Machine Vision and Applications (2014) 25:1631–1633
DOI 10.1007/s00138-014-0618-1

EDITORIAL

# Special issue on contextual vision computing

**Richang Hong · Qi Tian · Nicu Sebe**

Published online: 21 May 2014


The popularity of web 2.0 content brings the proliferation of social media in recent years. The intrinsic attributes of social media are to facilitate interactive information sharing, interoperability and collaboration on the internet. By virtue of that, web images and videos are generally accompanied by user-contributed contextual information such as tags, comments, etc. Massive emerging social media data offer new opportunities for resolving the long-standing challenges in computer vision. Fox example, how to jointly represent the visual aspect and user annotation of multimedia data and how can we build video indexing and enable search to benefit from contextual information? So we face both challenges and opportunities in the research on contextual vision computing. This special issue is organized with the purpose of introducing novel research work on contextual vision computing. Submissions are from an open call for paper. With the assistance of professional referees, ten papers out from seventeen submissions are accepted after two rounds of rigorous reviews. These papers cover a wide range of subtopics of contextual vision computing, including visual representation, image classification, tag localization, saliency detection, pedestrian detection, and so on.

The first part of the special issue contains three papers. These papers focus on the image representation, classification and local semantic analysis by directly leveraging user-generated context information. In the first paper, "Semi-supervised Unified Latent Factor Learning with Multi-view Data", Jiang et al. present a semi-supervised unified latent factor learning approach to learn a predictive unified latent representation. They claim that the web multimedia resources can be considered as multi-view data among which the complementary information and the supervision from the partially label information are used to learn the representation. Experimental results verify the more discriminating power of such representation. In the second paper, "Inductive Hierarchical Nonnegative Graph Embedding for Object–Verb Image Classification", Sun et al. introduce a scheme called Inductive Hierarchical Nonnegative Graph Embedding (IHNGE). They believe the real world images contain "verb–object" concepts rather than only "object" and the hierarchical structure embedded in these "verb–object" concepts can help enhance the performance of classification. In the third paper, "Localizing Relevant Frames in Web Videos Using Topic Model and Relevance Filtering", Li et al. describe a scheme to localize relevant frames by combining topic model and relevance filtering. The scheme is comprised of three steps: (1) use relevance filtering to get the top ranked frames, (2) separate the frames by topics using latent Dirichlet allocation-based semantic analysis, and (3) refine the raw relevant frame set through topic relevance. Experiments on a large-scale Web video dataset demonstrate the effectiveness of the scheme.

The next part contains two papers focusing on saliency detection and super-resolution image reconstruction. In this part, context refers to the auxiliary data such as eye-tracking and multi-view data. The paper "Image Visual Attention Computation and Application via the Learning of Object Attributes" introduces a framework to explore image visual attention via the learning of object attributes from eye-tracking data. This paper aims to solve three problems: pixel level attention computation, i.e., the saliency map, the image-level visual attention computation and the application of

R. Hong (✉)
Hefei University of Technology, Hefei, China
e-mail: hongrc.hfut@gmail.com

Q. Tian
University of Texas at San Antonio, San Antonio, USA

N. Sebe
University of Trento, Trento, Italy

these computation models in image categorization. Comprehensive evaluations of saliency detection and image categorization are conducted on publicly available benchmarks and the performance of their proposed framework is superior to the state-of-the-art methods. The paper "A New Closed Loop Method of Super-Resolution for Multi-view Images" presents a method to resolve the multi-view super-resolution. For the mixed resolution multi-view case where the input is one high-resolution view along with its neighboring low-resolution views, their method is able to produce the super-resolution image based on the depth map. The method consists of three steps of stereo matching, depth fusion and super-resolution. They formulate the super-resolution as an optimization problem under the guidance of the estimated depth information.

In the third part, we have three papers on improving the performance of object and person detection, identification and tracking. In the paper "A sparse Coding based Transfer Learning Framework for Pedestrian Detection", Liang et al. propose a transfer learning framework based on sparse coding to detect pedestrian in surveillance video. They employ generic detector to get the initial target samples and sparse coding to calculate the weights for source samples and target samples. By adding weights during retraining process, the outliers are removed from the source samples and the drift problem in the target samples is tackled. This finally works out a scene-specific pedestrian detector. The paper "Context-Based Person Identification Framework for Smart Video Surveillance" introduces a framework that leverages heterogeneous contextual information together with facial features to handle the person identification. They claim that the analysis of facial only is not sufficient to deal with poor quality data. Therefore, the heterogeneous context features including clothing, activity, human attributes, etc., are integrated into their framework. Experiments on the real surveillance videos demonstrate its superiority. Motivated by that traditional particle filter that uses simple geometric shapes for representation is not able to track objects with complex shape accurately. Sun et al. presents a refined particle filter method for contour tracking based on a determined binary level set model in the paper "A Refined Particle Filter Based on Determined Level Set Model for Robust Contour Tracking". In their method, some prior knowledge of the target model is taken into consideration in the update process of particle filter. Experiments conducted on several challenging video sequences show the effectiveness and the efficiency of the method.

