CrossMark

Guest Editorial: Learning Multimedia for Real World Applications

**Bing-Kun Bao[1] · Congyan Lang[2] · Tao Mei[3] ·
Alberto del Bimbo[4]**

Multimedia data, as vivid and comprehensive, exist everywhere in our daily lives, in communication, education, manufacturing and service industries, and so on. Thus, how to improve the learning of multimedia data for real world applications has attracted widespread interests in the academy circle. This issue consists of 12 papers, which are briefly discussed as follows.

Three out of these 12 papers focus on developing feature representation to improve the performance of image retrieval, near duplicate image detection and video classification respectively. Sketch-based Image Retrieval (SBIR), which uses simple edge or contour images, is one important branch of Content-based Image Retrieval. However, SBIR is more difficult than CBIR due to the lack of visual information, this makes the Bag-of-Words (BoW) or codebook in SBIR hard to construct. The paper entitled "Sketch4Image: A Novel Framework for Sketch-Based Image Retrieval Based on Product Quantization with Coding Residuals" (10.1007/s11042-015-2645-y) proposes a novel SBIR framework based on Product Quantization (PQ) with sparse coding (SC) to construct an optimized codebook. In "Efficient

✉ Bing-Kun Bao
   bkbao@nlpr.ia.ac.cn

   Congyan Lang
   cylang@bjtu.edu.cn

   Tao Mei
   tmei@microsoft.com

   Alberto del Bimbo
   alberto.delbimbo@unifi.it

1 Chinese Academy of Sciences, Beijing, China

2 Beijing Jiaotong University, Beijing, China

3 Microsoft Research Asia, Beijing, China

4 University of Florence, Florence, Italy

Near-Duplicate Image Detection with a Local-Based Binary Representation" (10.1007/s11042-015-2472-1), Local-based Binary Representation is presented to encode an image as a binary vector for online near-duplicate image detection. The proposed representation is efficient to compute as well as robust in performance. It is also highly compact and it does not require any training phase. "A Bag-of-regions Representation for Video Classification" (10.1007/s11042-015-2876-y) proposes a bag-of-regions representation for high-level sparse representation of a video sequence. The BoR representation of a video sequence is obtained by extracting regions that exist in the majority of its frames and largely correspond to a single object.

Two papers study on the application in transportation, one is visual railway detection, and the other is traffic anomaly detection. "Visual Railway Detection by Superpixel based Intra-cellular Decisions" (10.1007/s11042-015-2654-x) proposes to detect railways based on superpixels. An SVM classifier is learned based on features, on which a TF-IDF like transform is applied, and it greatly improves the performance of the classification. "Traffic Anomaly Detection Based on Image Descriptor in Videos" (10.1007/s11042-015-2637-y) introduces a new traffic anomaly detection algorithm based on image description technology. The experimental results show that the proposed algorithm has improved performances of anomaly detection on both intersection traffic videos and main road traffic videos.

Two papers are related to social media, one is for sentiment analysis of social network multimedia, and the other is photo collection summarization. To address sentiment analysis of micro-blogging content, such as Twitter short messages, "A Multimodal Feature Learning Approach for Sentiment Analysis of Social Network Multimedia" (10.1007/s11042-015-2646-x) investigates the use of a multimodal feature learning approach, using neural network based models such as Skip-gram and Denoising Autoencoders. For photo collection summarization, the existing methods mainly consider the low-level features for photo representation only, while ignore many other useful features. The paper entitled "Multi-modal and Multi-scale Photo Collection Summarization" (10.1007/s11042-015-2658-6) proposes a multi-modal and multi-scale photo collection summarization method by leveraging multi-modal features, including time, location and high-level semantic features. The key photo ranking algorithm also takes the importance of both events and photos into consideration, and the proposed method allows users to control the scale of event segmentation and number of key photos selected.

