

Guest Editorial: Knowledge-Based Multimedia Computing

Liang Li¹ · Zi Huang² · Zheng-Jun Zha³ ·
Shuqiang Jiang¹

Published online: 2 October 2017

© Springer Science+Business Media, LLC 2017

The rapid advances in multimedia technology have resulted in a proliferation of multimedia data on the Internet. This is reflected in the success of the social networks, such as Facebook, Twitter, YouTube, Flickr, and Pinterest, which dramatically increased the volume of community-shared media, including images and videos. Although these websites allow users to annotate and rate them, the accurate annotations of online media are very rare and unsatisfactory. Thus, accurately understanding this multimedia content is a very significant and challenging issue.

The scenario of online multimedia understanding has usually large number of categories with unconstrained domains and noise. Recent progress on visual genome dataset and deep model open an exciting new era of knowledge-based multimedia computing, which can provide a knowledge base of images and capture the complex content with domain-specific knowledge. Moreover, some works about dense image captioning, visual relationship detection, visual question answering, knowledge inference, and social network knowledge graph also provide insight into tackling the knowledge-based multimedia computing. The prior

✉ Liang Li
liang.li@vipl.ict.ac.cn

Zi Huang
huang@itee.uq.edu.au

Zheng-Jun Zha
zhazj@ustc.edu.cn

Shuqiang Jiang
sqjiang@ict.ac.cn

¹ Institute of Computing Technology, Chinese Academy of Sciences, Haidian Qu, Beijing 100080, China

² School of Information Technology & Electrical Engineering, University of Queensland, St Lucia, QLD 4072, Australia

³ School of Information Science and Technology, University of Science and Technology of China, Hefei 230000, China

knowledge, experience, and human perceptual theory are playing a critical pole in accurately understanding the multimedia under the network environment.

This special issue aims at providing a forum to present the original and innovative research works that report the latest results and advances in the field of knowledge-based multimedia computing. This special issue contains 11 papers, selected from the initial 23 submissions, after two rounds of blind review.

The paper “CrossbowCam: A cost-effective multi-camera system for advanced applications” by Che-Hao Hsu et al. [4] proposes a novel multi-functional, low-cost handheld multi-camera system, “CrossbowCam.” The system is suitable for multi-viewpoint image acquisition, smooth switching, alignment, and seamless stitching applications. With the proposed system, the users can push one single button to change the configuration of the camera array rapidly to divergence (convex arc), parallel (linear), or convergence (concave arc). The three camera configurations can each be suitable for applications such as panorama image stitching, auto stereoscopic 3D display, bullet-time (time-freeze) visual effect, 3D scene reconstruction, etc.

The paper “Deep Learning based Basketball Video Analysis for Intelligent Arena Application” by Wu Liu et al. [7] proposes a deep learning based video analysis scheme for intelligent basketball arena applications. First of all, with multiple cameras or mobile devices capturing the activities in arena, the proposed scheme can automatically select the camera to give high-quality broadcast in real time. Furthermore, with basketball energy image based deep conventional neural network, the scoring clips as the highlight video reels are detected to support the wonderful actions replay and online sharing functions.

The paper “Hyperspectral Image Compression Based on Online Learning Spectral Features Dictionary” by Worku Jifara et al. [5] proposes a novel method of lossy hyperspectral image compression using online learning dictionary. From the perspective of sparse coding, learning a sparse dictionary could achieve a better result of data decorrelation. In order to compress the hyperspectral data, an online learning sparse coding dictionary which could describe the characteristics of spectral curve was created to represent and reconstruct hyperspectral data. In the online learning phase, effective clustering algorithm is applied to generate and update the dictionary more properly.

The paper “Sparse Representations based Distributed Attribute Learning for Person Re-identification” by Keyang Cheng et al. [1] is to solve the person re-identification task, and proposes a novel Sparse Representations based distributed attribute learning model (SRDAL) to encode targets into semantic topics. Compared to other models such as ELF, the proposed model performs best by imposing semantic restrictions onto the generation of human specific attributes and employing deep convolutional neural network to generate features without supervision for attributes learning model.

The paper “An Auto-Encoder-Based Summarization Algorithm for Unstructured Videos” by Meng-Xiong Han et al. [2] proposes an Auto-encoder-based summarization algorithm for unstructured videos. Each video structure is detected by an auto-encoder and both of the interestingness and representativeness of each video segment are predicted by the reconstruction errors of the segment. Meanwhile, the most interesting and representative summarization is generated with the limited summary length.