The final part of the special issue is composed of two papers on video relighting and geometry completion. In the paper "Free-viewpoint Video Relighting from Multi-View Sequence under General Illumination", Li et al. propose an approach to create plausible free-viewpoint relighting video using multi-view camera array under general illumination. In their method, they construct 3D model of the captured target using multi-view stereo approach, and estimate the spatially varying surface reflectance in the spherical harmonics domain. 3D target is relit by a flow- and quotient-based transfer strategy based on the estimated geometry and reflectance. The free-viewpoint video is generated using a view-dependent rendering strategy. Extensive experiments demonstrate their approach enables plausible free-view video relighting. In the paper "Detail-Generating Geometry Completion for Point-Sampled Geometry", Wang et al. presents a method for detail-generating geometry completion over point-sampled geometry. The motivation of this work is to convert the context-based geometry completion into the detail-based texture completion on the surface. They construct a smooth patch covering the hole and perform region-growing clustering to produce the patching units. The geometry details on the smooth patches are finally generated by optimizing a constrained global texture energy function on the point-sampled surfaces. Experiments verify that the method is able to produce efficient patches that conform to their boundaries and meanwhile contain plausible 3D surface details.

In conclusion, the papers in this special issue cover the techniques addressing different challenges in contextual vision computing. We believe this special issue will benefit researchers and practitioners working in this area.

**Acknowledgments** We would like to thank Prof. Mubarak Shah for providing us the chance 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.

**Dr. Richang Hong** is currently a Professor at School of Computer and Information, Hefei University. Dr. Richang Hong received his Ph.D. degrees in July 2008 from the University of Science and Technology of China (USTC). His current research interests include multimedia question answering, social media analysis and multimedia data management. He is a recipient of the Best Paper Award in ACM Multimedia 2010.

**Dr. Qi Tian** is currently an Associate Professor in the Department of Computer Science, University of Texas at San Antonio (UTSA). During 2008–2009, he took one-year Faculty Leave at Microsoft Research Asia (MSRA) in the Media Computing Group (former Internet Media Group). He received his Ph.D. in 2002 from UIUC. Dr. Tian's research interests include multimedia information retrieval and computer vision. He has published about 160 refereed journal and conference papers in these fields. His research projects were funded by NSF, ARO, DHS, Google, NEC Laboratories of America, HP Lab, FXPAL, SALSI, and Akiira Media System, Inc. He is the co-author of a Best Student Paper in ICASSP 2006, and co-author of a Best Paper Candidate in PCM 2007. He was a nominee for 2008 and 2010 UTSA President Distinguished Research Award. He received 2010 ACM Service Award. He has been serving as Program Chair, Organization Committee Member, Session Chair and TPC for over 120 IEEE and ACM Conferences including ACM Multimedia, SIGIR, ICCV, ICME, ICASSP, ICPR, MIR, VCIP, PCM, etc. He is the Guest Co-Editor of IEEE Transactions on Multimedia, ACM Transactions on Intelligent Systems and Technology, Journal of Computer Vision and Image Understanding, and EURASIP Journal on Advances in Signal Processing and is the associate editor of Journal of Machine Vision and Applications, the associate editor of IEEE Transaction on Circuits and Systems for Video Technology and in the Editorial Board of Journal of Multimedia. He is a Senior Member of IEEE (2003), and a Member of ACM (2004).

**Dr. Nicu Sebe** is a professor with the Faculty of Cognitive Sciences, University of Trento, Italy, where he is leading the research in the areas of multimedia information retrieval and human-computer interaction in computer vision applications. He was involved in the organization of the major conferences and workshops addressing the computer vision and human-centered aspects of computer vision and human-centered aspects of multimedia information retrieval, among which as a General Co-Chair of the IEEE Automatic Face and Gesture Recognition Conference, FG 2008, ACM International Conference on Image and Video Retrieval (CIVR) 2007 and 2010, and WIAMIS 2009 and as one of the initiators and a Program Co-Chair of the Human-Centered Multimedia track of the ACM Multimedia 2007 conference. He is the general chair of ACM Multimedia 2013 and was a program chair of ACM Multimedia 2011. He has served as the guest editor for several special issues in IEEE Transactions on Multimedia, IEEE Computer, Computer Vision and Image Understanding, Image and Vision Computing, Multimedia Systems, and ACM TOMCCAP. He has been a visiting professor in Beckman Institute, University of Illinois at Urbana-Champaign and in the Electrical Engineering Department, Darmstadt University of Technology, Germany. He is the Co-Chair of the IEEE Computer Society Task Force on Human-centered Computing and is an associate editor of IEEE Transactions on Multimedia, Machine Vision and Applications, Image and Vision Computing, Computer Vision and Image Understanding, and Journal of Multimedia.