One paper raises an interesting discussion on the importance of location information in saliency detection. "How Important is Location Information in Saliency Detection of Natural Images" (10.1007/s11042-015-2875-z) provides direct and quantitative analysis of the importance of location information for saliency detection in natural images. A location based saliency detection approach is proposed to completely initialize saliency maps with location information and propagate saliency among patches based on color similarity. The proposed method can handle natural images with different object positions and multiple salient objects.

"Compressive Sensing Reconstruction for Compressible Signal Based on Projection Replacement" (10.1007/s11042-015-2578-5) proposes the projection replacement (PR) algorithm by building the measurement space and its orthogonal complement space with singular value decomposition, and replacing the projection in measurement space of the reconstructed result with the pseudo-inverse one. The proposed PR algorithm eliminates the hypothetic measurement error in OMP and TSW-CS reconstructed model, and it guarantees theoretically that the PR results have a smaller error. Its effectiveness is verified experimentally with OMP and TSW-CS. The proposed algorithm can serve as a good reconstruction algorithm for the CS-based applications such as image coding, super-resolution, video retrieval etc.

In image denoising, we accepted the paper entitled "Non-local Sparse Regularization Model with Application to Image Denoising" (10.1007/s11042-015-2471-2), which studies on the denoising of natural images corrupted by Gaussian white noise. A framework is proposed to explore two sets of ideas involving on the one hand, locally learning a dictionary and estimating the sparse regularization signal descriptions for each coefficient; and on the other hand, nonlocally enforcing the invariance constraint by introducing patch self-similarities of natural images into the cost functional. The proposed framework outperforms the state-of-the-art, and makes it possible to effectively restore raw images from digital cameras at a reasonable speed and memory cost.

"An Implicit Relevance Feedback Method for CBIR with Real-Time Eye Tracking" (10.1007/s11042-015-2873-1) proposes a novel image retrieval system with implicit relevance feedback, named eye tracking based relevance feedback system (ETRFs). ETRFs is composed of three main modules: image retrieval subsystem based on bag-of-word architecture; user relevance assessment that implicitly acquires relevant images with the help of a modern eye tracker; and relevance feedback module that applies a weighted query expansion method to fuse users' relevance feedback. ETRFs is implemented online and real time, which makes it remarkably distinguish from other online systems.

Last but not the least, we also accepted one paper studying on music. "Note Onset Detection Based on Sparse Decomposition" (10.1007/s11042-015-2656-8) studies on music onset detection, which is significant and essential for obtaining the high-level music features such as rhythm, beat, music paragraph and structure. This paper proposes a new algorithm for note onset detection based on sparse decomposition to solve the problem of lacking adaptiveness for representing the stationary and non-stationary part of the music signal in traditional onset detection methods. This is the first attempt to employ sparse decomposition into music onset detection.

Bing-Kun Bao Dr. Bao received the Ph.D. degree in Control Theory and Control Application, Department of Automation,University of Science and Technology of China (USTC), China, in 2009. She is currently an assistant researcher at Institute of Automation, Chinese Academy of Sciences, and a researcher at China-Singapore Institute of Digital Media. Her research interests include cross-media cross-modal image search, social event detection, image classification and annotation, and sparse/low rank representation.

She received the Best Paper Award from ICIMCS'09. She served as a technical program committee member of several international conferences (MMM2013, ICIMCS2013, etc.), and a reviewer of over 10 prestigious international journals (TMM, TSMC-B, Multimedia Systems Journal, Information Science, Neurocomputing, etc.). She served as a special session organizer in MMM2013, and a guest editor in Multimedia System Journal.
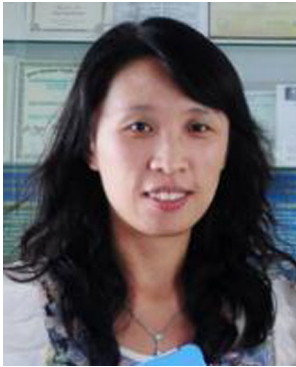
Congyan Lang Dr. Lang received the Ph.D. degree from Beijing Jiaotong University, Beijing, China, in 2006. She is currently a professor with the School of Computer and Information Technology, Beijing Jiaotong University. From December 2010 to December 2011, she was a Visiting Professor at the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. From Feb. to Aug. 2013, she was a visiting Professor at Multimedia Computing Lab, MSRA. She has authored or coauthored more than 40 technical papers over a wide range of research topics. Her current research interests include multimedia information retrieval and analysis and computer vision.

