic due to its application in video indexing and retrieval. In the paper entitled “Multimedia Event Detection Using Segment-based Approach for Motion Feature,” Phan et al. propose a new approach for multimedia event detection by partitioning each video into segments for feature extraction and classification. The experimental results on the TRECVID Multimedia Event Detection datasets show promising results of their approach.

Saliency detection can be used for many real world applications such as object detection and recognition. In the paper

entitled “Top-Down Saliency Detection via Contextual Pooling,” Zhu et al. propose a new top-down approach for saliency detection by additionally using spatial context information. The results demonstrate their method achieves state-of-the-art performance for saliency detection.

Hand gesture recognition is an important component for human computer interaction. In the paper entitled “Real-time Hand Gesture Recognition from Depth Images Using Convex Shape Decomposition Method,” Qin et al. propose a new hand gesture recognition system by using depth images. New hand detection and segmentation methods were proposed before conducting gesture recognition. The experiments demonstrate that their system can accurately recognize hand gestures in real-time.

Colorizing the gray scale facial images is an interesting research topic in multimedia. In the paper entitled “Colorization for Gray Scale Facial Image by Locality-constrained Linear Coding,” Liang et al. develop a new colorization system for gray scale facial images by using Locality-constrained Linear Coding (LLC) and Markov Random Field (MRF). The experiments demonstrate the effectiveness of their proposed system.

B) Image and Audio Processing

Noise estimation is a fundamental issue for image denoising and many other image processing applications. In the paper entitled “Robust Noise Estimation Based on Noise Injection,” Tang et al. describe a new noise level estimation algorithm by investigating the distribution of local variances in natural images. They have used a wavelet decomposition based preliminary estimation stage to alleviate the influence of an image's textural or structural information, and then a noise injection based estimation stage to find the impact of image content on the variance distribution. Proper experiments and comparative analysis demonstrate that the proposed algorithm

D. Xu (✉) · W. Lin
School of Computer Engineering, Nanyang Technological
University, Blk N4, 2A-29, Nanyang Avenue, Singapore 639798
e-mail: DongXu@ntu.edu.sg

A. T. S. Ho
Department of Computing, University of Surrey, Guildford,
Surrey GU2 7XH, UK

can infer noise levels and has robust performance over a wide range of visual content, as compared to the existing relevant methods.

Audio is indispensable in multimedia applications. In “Audio Quality Requirements and Comparison of Multimodal vs. Unimodal Perception of Impairments for Long Duration Content” by Borowiak et al., the effect of the time dimension on quality ratings and user responses, as well as that of audio artifacts related to different compression rates over extended periods of time, are investigated. The study gives useful insights toward user quality expectations, user reaction time to quality degradation, user sensitivity to quality changes when he is able to influence the quality himself, and cross-modal effect between visual and auditory modalities.

In the next paper, “An Adaptive Non Reference Anchor Array Framework for Audio Retrieval in Teleconferencing Environment,” Nathwani et al. discuss an adaptive method for audio retrieval for live teleconferencing with multiple participants. A non reference anchor array (NRA) was used for capturing the interfering speech, in addition to the primary array that captures the speech source of interest (SOI). This method claimed to be computationally efficient since it does not require the computation of acoustic impulse response (AIR) of the teleconferencing room, and the NRA is able to remove correlated noise in the direction of the SOI. The proposed method was evaluated by conducting experiments with clean speech acquisition from distant microphone arrays, as well as using the existing databases.

C) Video Coding and Compression

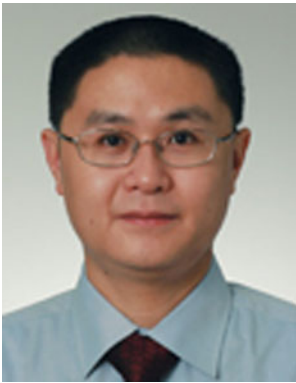
Video compression continues to be an important research area due to the fact that visual signal still consumes a dominant portion of system resources (like computation, bandwidth, storage, and so on) in multimedia services. This special issue includes three papers for 2D, multi-view and 3D video compression, respectively. The paper “SSIM-based Error Resilient Video

Coding over Packet-Switched Networks” by Zhang et al. proposes a structural similarity (SSIM) (rather than the traditional MSE) based error resilient video coding scheme, in order to improve perceived visual quality of compressed videos over packet-switched networks. To be more specific, a SSIM-based distortion model was first developed to estimate the perceptual distortion due to quantization, error concealment and error propagation; an adaptive mode selection strategy was then presented to enhance the robustness of the resultant algorithm.

In the next paper “Adaptive Learning Based View Synthesis Prediction for Multi-View Video Coding,” Hu et al. devise an adaptive learning based view synthesis prediction algorithm to enhance the prediction of virtual view picture for free-view TV by integrating the least square prediction with backward warping. It utilizes both adjacent view and temporal decoding information to adaptively learn the prediction coefficients. The proposed method has been demonstrated to save bitrates with up to 11 %–18 % when compared with the relevant existing approaches.

