

Guest editorial: selected papers from ICIMCS 2012

Zheng-Jun Zha¹ · Yan Liu² · Shin'ichi Satoh³ · Xinguo Yu⁴ · Rainer Lienhart⁵

International Conference on Internet Multimedia Computing and Services (ICIMCS) is an annual conference sponsored by ACM SIGMM China Chapter. The conference is especially interested in the latest technologies and applications that deal with the Web-scale processing and management of heterogeneous data from the Internet for multimedia computing and service. ICIMCS 2012 was held in Wuhan, China. The conference attracted around 90 participants, including researchers from academia and industries across ten countries/regions, for sharing their recent works on the topics ranging from visual feature representation and analysis to visual recognition and classification, and from social and mobile media analysis to multimedia services.

This special issue comprises the extended versions of seven papers, including three best papers and four papers from the regular and special sessions of ICIMCS 2012.

✉ Zheng-Jun Zha
junzzustc@gmail.com

Yan Liu
csyliu@comp.polyu.edu.hk

Shin'ichi Satoh
satoh@nii.ac.jp

Xinguo Yu
xgyu@mail.ccnu.edu.cn

Rainer Lienhart
Rainer.Lienhart@informatik.uni-augsburg.de

¹ Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, China

² Hong Kong Polytechnic University, Hong Kong, China

³ National Institute of Informatics, Tokyo, Japan

⁴ Central China Normal University, Wuhan, China

⁵ University of Augsburg, Augsburg, Germany

These papers cover key issues in Internet multimedia computing and services, including large-scale visual search, mobile media processing, Internet image tagging, video transmission, as well as fundamental technologies for intelligent multimedia computing, such as image segmentation, logo recognition, and object tracking.

The first article “Visual word expansion and BSIFT verification for large-scale image search” by Zhou et al. presents a visual word expansion method improving the recall of retrieval on large-scale image collections and a binary SIFT verification method boosting the retrieval precision. The former represents a query image as the aggregation of visual words of local features in the query and the neighboring words of these query words. The latter verifies feature matching between images based on the binarized SIFT feature. The work is experimented on two public image data sets. The second article “Accurate sensing of scene geo-context via mobile visual localization” by Liu et al. proposes a new mobile image geo-tagging approach toward accurate location tagging for images captured in urban areas. For a landmark image, the approach infers comprehensive and accurate geo-context of the image, including camera location, view direction, and scene location, by exploiting large-scale image retrieval and 3D reconstruction techniques. Experimental results on the San Francisco street view data set, which consists of approximately 150 k panoramic images in San Francisco at about 4-m intervals, have demonstrated the effectiveness of the approach. The third article entitled “Tag ranking based on salient region graph propagation” by Tang et al. presents a new tag ranking solution to rank user-contributed tags associated with Internet images. The solution exploits salient regions in images and ranks tags by integrating visual clues from entire images and salient regions of images. It constructs two sparse graphs over images and salient sub-images,

respectively, based on visual affinities. The relevance of each tag to images is propagated over the image graph and sub-image graph successively, leading to refined relevance scores. The tags associated with an image are then ranked according to their relevance scores. The approach is evaluated on the NUS-WIDE image set. The fourth article “Transmission of multimedia contents in opportunistic networks with social selfish nodes” by Pan et al. proposes a new transmission scheme toward facilitating multimedia content delivery in opportunistic networks. Different from existing works that do not consider the social selfish nodes in network, this work boosts the social selfish nodes to forward messages for other nodes. It makes use of historical records, cache resources, and social ties of nodes to improve the cooperation between nodes and guarantee the integrity of multimedia transmission in opportunistic networks with selfish nodes. The results from simulations have shown that the proposed transmission scheme performs better than existing solutions.

The next three articles in this issue focus on fundamental technologies for intelligent multimedia computing, including image segmentation, logo recognition, and object tracking. In “Nonlocal variational image segmentation models on graphs using the Split Bregman,” Lu et al. propose a new nonlocal variational segmentation technology, which extends conventional active contour segmentation model into a nonlocal means framework based on a graph structure. It defines an energy function that exploits a nonlocal variational regularization term and a modified local binary fitting term. While the former can preserve the structure of objects, the latter is able to address the intensity in homogeneity in images. For minimizing the energy function, the

split Bregman method is employed, leading to segmentation results. Experimental results on medical and remote sensing images have demonstrated the effectiveness of this work. Wang et al. in “Finding logos in real-world image with point-context representation-based region search” address the problem of logo detection in real-world images. They combine contextual shape and patch information around feature points in images into a new feature representation, i.e., point-context representation, improving the discriminability of single point features. To alleviate the inference of image background, they segment images into region trees and transfer log recognition to a region-to-image search problem. The method is evaluated on the Flickr Logos 27 and CASIA-Logo datasets. In “Multi-object tracking via MHT with multiple information fusion in surveillance video,” Ying et al. focus on tracking multiple objects in surveillance videos. They propose a multiple hypotheses tracking algorithm that exploits multiple information including appearance feature, local motion pattern feature, and repulsion-inertia model. Experimental results on five video sequences from TRECVID 2011 and two video sequences from PETS 2009 S2.L1 have shown that the proposed multi-object tracking algorithm generates better trajectories with less missing objects and identity switches than existing solutions.

Acknowledgments We would like to thank Prof. Thomas Plagemann for providing us the opportunity to organize this special issue. We thank the reviewers for their great efforts. Their professional evaluations and constructive comments are vital for securing the high quality of the special issue. Finally, we express our thanks to all the authors who have contributed to the special issue.