

Guest Editorial: Large-scale 3D Multimedia Analysis and Applications

Sicheng Zhao¹ · Jun Zhang² · Yi Zhen³

Published online: 10 July 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

With the rapid advancement of networks, graphics hardware, and computing techniques, the 3D information has been widely applied in various domains, such as virtual reality and medical industry. The proliferation of such applications has produced explosively growing 3D multimedia data, which lead to the requirement for large-scale 3D data analysis. The traditional 3D feature representation and learning algorithms may not work well for large-scale analysis, such as 3D retrieval and recognition, since the computational cost in such cases is much larger. How to design efficient and effective 3D feature representation and learning techniques to deal with the 3D big data is desirable and meaningful. This special issue targets the most recent technical progress on the analysis and applications of large-scale 3D multimedia.

Submissions came from an open call for papers and with the assistance of professional referees. Eight papers are finally selected out from in total 24 submissions after two rounds of rigorous reviews. These accepted papers cover several popular topics of large-scale 3D multimedia analysis and applications, including 3D recognition, retrieval, compression, pre-processing, etc. We summarize these papers as follows:

In “3D skeleton based action recognition by video-domain translation-scale invariant mapping and multi-scale dilated CNN” [4], Li et al. propose an action recognition approach from 3D skeleton videos. By a video domain translation-scale invariant image mapping, the 3D skeleton videos are transformed to skeleton color images, which are input to a multi-scale

✉ Sicheng Zhao
schzhao@gmail.com

Jun Zhang
xdzhangjun@gmail.com

Yi Zhen
zhenyisx@gmail.com

¹ Tsinghua University, Beijing, China

² Duke University, Durham, NC, USA

³ LinkedIn, Sunnyvale, CA, USA

dilated CNN for action recognition. Further, different data augmentation strategies are designed to improve the generalization and robustness of the method. Experimental results on several popular benchmark datasets demonstrate the superiority of the approach with significant performance improvements.

The paper “Dense matching for multi-scale images by propagation” [3] presents a new dense matching algorithm between two non-stereoscopic images. There are no specific restrictions on the capture conditions. Based on the identified points of interest, the propagation process is guided according to the geometric constraints. The matching algorithm is verified on different tasks, i.e. 3D reconstruction. This dense matching method can be useful for 3D retrieval. Instead of identifying points of interest first, Nie et al. propose an end-to-end multi-scale CNN (MSCNN) for 3D model retrieval in the paper “Multi-scale CNNs for 3D model retrieval” [6]. Multiple rendered images are extracted from a 3D object, which are combined into one representative view and input to the network. The multi-layer design can effectively save local and global information for each 3D model. To compute the similarity between two different 3D models for retrieval, the Euclidean metric is employed. The retrieval and classification results on the NTU dataset show the superiority.

It is a challenging task for network service providers to deliver the multi-view plus depth (MVD) 3D videos over Internet with the best user’s Quality of Experience in dynamic network condition. In “User-perceived quality aware adaptive streaming of 3D multi-view video plus depth over the internet” [2], Karn et al. comprehensively analyze the dynamic network environment for streaming of 3D MVD over Internet. A new multi-view video streaming method is further designed to concentrate on decision strategy by user-perceived quality responsive adaptation. The simulation results show significant quality improvements under challenging network conditions. Another way to transmit and store 3D video efficiently is compression. In “Perceptual rate distortion optimization of 3D-HEVC using PSNR-HVS” [8], Valizadeh et al. propose to integrate a perceptual video quality metric inside the rate distortion optimization process of the 3D-HEVC. Based on the characteristics of the human visual system, PSNR-HVS is used as a measure for distortion in the coding unit mode selection process. The proposed method achieves 2.78% 3D-HEVC compression efficiency improvement tested on different standard multiview video sequences.

When dealing with the 3D data with complex scenes, some preprocessing techniques are required. One typical preprocessing is denoising. The paper “Feature-preserving mesh denoising based on guided normal filtering” [5] proposes a feature-preserving mesh denoising algorithm based on face classification. When denoising, the sharp features which play a key role in 3D models are kept unchanged. The multi-scale tensor voting is used to classify the faces into feature faces and non-feature faces. Based on the sub-neighborhood of the feature faces, the second-order joint bilateral filter is employed for the face normal field and vertex position update. Experiments on various synthetic datasets demonstrate the effectiveness of the proposed method. Another typical preprocessing is to subtract the background to extract the foreground objects. In “End-to-end video background subtraction with 3d convolutional neural networks” [7], Sakkos et al. propose an end-to-end temporal-aware background subtraction approach with 3D convolutional neural networks. By performing 3D convolutions on the 10 most recent frames of the video, the changes on the spatial and temporal dimension are simultaneously tracked. In this way, the

model can effectively track the movement of the foreground and the relations between neighboring pixels using multi-modal features. The experiments show a significant improvement as compared with several baseline methods.