Tao Mei Dr. Tao is a Lead Researcher with Microsoft Research Asia, Beijing, China. He received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2001 and 2006, respectively. His current research interests include

multimedia information retrieval and computer vision. He has authored or co-authored over 140 papers in journals and conferences, eight book chapters, and edited two books. He holds seven U.S. granted patents and more than 20 in pending.

Dr. Mei was the recipient of several paper awards from prestigious multimedia conferences, including the Best Paper Award and the Best Demonstration Award at ACM Multimedia in 2007, the Best Poster Paper Award at the IEEE MMSP in 2008, the Best Paper Award at ACM Multimedia in 2009, the Top 10% Paper Award at the IEEE MMSP in 2012, the Best Paper Award at ACM ICIMCS in 2012, the Best Student Paper Award at the IEEE VCIP in 2012, and the IEEE Trans. on Multimedia Prize Paper Award 2013. He was the principle designer of the automatic video search system that achieved the best performance in the worldwide TRECVID evaluation in 2007. He received Microsoft Gold Star Award in 2010, and Microsoft Technology Transfer Awards in 2010 and 2012. He is an Associate Editor of Neurocomputing and the Journal of Multimedia, a Guest Editor of the IEEE Trans. on Multimedia, the IEEE Multimedia Magazine, the ACM/Springer Multimedia Systems, and the Journal of Visual Communication and Image Representation. He is the Program Co-Chair of MMM 2013, and the General Co-Chair of ACM ICIMCS 2013. He is a Senior Member of the IEEE and the ACM.

Alberto del Bimbo Dr. Bimbo is the Director of Media Integration and Communication Center and the Master in Multimedia program of the University of Florence, Italy. He was the Deputy Rector for Research and Innovation transfer at the University of Florence and the President of the Foundation for Research and Innovation. His scientific interests and activities have been in the field of Computer Vision and Multimedia Systems and is the author of over 300 publications that have appeared in the most distinguished international journals and conference proceedings. He is the author of "Visual Information Retrieval" a monography on content-based retrieval from image and video databases.

He is Member of IEEE and Fellow of IAPR. He was the Vice-President and President of the IAPR Italian Chapter, and Member of IEEE Publication Board. He is Associate Editor of Multimedia Tools and Applications Journal, Pattern Analysis and Applications, Journal of Visual Languages and Computing and International Journal of Image and video Processing. He was Associate Editor of IEEE Transactions on Multimedia, IEEE Transactions on Pattern Analysis and Machine Intelligence and Pattern Recognition. He has been the Guest Editor of many Special Issues on image and video retrieval and multimedia systems on some of the most prestigious scientific journals.

He served as the General Chair and Member of the Program Committee of many international conferences and workshops in the fields of image analysis and computer vision, multimedia computing and systems. Among the most imporetant and prestigeous ones, he was the General Chair of the 9th IAPR International Conference on Image Analysis and Processing ICIAP 1997, the 6th IEEE International Conference on Multimedia Computing and Systems ICMCS 1999, the ACM MULTIMEDIA 2010, the International Conference on Multimedia Retrieval ICMR 2011, and the European Conference on Computer Vision ECCV2012. He was also the Program Co-Chair of the ACM International Conference on Computer Human Interaction CHI 2008, the ACM MUL-TIMEDIA 2008 and the International Conference on Pattern Recognition ICPR 2012.