Guest Editorial

Advanced Understanding and Modelling of Human Motion in Multidimensional Spaces

Hui Yu· Junyu Dong · Tuan D. Pham · Honghai Liu

Special Issue – Multimedia Tools and Applications – Springer-Verlag 2015

Abstract With increasing applications such as human-computer interaction, surveillance, medical diagnosis and computer animation, automatic understanding and modelling of human motion have attracted more attention from researchers and industries. Extensive research in human motions especially facial behavior and human body actions has been explored in recent years. The bulk of research for human motion understanding concentrates on using low-level machine-understanding features. Despite the huge amount of excellent research in these fields, the effective and efficient description and representation of human motions remain challenging in many scenarios. Therefore, the aim of this special issue is to survey the-state-of-the-art methodologies, algorithms and concepts in advanced human motion understanding with any kind of data types. It intends to bridge the gap between the low-level features and high-level semantics of human motion.

H. Yu (✉)
University of Portsmouth, UK
E-mail: hui.yu@port.ac.uk

J. Y. Dong
Ocean University of China
Email: dongjunyu@ouc.edu.cn

T. D. Pham
Linköping University, Sweden
Email: tuan.pham@liu.se

H. Liu
University of Portsmouth, UK
Email: honghai.liu@port.ac.uk

1 Introduction

Motion is intimately tied with our behavior. For example, we communicate through facial expressions and gestures etc. to interact with each other in our daily lives. The understanding and modelling of human motion has been a subject of interest in the scientific community for more than one century. The long history of human motion analysis comes from the large scope of applications of such measurement that can be found in medicine, biomechanics, sport, ergonomics, and so forth. It is getting increasingly important with the development of computing technologies and applications. More recently, those technologies have also been widely exploited for the development of human-machine interaction, surveillance, medical diagnosis and computer animation, social interaction, and nonverbal communication.

There is extensive research in human motions especially facial behavior and human body actions in recent years. Great efforts have been made by researchers from various areas including computer scientists and phycologists etc. Complex methods have been proposed for human motion understanding based on low-level machine-understanding features. Approaches of human motions understanding normally use motion primitives based on the statistical evaluation of motion clusters or dynamic models or discriminative methods. Moreover, the fast-increasing volume of multimedia data has emerged in recent years. The data related to human motions includes different kinds of sensory data such as visual (including 2D, 3D and RGB data), audial, Electroencephalography (EEG) and touch sensory data etc. The multimodal data provides overwhelming channels for advanced human motion understanding.

As a matter of fact, human beings are able to response effectively to emotional state in social interaction from subtle human motions. Therefore, it is still challenging to effectively and efficiently describe and represent human motions. We dedicate this special issue to these challenges as a venue to advance and disseminate the most recent research on this theme. The aim of this special issue is to survey a state of art of methodologies, algorithms and concepts in advanced human motion understanding with any kind of data types. It intends to bridge a gap between the low-level features and high-level semantics of human motions.

We have received thirty three manuscripts with each manuscript being blindly reviewed by at least three reviewers consisting of guest editors and anonymous reviewers. After the review processes, twenty two manuscripts were finally selected to be included in this special issue. Those accepted papers represent a wide spectrum of research under the theme of the special issue ranging from human action, facial expression, to gaze estimation. In Section 2, we present a brief summation of selected papers for advanced understanding and modelling of human motions in this special issue.

2 Summary of accepted papers

This special issue has accepted twenty two carefully selected papers. Here is a brief summary of those papers in this section.

Falls at homes are one of the major risks for elderly and an immediate alarming and helping is essential to reduce the rate of morbidity and mortality. Therefore, human fall detection is necessary and useful in smart home system and health care system. The paper "Human Fall Detection in Surveillance Video based on PCANet" authored by Shengke Wang et al. [1] highlights the importance of human motion understanding in the healthcare environment. It proposes a new framework for fall detection based on automatically feature learning methods. The authors train the samples by PCANet, which is a deep learning approach to predict the label of each frame that is further used to train an action model with SVM. They report that the proposed method can achieve a reliable result comparing with the state-of-the-art methods.

Recognition in a video stream of air-drawn gestures and characters is an important technology in realizing verbal and nonverbal communication in human-computer interaction. The paper "Time-Segmentation and Position-Free Recognition of Air-Drawn Gestures and Characters in Videos" is submitted by Yuki Niitsuma et al [2]. The paper presents an algorithm called time-space continuous dynamic programming (TSCDP) to realize both time- and location-free (spotting) recognition in video streams of isolated alphabetic characters and connected cursive textual characters. The authors test the TSCDP algorithm in recognizing 26 isolated alphabetic characters and 23 Japanese hiragana and kanji air-drawn characters. Moreover, the authors report that the TSCDP algorithm can also be well performed in gesture recognition.