To better understand the pathophysiological processes of ischemic heart diseases, Gao et al. present a robust motion estimation framework with an adaptive biomechanical model constraint using dual H_∞ criteria in the paper “Robust recovery of myocardial kinematics using dual H_∞ criteria” [1]. The design of two iterative H_∞ filters, i.e. the kinematics estimation filter and the elasticity estimation filter, ensures the robustness to noise changes and the flexibility to new algorithm incorporation. The results on the motion parameter estimation from MRI cardiac images demonstrate its effectiveness.

To conclude, the eight papers contained in this special issue cover several emerging topics of large-scale 3D multimedia analysis and applications. We sincerely hope that these papers can provide interesting insights and inspirations for the researchers and engineers in the related fields. We would like to thank Prof. Borko Furht for providing us the opportunity to organize this special issue. We also thank the reviewers for their generous efforts in reviewing the papers, which guarantee the high quality of accepted papers. Finally, we thank all the authors who have contributed to this special issue. Thanks to all the people who help us to make this special issue a successful one.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Gao Z, Zhang H, Wang D, Guo M, Liu H, Zhuang L, Shi P (2017) Robust recovery of myocardial kinematics using dual H_∞ criteria. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-017-5395-1>
2. Kam NK, Zhang H, Jiang F (2018) User-perceived quality aware adaptive streaming of 3d multi-view video plus depth over the internet. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-018-5744-8>
3. Laraqui M, Saaidi A, Mouhib A, Abarkan M (2018) Dense matching for multi-scale images by propagation. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-018-5644-y>
4. Li B, He M, Dai Y, Cheng X, Chen Y (2018) 3d skeleton based action recognition by video-domain translation-scale invariant mapping and multi-scale dilated cnn. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-018-5642-0>
5. Liu S, Rho S, Wang R, Jiang F (2018) Feature-preserving mesh denoising based on guided normal filtering. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-018-5735-9>
6. Nie W, Xiang S, Liu A (2018) Multi-scale cnns for 3d model retrieval. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-018-5641-1>
7. Sakkos D, Liu H, Han J, Shao L (2017) End-to-end video background subtraction with 3d convolutional neural networks. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-017-5460-9>
8. Valizadeh S, Nasiopoulos P, Ward R (2017) Perceptual rate distortion optimization of 3d-hevc using psnr-hvs. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-017-5486-z>

Sicheng Zhao received the Ph.D. degree from Harbin Institute of Technology, China, in 2016. He has been a visiting student in National University of Singapore, Singapore, from July 2013 to June 2014. He has worked as a Research Fellow in Tsinghua University, China, from August 2016 to September 2017. He is currently a Research Fellow in University of California, Berkeley, USA. His research interests include computer vision, multimedia, and affective computing. He has served / will serve as a leading guest editor of ACM TOMM, NPL, and MTAP, a co-session chair of ACM ICMR 2018, IEEE VCIP 2018, PCM 2017, and ICIG 2017, and a reviewer for over 20 prestigious international journals and conferences. He has published over 50 scientific papers in leading journals and conferences, like IEEE TMM, IEEE TAFPC, IEEE TCYB, IEEE TITS, IEEE TCSVT, ACM TOMM, ACM MM, IJCAI, AAAI, CVPR, etc. He was a co-recipient of the MMM 2015 Best Paper Runner-up Award, MS COCO 2016 Image Captioning Challenge 2nd Place Award, and IBM China Excellent Student Scholarship.

Jun Zhang is a research associate with the Department of Radiology in Duke University. He was a research associate with the Department of Radiology in University of North Carolina at Chapel Hill during 2014–2017. He received his Ph.D. degree in pattern recognition and intelligent systems from Xidian University, Shaanxi, China, 2014. His research interests include pattern recognition, optical imaging, and medical image analysis. He has published peer-reviewed journals and conferences in IEEE Transactions on Medical Imaging (TMI), IEEE Transactions on Image Processing (TIP), IEEE Transactions on Biomedical Engineering (TBME), Computer Vision and Image Understanding (CVIU), Biomedical Optical Express (BOE), Pattern Recognition Letters (PRLs), International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), IEEE International Symposium on Biomedical Imaging (ISBI), etc. His current research work is geared to medical image segmentation, landmark detection and Alzheimer's disease classification. He serves as reviewers for TIP, TBME, PRLs, CVIU, Neurocomputing, MICCAI, and Frontiers of Computer Science. He also serves as guest editors for Multimedia Tools and Applications, Multimedia Systems, and BioMed Research International. He also serves as a session chair for International Conference on Internet Multimedia Computing and Service 2016.

Yi Zhen is a Software Engineer with LinkedIn. He received his Bachelor degree from Nankai University, Master degree from Tsinghua University, and PhD degree from the Hong Kong University of Science and Technology. His research interests are machine learning, data mining, and their applications to computer vision, multimedia, and healthcare informatics. Yi Zhen served as the TPC member/reviewer for several outstanding conferences and journals. He is also the guest editor of Neurocomputing.