The paper “Efficient PCIe Transmission for Multi-Channel Video Using Dynamic Splicing and Conditional Prefetching” by Tingshan Liu et al. [6] proposes an efficient PCIe transmission method for multi-channel video. Firstly, a dynamic splicing mechanism is introduced to combine the video analyzed data and the compressed stream with the raw video to avoid the individual transmission of the auxiliary data. Secondly, a conditional prefetching mechanism is

employed to determine whether there exists any entire video frame in other channel buffers. Finally, in the host-side driver, direct kernel buffer access technique is used to improve the application I/O request packet (IRP) performance.

The paper “Pedestrian Detection based on Multi-Convolutional Features by Feature Maps Pruning” by Ting Rui et al. [10] proposes a feature map selection method to reduce the number of high dimensional feature maps in shallow layers, which cuts the feature map number by correlation coefficient between kernels and finishes detection by HOG + SVM method. Firstly, the feature maps of shallow layers from trained CNN are extracted. Then, strongly relevant feature maps are merged and all maps among weakly relevant feature maps are chosen by analyzing correlation coefficient of kernels. Finally, HOG features of the chosen feature maps are extracted and SVM is used to complete the training and classification.

The paper “Salient Object Detection via Multiple Saliency Weights” by Weimin Tan et al. [11] proposes a novel bottom-up approach to automatically detect salient objects of an image via multiple visual cues. The key idea is to represent a saliency map of an image as an integration of multiple visual cues, i.e., local contrast weight, superpixel clarity weight, background probability weight, and central bias weight. To obtain the saliency map, the four resulting saliency weights are integrated in a principled way via multiplication and summation based fusion. Furthermore, a new superpixel-level saliency smoothing approach is proposed to optimize the integrated results for producing clean and consistent saliency maps.

The paper “Cross-media Similarity Metric Learning with Unified Deep Networks” by Jinwei Qi et al. [9] proposes the Unified Network for Cross-media Similarity Metric (UNCSM) to associate cross-media shared representation learning with distance metric in a unified framework. First, a two-pathway deep network pre-trained with contrastive loss is designed, and double triplet similarity loss for fine-tuning is employed to learn the shared representation for each media type by modeling the relative semantic similarity. Second, the metric network is designed for effectively calculating the cross-media similarity of the shared representation, by modeling the pairwise similar and dissimilar constraints.

The paper “Cross-lingual event-centered news clustering based on elements semantic correlations of different news” by Xudong Hong et al. [3] proposes a new method based on semantic correlations of news elements to solve the problem of similarity computation between bilingual documents. First, use bilingual entity lexical and terms co-occurrences in news to acquire the semantic correlation of news elements in different language. Then, the similarity between news in different languages is computed using the GVSM model on this basis. Finally, spectral clustering is applied to categorize news stories.

The paper “Justify Role of Similarity Diffusion Process in Cross-Media Topic Ranking: An Empirical Evaluation” by Junbiao Pang et al. [8] empirically revisits the correlations between the types of noises and modalities for cross-media topic ranking, in order to provide the necessary insights to understand when to choose a noise. This paper reviews the existing unsupervised ranking methods and compares them in a unified evaluation criterion. We evaluate different noises, Poisson noise and Gaussian noise, for different modalities, i.e., texts and images.

These 11 papers cover a wide range of knowledge-based multimedia computing, and we hope the readers will find interesting ideas in them. Finally, the guest editorial team would like to thank all the authors for contributing their work to this special issue, and to the reviewers for their hard work and constructive comments. We would also like to express our gratitude to Prof. Borko Furht, editor-in-chief for providing an opportunity to organize this special issue, as well as for his helpful guidance in the reviewing process of this special issue.

References

1. Cheng K, Hui K, Zhan Y, Li M (2017) Sparse representations based distributed attribute learning for person re-identification. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-017-4967-4>
2. Han M-X, Hu H-M, Liu Y, Zhang C, Tian R-P, Zheng J (2017) An auto-encoder-based summarization algorithm for unstructured videos. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-017-4485-4>
3. Hong X, Yu Z, Tang M, Xian Y (2017) Cross-lingual event-centered news clustering based on elements semantic correlations of different news. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-017-4838-z>
4. Hsu C-H, Cheng W-H, Wu Y-L, Huang W-S, Mei T, Hua K-L (2017) CrossbowCam: a cost-effective multi-camera system for advanced applications. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-017-4852-1>
5. Jifara W, Jiang F, Zhang B, Wang H, Li J, Grigorev A, Liu S (2017) Hyperspectral image compression based on online learning spectral features dictionary. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-017-4724-8>
6. Liu T, Jiang H, Li H, Li B, Duan M (2017a) Efficient PCIe transmission for Multi-Channel video using dynamic splicing and conditional prefetching. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-017-4410-x>
7. Liu W, Yan CC, Liu J, Ma H (2017b) Deep learning based basketball video analysis for intelligent arena application. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-017-5002-5>
8. Pang J, Huang J, Zhang W, Huang Q, Yin B (2017) Justify role of Similarity Diffusion Process in cross-media topic ranking: an empirical evaluation. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-017-5037-7>
9. Qi J, Huang X, Peng Y (2017) Cross-media similarity metric learning with unified deep networks. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-017-4726-6>
10. Rui T, Zou J, Zhou Y, Fang H, Gao Q (2017) Pedestrian detection based on multi-convolutional features by feature maps pruning. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-017-4837-0>
11. Tan W, Yan B (2017) Salient object detection via multiple saliency weights. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-017-4725-7>