There are seven papers related to human actions or activities are included in this issue. The paper "Exploiting Stereoscopic Disparity for Augmenting Human Activity Recognition Performance" from Ioannis Mademlis at al. [3] investigates several ways to exploit scene depth information, implicitly available through the modality of stereoscopic disparity in 3D videos, with the purpose of augmenting performance in the problem of recognizing complex human activities in natural settings. The investigated approaches are flexible and able to cooperate with any monocular low-level feature descriptor. Qi Jia et al. propose the temporal characteristic number (TCN) for individual joint point of a human body in temporal series [4]. In [5], Shaomei Li et al treat the motion primitive learning as a temporal problem and propose a multi-layered multi-exemplar affinity propagation (MLMEAP) method. Yidong Li et al. [6] develop a learning-based comprehensive evaluation model for traffic data quality (TDQ) in intelligent transportation systems (ITS). Andrés Adolfo Navarro-Newball et al. [7] present a novel approach for recreating life-like experiences through an easy and natural gesture-based interaction. In [8], Xin Wang et al. present a method for the efficient retrieval and browsing of immense amounts of realistic 3D human body motion capture data. Kang Wang et al. [9] propose a base-points-driven shape correspondence (BSC) approach to extract skeletons of articulated objects from 3D mesh shapes.

Eye and its movement as a part of the most significant facial features, plan an important role in reflecting a person's mental and emotional state. The paper titled "Comparison of Random Forest, Random Ferns and Support Vector Machine for Eye State Classification" is authored by Yanchao Dong et al [10]. The paper presents an eye state estimation framework with various feature sets using random forest\ferns based on its superior performance on time consumption. The comparison of different classifiers indicates that random forest\ferns outperform the SVM in terms of time consumption. The authors incorporate HOG feature into the classifier to reduce noise influence. The correctly recognition rate reported in this paper is above 93%.

Remote gaze estimation under natural light with free head movements is still a challenging problem. The paper "Remote Gaze Estimation Based on 3D Face Structure and Iris Centers Under Natural Light" authored by Chunshui Xiong et al. [11] proposes a novel feature-based gaze estimation method without use of cornea reflections. The authors utilize 3D active shape models to represent 3D face structure and present a 3D Iris-Eye-Contours descriptor to represent human gaze information. They improve the ability of tolerance to head movements with rectified 3D iris centers and eye contours obtained by the head poses. The proposed gaze estimation system is tested on several subjects and the results demonstrate that the system can achieve low estimation error and allow natural head movements under natural light.

Face tracking often encounters drifting problems, especially when a significant face appearance variation occurs. The paper "A Fusion Method for Robust Face Tracking" authored by Xiaodong Jiang at al. [12] proposes a novel and efficient fusion strategy for robust face tracking. A Supervised Descent Method (SDM) and a Compressive Tracking method (CT) are employed at the same time. The novel on-line fusion method can remarkably alleviate the drifting problem in robust face tracking. Another paper related to face features is "Positioning corners of human mouth based on local gradient operator" authored by Yulin Wang at al [13]. This paper presents a method to detect and locate accurately facial feature points based on local gradient operator.

The paper titled "Ensemble based Extreme Learning Machine for Cross-modality Face Matching" authored by Yi Jin et al. [14] proposes a new ensemble-based extreme learning machine (ELM) approach for cross-modality face matching, which integrates the voting ELM with a discriminant feature descriptor. ELM is one of the most important and efficient machine learning algorithms for pattern classification due to its fast learning speed. The authors demonstrate the effectiveness of the proposed approach on two heterogeneous face recognition scenarios.

View-invariant human action recognition is a challenging research topic in computer vision. The paper "Multi-view Transition HMMs based View-invariant Human Action Recognition Method" authored by Xiaofei Ji at al. [15] presents a novel graphical structure based on Hidden Markov Models with multi-view transition to model the human action with viewpoint changing. The proposed model can not only simplify the model training process by decomposing the parameter space into multiple sub-spaces, but also improve the performance of the algorithm by constraining the possible viewpoint changing.

The paper “Cultural-based Visual Expression: Emotional Analysis of Human Face via Peking Opera Painted Faces (POPF)” authored by Ding Wang et al. [16] takes the study of Peking Opera Painted Faces (POPF) as an example to see how information and meanings can be effectively expressed through the change of facial expressions based on the facial motion within natural and emotional aspects. The paper titled “Automatic Evaluation of The Degree of Facial Nerve Paralysis” authored by Ting Wang et al. [17] presents a novel method for evaluating the degree of facial paralysis considering both static facial asymmetry and dynamic transformation factors. A quantitative approach of static facial asymmetry based on local mirror asymmetry is proposed.

