



## Editorial

## Learning for 3D understanding



3D data has been widely applied in various domains, such as 3D graphics, entertainment and the architecture design, due to the advances in computing techniques, graphics hardware, and networks. Along with the proliferation of aforementioned applications, it is emerging for the effective 3D processing technique to manipulate such massive scale 3D data corpus, bringing novel and significant challenges to the state-of-the-art 3D processing techniques. In many recent 3D processing systems, machine learning has been widely investigated and regarded as one of the most fundamental components in next-generation 3D processing techniques. We believe it is the time to organize a special issue on such a topic. The primary objective of this special issue fosters focused attention on the latest research progress in this interesting area. In this special issue, we target at introducing cutting-edge learning algorithms and techniques for 3D object and scene understanding, such as 3D semantic analysis, 3D object retrieval and recognition and other applications.

Submissions came from an open call for paper and with the assistance of professional referees, 20 papers are finally selected out from in total 39 submissions after rigorous reviews. These papers cover several popular subtopics of learning methods for 3D understanding, including 3D action recognition and retrieval, 3D object recognition and retrieval, etc. We summarize papers into four subtopics according to the themes of the papers.

The first part contains 5 papers that are related to 3D action recognition and retrieval. The first paper, “*Strategy for Dynamic 3D Depth Data Matching Towards Robust Action Retrieval*”, introduces a generalized strategy to match 3D depth data dynamically for action retrieval. In this method, a 3D shape context descriptor is used to extract static depth frame features. Then the temporal similarity between two 3D sequences is measured by dynamic time warping. The second paper, “*Single/Multi-view Human Action Recognition via Regularized Multi-Task Learning*”, introduces a pyramid part-wise bag of words representation to implicitly encode both local visual characteristics and human body structure and the single/multi-view human action recognition is conducted under a multi-task learning framework to discover the latent correlation between multiple views and body parts. In the third paper “*Multi-perspective and Multi-modality Joint Representation and Recognition Model for 3D Action Recognition*”, the 3D action recognition task is exploited. Gao et al. propose a multi-perspective and multi-modality discriminated and joint representation and recognition model. In the fourth paper “*Feature Learning based on SAE-PCA Network for Human Gesture Recognition in RGBD images*”, Li et al. propose to recognize 3D human actions via sparse auto-encoder and PCA based feature learning method. In this method, both the RGB channel and the depth channel are used

to learn features via sparse auto-encoder with convolutional neural networks. Then these features are concatenated in a multiple layer PCA procedure to conduct human action recognition. The fifth paper, “*Hand Fine-Motion Recognition based on 3D Mesh MoSIFT Feature Descriptor*”, addresses the hand fine-motion recognition task. In this method, the hand location is first located using the improved graph cuts method. With 3D geometric characteristics and hand behavior prior information, Ming et al. propose a 3D mesh MoSIFT feature descriptor for hand activity representation. The features are extracted by using simulation orthogonal matching pursuit.

The second part contains 5 papers that are related to 3D object retrieval and recognition. The paper “*Locality-constrained Sparse Patch Coding for 3D Shape Retrieval*” proposes to employ low-level patches of 3D shapes, similar to the superpixels of images, for 3D object retrieval. In the paper “*A 3D Model Recognition Mechanism based on Deep Boltzmann Machines*”, Leng et al. employs the deep Boltzmann machines to represent the distributions of multiple depth images of 3D data towards 3D model recognition. In this way, the feature of 3D models is generated via deep learning and a graph-based semi-supervised learning procedure is used for 3D model recognition. In the paper “*Supervised Feature Learning via  $l_2$  norm Regularized Logistic Regression for 3D Object Recognition*”, Zou et al. exploit a group of classifiers to construct feature extraction method. This method enjoys the benefits from the classifier-based feature extraction method which employs the label information and can be more discriminative. In the paper “*Hypergraph based Feature Fusion for 3-D Object Retrieval*”, Wang et al. propose a hypergraph-based feature fusion method for view-based 3D object retrieval. The paper “*3D Model Retrieval with Weighted Locality-constrained Group Sparse Coding*” introduces a weighted locality-constrained group sparse coding method for 3D model retrieval.

The third part contains 6 papers that are related to depth reconstruction and 3D segmentation. In the paper “*Estimation of Human Body Shape and Cloth Field In Front of a Kinect*”, Zeng et al. introduce an easy-to-use system for human body and clothes shape estimation, which employs a Kinect to capture the human's RGB and depth information from different views. The paper “*Graph-Based Learning for Segmentation of 3D Ultrasound Images*” focuses on the segmentation method to extract objects of interest from 3D ultrasound images and the paper “*Automatic Stereoscopic Video Generation based on Virtual View Synthesis*” concentrates on automatically and robustly synthesizing stereoscopic videos from casual 2D monocular videos. In the paper “*Depth Map Reconstruction and Rectification Through Coding Parameters for Mobile 3D Video System*”, Yang et al. focus their work on dense depth map generation with high quality and high resolution, which is important but challenge

for all 3D video and vision applications due to the limitation of depth sensing systems. This work shows its great potential in high quality portable 3D applications. In the paper “3D Face Reconstruction from Skull by Regression Modeling in Shape Parameter Spaces”, Duan et al. propose a regression modeling method for craniofacial reconstruction, in which for each skull and face, a statistical shape model is built respectively.

The fourth part focuses on other applications, which consists of 4 papers. In the paper “Natural Image Statistics based 3D Reduced Reference Image Quality Assessment in Contourlet Domain”, Wang et al. introduce a novel 3D reduced reference image quality assessment metric based on 3D natural image statistics in Contourlet domain. The paper “3D Orbit Selection for Regional Observation GEO SAR” considers the application in remote sensing, which introduces a 3D orbit selection method for regional observation GEO SAR. The paper “Integrating 3D Structure into Traffic Scene Understanding with RGB-D Data” focuses on the application in traffic scene understanding and the paper “Online Learning 3D Context for Robust Visual Tracking” introduces a 3D context-based object tracking method by learning a set of 3D context key-points.

To conclude, papers in this special issue cover several emerging topics of how learning based methods can benefit the existing 3D processing and understanding techniques. We highly hope this special issue can attract concentrated attentions in the related field to benefit and inspire researchers and practitioners in this emerging topic. We would like to thank Prof. Tom Heskes for

providing the chance to organize this special issue. We also thank the reviewers for their efforts to guarantee the high quality of this special issue. Finally, we thank all the authors who have contributed to this special issue.

Yue Gao\*

Tsinghua University, China

E-mail address: kevin.gaoy@gmail.com

Rongrong Ji

Xiamen University, China

Wei Liu

IBM T. J. Watson Research Center, USA

Qionghai Dai

Tsinghua University, China

Received 13 October 2014; accepted 19 October 2014

Available online 27 October 2014

---

\* Corresponding author at: Main building 724, Tsinghua University, Beijing, China, 100086.