Liang Li is an assistant professor with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. He received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences in 2013. He received the B.S. degree in computer science from Xi'an Jiaotong University, Shaanxi, China, in 2008. He was a Post-doctoral Fellow with the University of Chinese Academy of Sciences from 2013 to 2015. His research interests include image processing, large-scale image retrieval, image semantic understanding, multimedia content analysis, computer vision, and pattern recognition. He has authored or coauthored more than 30 papers on the related research topics, including IEEE TNNLS, TMM, ACM Multimedia, CVPR, IJCAI, DCC, etc. He won the best paper award of Gamenets2014.

Zi Huang is an ARC Future Fellow in School of ITEE, The University of Queensland. She received her BSc degree from the Department of Computer Science, Tsinghua University, China, in 2001, and her PhD in Computer Science from School of ITEE, The University of Queensland, in 2007. Huang's research interests mainly include multimedia indexing and search, social data analysis, and knowledge discovery. She has authored or coauthored more than 80 papers, including IEEE TIP, TKDE, TMM, TOC, ACM TOIS, ACM Multimedia, CVPR, VLDB, ICDE, WWW, and SIGMOD. She was the area chair of ACM Multimedia2015, Workshop Chair of ICDE2015, Demo Chair (2013) and Publication Chair (2010) of Asia-Pacific Web Conference. She has also served as a TPC member of ACM Multimedia 2014, 2012; VLDB 2013; SIGIR 2008, 2010; and ICDE 2012, 2013, 2014. She won the best paper awards of WISE2013, DASFAA2012, and APWEB2010. She also won the Chris Wallace Award for Outstanding Research of Australia in 2016, Women in Technology (WiT) Infotech Research Award in 2015, and the Early Career Researcher Award of the University of Queensland in 2011.

Zheng-Jun Zha is a Professor with University of Science and Technology of China, Hefei, China. His current research interests include multimedia big data computing and computer vision. He has published over 100 referred papers in these areas. He was the recipient of several paper awards from prestigious multimedia conferences, including the best paper award in ACM Multimedia 2009, the Best Demo Runner-Up award in ACM Multimedia 2012, the Best Student Paper Award in ACM Multimedia 2013, the Best Paper Award in ICIMCS 2013, and the Best Paper Candidate Award in PCM 2014, etc. He served as a guest editor, organizer, and area chair for many prestigious international journals and conferences. Dr. Zha received the B.E. degree (2004) in automation and the Ph.D. degree (2009) in pattern recognition and intelligent system from the University of Science and Technology of China (USTC), Hefei, China, respectively. He worked as a Senior Research Fellow (2011–2013) and Research Fellow (2009–2010) in the School of Computing, National University of Singapore (NUS).

Shuqiang Jiang is a professor with the Institute of Computing Technology and the Key Laboratory of Intelligent Information Processing, Chinese Academy of Sciences, Beijing, China. He has authored or coauthored more than 100 papers. His current research interests include multimedia processing and semantic understanding, pattern recognition, and computer vision. Dr. Jiang is a member of the ACM, CCF, and YOCSEF. He is the Executive Committee member of ACM SIGMM, China chapter. He has been a guest editor for the special issues for Pattern Recognition, Multimedia Tools and Applications, and Multimedia Systems. He is the General Chair of ICIMCS2015, Program Chair of ICIMCS2010, Special Session Chair of PCM2008 and ICIMCS2012, Area Chair of PCIVT2011, Publicity Chair of PCM2011, and Proceedings Chair of MMSP2011. He has also served as a TPC member for more than 20 conferences, including ACM Multimedia, CVPR, ICCV, ICME, ICIP, and PCM. He won the Lu Jiaxi Young Talent Award from Chinese Academy of Sciences in 2012, and the CCF Award of Science and Technology in 2012. He was supported by the New-Star Program of Science and Technology of Beijing Metropolis in 2008.