Hand postures provide attractive means of interface devices for human-computer interaction. In the paper “Hand posture recognition based on heterogeneous features fusion of multiple kernels learning” authored by Jiangtao Cao et al. [18], a novel hand posture recognition method is proposed by integrating the multiple image features and multiple kernels learning support vector machine(SVM). The paper “A Novel Approach to Extract Hand Gesture Feature in Depth Images” authored by Zhaojie Ju et al. [19] proposes a novel approach to extract human hand gesture features in real-time from RGB-D images based on the earth mover’s distance and Lasso algorithms.

The paper “Small Scale Crowd Behavior Classification by Euclidean Distance Variation-Weighted Network” authored by Xuguang Zhang at al. [20] introduces microscopic-based method and macroscopic-based method in crowd behavior analysis. By exploring the connection between the microscopic and macroscopic properties of a crowd, a method which use Euclidean distance variation-weighted network to recognize the crowd behavior is proposed in this paper.

The paper “Video Parsing via Spatiotemporally Analysis with Images” authored by Xuelong Li at al. [21] proposes to transfer or propagate such labels from images to videos. The proposed approach consists of three main stages: I) the posterior category *probability density function* (PDF) is learned by an algorithm which combines frame relevance and label propagation from images. II) the prior contextual constraint PDF on the map of pixel categories through whole video is learned by the Markov Random Fields (MRF). III) finally, based on both learned PDFs, the final parsing results are yielded up to the maximum a posterior (MAP) process which is computed via a very efficient graph-cut based integer optimization algorithm.

Finally, the paper “Age Estimation Based on Improved Discriminative Gaussian Process Latent Variable Model” authored by Lijun Cai at al. [22] proposes a novel age estimation method based on improved discriminative Gaussian process latent variable model (DGPLVM) to discover the underlying trend of aging patterns. To consider age estimation as a complex and nonlinear problem, the authors employ the improved DGPLVM to obtain low-dimensional representations and use the Gaussian process regression model to find the age regressor that maps the low-dimensional representations to ages. They conduct experiments on two widely used databases FG-NET and MORPH, and the results show that the proposed improved DGPLVM age estimation method is effective and comparable to the state-of-the-art methods.

3 Conclusion

The papers included in this special issue are representative of the current research challenges in advanced understanding and modeling of human motions. It is expected that these papers can provide researchers with valuable resources and motivations to work on the challenging issues in this research theme.

We would like to thank the Editor-in-Chief of the journal, Prof. Borko Furht for his huge support for this issue. Our special thanks go to all editorial staff, especially Mr. Jay-Y Banua and Ms. Courtney Clark for their valuable and prompt support throughout the preparation and publication of this special issue. We would like to thank all authors for their contributions to this special issue. We also extend our thanks to all reviewers for their hard work to ensure the high quality of accepted papers.

References

1. Shengke Wang, Long Chen, Zixi Zhou, Junyu Dong (2015) "Human Fall Detection in Surveillance Video based on PCANet," Multimedia Tools and Applications (MTAP), Online Published, June 12 2015. (doi: 10.1007/s11042-015-2698-y)
2. Yuki Niitsuma, Syunpei Torii, Yuichi Yaguchi, Ryuichi Oka (2015) "Time-segmentation and position-free recognition of air-drawn gestures and characters in videos," Multimedia Tools and Applications (MTAP), Online Published, May 22 2015. (doi: 10.1007/s11042-015-2669-3)
3. Ioannis Mademlis, Alexandros Iosifidis, Anastasios Tefas, Nikos Nikolaidis, Ioannis Pitas (2015) "Exploiting Stereoscopic Disparity for Augmenting Human Activity Recognition Performance," Multimedia Tools and Applications (MTAP), Online Published, July 04 2015. (doi: 10.1007/s11042-015-2719-x)
4. Qi Jia, Xin Fan, Zhongxuan Luo, Haojie Li, Kang Huyan, Zezhou Li (2015) "Cross-view action matching using a novel projective invariant on non-coplanar space-time points," Multimedia Tools and Applications (MTAP), Online Published, June 10 2015. (doi: 10.1007/s11042-015-2704-4)
5. Shao-Mei Li, Hui Yu, Ya-Wen Wang, Chao Gao (2015) "Multi-layered Multi-Exemplar Affinity Propagation for Temporal Clustering of Human Motion," Multimedia Tools and Applications (MTAP), Online Published, April 18 2015. (doi: 10.1007/s11042-015-2608-3)
6. Yidong Li, Dewang Chen (2015) "A Learning-based Comprehensive Evaluation Model For Traffic Data Quality In Intelligent Transportation Systems," Multimedia Tools and Applications (MTAP), Online Published, June 11 2015. (doi: 10.1007/s11042-015-2676-4)
7. Andrés Adolfo Navarro-Newball, Isidro Moreno, Edmond Prakash, Ali Arya, Victoria E. Contreras, Victor A. Quiceno, Santiago Lozano, Juan David Mejía, Diego Fernando Loaiza (2015) "Gesture based human motion and game principles to aid understanding of science and cultural practices," Multimedia Tools and Applications (MTAP), Online Published, May 31 2015. (doi: 10.1007/s11042-015-2667-5)
8. Xin Wang, Liangxiu Chen, Jiali Jing, Herong Zheng (2015) "Human motion capture data retrieval based on semantic thumbnail," Multimedia Tools and Applications (MTAP), Online Published, May 29 2015. (doi: 10.1007/s11042-015-2705-3)
9. Kang Wang, Abdul Razzaq, Zhongke Wu, Feng Tian, Sajid Ali, Taorui Jia, Xingce Wang, Mingquan Zhou (2015) "Novel Correspondence-based Approach for Consistent Human Skeleton Extraction," Multimedia Tools and Applications (MTAP), Online Published, April 30 2015. (doi: 10.1007/s11042-015-2629-y)
10. Yanchao Dong, Yan Zhang, Jiguang Yue, Zhencheng Hu (2015) "Comparison of Random Forest, Random Ferns and Support Vector Machine for Eye State Classification," Multimedia