Because different parts of a depth map have different impacts on the synthesized image quality for 3-D video represented by texture plus depth map, Xiao et al. suggest a macroblock-level bit allocation method in the paper entitled “Macroblock Level Bits Allocation for Depth Maps in 3-D Video Coding.” In this work, different macroblocks of a depth map were encoded with different quantization parameters and coding modes. With the fine bit-allocation granularity, the proposed approach outperformed other bits allocation approaches, maintaining the synthesized view quality with little price in terms of pre-encoding delay.

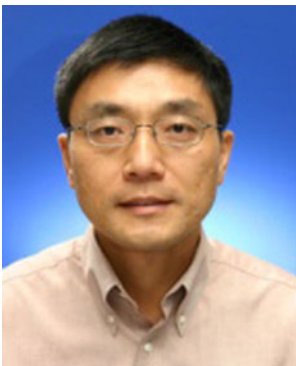
Acknowledgments We would like to thank all authors for submitting their high-quality papers to this special issue and all reviewers for spending their invaluable time to provide insightful and timely reviews. In addition, we would like to thank Prof. S. Y. Kung, the current Editor-in-Chief, Prof. Shuvra Bhattacharyya and Prof. Jarmo Takala, the current Co-Editors-in-Chief, for their support and Ms. Courtney Clark for her professional assistance in the paper review and production process.



Dong Xu received the B.Eng. and PhD degrees from University of Science and Technology of China, in 2001 and 2005, respectively. While pursuing the PhD degree, he worked at Microsoft Research Asia and The Chinese University of Hong Kong for more than two years. He also worked at Columbia University for one year as a postdoctoral research scientist. In May 2007, he joined Nanyang Technological University in Singapore, where he is currently an associate professor

with School of Computer Engineering.

His research focuses on new theories, algorithms and systems for intelligent processing and understanding of visual data such as images and videos. He was the co-author of a paper that won the Best Student Paper Award in CVPR 2010.



Weisi Lin received his Ph.D. from King's College London. He was the Lab Head and Acting Department Manager for Media Processing, in Institute for Infocomm Research, Singapore. Currently, he is an Associate Professor in Computer Engineering, Nanyang Technological University, Singapore. His research areas include image processing, perceptual multimedia modeling and evaluation, and video compression. He published 250+ refereed papers in international journals and conferences.

He is on the editorial boards of IEEE Trans. on Multimedia, IEEE SIGNAL PROCESSING LETTERS and *Journal of Visual Communication and Image Representation*. He chairs the IEEE MMTC IG on Quality-of-Experience. He has been elected as an APSIPA Distinguished Lecturer (2012/3). He is the Lead Technical-Program Chair for Pacific-Rim Conference on Multimedia (PCM) 2012 and International Workshop on Quality of Multimedia Experience (QoMEX) 2014, and a Technical-

Program Chair for *IEEE International Conference on Multimedia and Expo (ICME)* 2013. He is a fellow of Institution of Engineering Technology, and an Honorary Fellow, Singapore Institute of Engineering Technologists.



Professor Anthony T.S. Ho holds a Personal Chair in Multimedia Security and is currently Head of Department of Computing, University of Surrey since 2010. He also leads the Multimedia Security and Forensics research group in the Department. He was the recipient of the Institution of Engineering and Technology (IET) Innovation in Engineering Award under the Security category for his research and commercialization work on digital watermarking in 2006.

Professor Ho obtained his BSc (Hons) in Physical Electronics from Northumbria University in 1979, his MSc in Applied Optics from Imperial College London in 1980 and his PhD in Digital Image Processing from King's College London in 1983. After graduation, he worked in technical management positions in industry for 11 years in the UK and Canada.

From 1994 to 2005, Professor Ho was a Senior Lecturer and then Associate Professor at Nanyang Technological University (NTU), Singapore. While at NTU, he co-founded and became CEO of DataMark Technologies (DMT) in 1998, a company specialises in digital watermarking and steganography. DMT is now a subsidiary of Singapore Technologies Engineering. Professor Ho has published more than 140 articles in international journals and conference proceedings as well as 8 international patents granted related to watermarking and steganography. He recently received a best paper award for a co-authored conference paper on camera model identification from IEVC 2012. He is the Editor-in-Chief of the International Journal of Information Security and Applications (JISA) published by Elsevier.

Professor Ho was General Chair for the 8th International Workshop on Digital Watermarking (IWDW) held at the University of Surrey in 2009 and co-Editor for two Springer Proceedings on Digital Watermarking from IWDW08 and IWDW09. He is a Fellow of Institution of Engineering and Technology (FIET), Fellow of Institute of Physics (FInstP), Fellow of British Computer Society (FBCS) and Senior Member of IEEE.