Tools and Applications (MTAP), Online Published, May 07 2015. (doi: 10.1007/s11042-015-2635-0)

11. Chunshui Xiong, Lei Huang, Changping Liu (2015) "Remote gaze estimation based on 3D face structure and iris centers under natural light," Multimedia Tools and Applications (MTAP), Online Published, April 24 2015. (doi: 10.1007/s11042-015-2600-y)
12. Xiaodong Jiang, Hui Yu, Yang Lu, Honghai Liu (2015) "A Fusion Method for Robust Face Tracking," Multimedia Tools and Applications (MTAP), Online Published, May 03 2015. (doi: 10.1007/s11042-015-2659-5)
13. Yulin Wang, Wenjia Ding, Yixin Chen (2015) "Positioning corners of human mouth based on local gradient operator," Multimedia Tools and Applications (MTAP), Online Published, April 22 2015. (doi: 10.1007/s11042-015-2627-0)
14. Yi Jin, Jiuwen Cao, Yizhi Wang, Ruicong Zhi (2015) "Ensemble based Extreme Learning Machine for Cross-modality Face Matching," Multimedia Tools and Applications (MTAP), Online Published, May 10 2015. (doi: 10.1007/s11042-015-2650-1)
15. Xiaofei Ji, and Zhaojie Ju, Ce Wang, Changhui Wang (2015) "Multi-view Transition HMMs based View-invariant Human Action Recognition Method," Multimedia Tools and Applications (MTAP), Online Published, May 10 2015. (doi: 10.1007/s11042-015-2661-y)
16. Ding Wang, Jinsheng Kang, Sheng-Feng Qin, Johannes Birringer (2015) "Cultural-based Visual Expression: Emotional Analysis of Human Face via Peking Opera Painted Faces (POPF)," Multimedia Tools and Applications (MTAP), Online Published, May 15 2015. (doi: 10.1007/s11042-015-2665-7)
17. Ting Wang, Shu Zhang, Junyu Dong, Li'an Liu, Hui Yu (2015) "Automatic evaluation of the degree of facial nerve paralysis," Multimedia Tools and Applications (MTAP), Online Published, June 05 2015. (doi: 10.1007/s11042-015-2696-0)
18. Jiangtao Cao, Siquan Yu, Honghai Liu, Ping Li (2015) "Hand posture recognition based on heterogeneous features fusion of multiple kernels learning," Multimedia Tools and Applications (MTAP), Online Published, May 14 2015. (doi: 10.1007/s11042-015-2628-z)
19. Zhaojie Ju, Dongxu Gao, Jiangtao Cao, Honghai Liu (2015) "A novel approach to extract hand gesture feature in depth images," Multimedia Tools and Applications (MTAP), Online Published, April 24 2015. (doi: 10.1007/s11042-015-2609-2)
20. Xuguang Zhang, Meiling Ouyang, Xufeng Zhang (2015) "Small scale crowd behavior classification by Euclidean distance variation-weighted network," Multimedia Tools and Applications (MTAP), Online Published, May 21 2015. (doi: 10.1007/s11042-015-2670-x)
21. Xuelong Li, Lichao Mou, Xiaoqiang Lu (2015) "Video parsing via spatiotemporally analysis with images," Multimedia Tools and Applications (MTAP), Online Published, July 07 2015. (doi: 10.1007/s11042-015-2735-x)
22. Lijun Cai, Lei Huang, Changping Liu (2015) "Age estimation based on improved discriminative Gaussian process latent variable model," Multimedia Tools and Applications (MTAP), Online Published, May 21 2015. (doi: 10.1007/s11042-015-2668-4)