BAVC: Human Computer Interaction and Behavior Analysis Using Vision Cameras Editorial

Special Track along with ACHI 2020
The 13th International Conference on Advances in Computer-Human Interactions
November 21, 2020 to November 25, 2020 - Valencia, Spain
http://www.iaria.org/conferences2020/ACHI20.html

Oky Dicky Ardiansyah Prima
Iwate Prefectural University
Department of Software and Information Science
152-52 Sugo Takizawa Iwate 020-0693, Japan.
email: prima@iwate-pu.ac.jp

Abstract— **Recent advances in computer vision technology have made it possible to facilitate the automatic detection of facial and body features from images. Real-time analysis of facial features allows us to observe changes in facial emotions and facial muscle movements, which are potentially important for behavioral analysis. Similarly, the analysis of body movement features enables us to easily facilitate human-computer interaction without the use of motion capture. In this special track, four presentations were given, including estimation of engagement using facial features, analysis of expression similarity using 3D facial models, communication mirroring detection from body movements, and a stylus pen with six degrees of freedom for 3D interaction.**

Keywords-behavior analysis;engagement estimation; expression similarity; communication mirroring; 3D interaction.

I. INTRODUCTION

In the last decade, the rapid development of machine learning and artificial intelligence has allowed computer vision to perform various analyses from images and videos. Several methods for analyzing faces and bodies have been proposed, including face detection, facial landmark detection, and human pose estimation.

The detection of faces has shifted from classical Haar-like to Histogram of Oriented Gradients (HOG) [1] features and then to methods using Convolutional Neural Networks (CNN) [2], which provide more robust detection results. The detection of facial landmarks has evolved from Active Shape Models (ASMs) [3] to Ensemble of Regression Trees (ERT) [4] and then to CNN-based methods [5], which allow us to detect features more efficiently and in greater detail. The head pose can be estimated by fitting the coordinates of face landmarks to a 3-Dimensional (3D) face model.

In the field of human pose estimation, the introduction of the OpenPose library [6] has provided the ability to detect 2-Dimensional (2D) pose of multi-person in an image. Moreover, methods for estimating 3D human poses [7] from the 2D human poses have been developed, which enables us to obtain 3D human poses from a single camera, comparable to the information obtained from motion capture.

While computer vision can be used to acquire various facial and body features, the uses of these features in behavioral analysis and Human Computer Interface (HCI) are still in the developmental stages. This special session invites the researchers working in these fields to introduce and discuss cases of study that are applicable to the real life. Some of the most relevant topics relevant to the objectives of the special track include: facial analysis, body motion analysis, behavior analysis, interactive human computer interaction, flexibility training, human tracking, and other applications such as healthcare, assistive-robotics, etc.

The rest of this editorial is organized as follows: the following Section II summarizes the submissions accepted for presentation and publication in the special track. Section III concludes and presents future perspectives and challenges for this topic.

II. SUBMISSIONS

The first paper entitled “Engagement Estimation for an E-Learning Environment Application” by Khine et al. [8] took advantage of style transfer techniques to obtain the basic facial features and remove other features that are not useful for estimating engagement through the virtual environment. Based on the high-dimensional data from the 6th fully connected layer of the VGG-16 model compressed with t-Distributed Stochastic Neighbor Embedding (t-SNE), the fully expressive (peak) and neutral facial expressions were successfully classified for each frame using k-means. This paper also found individual differences in the peak-like and neutral-like frames. In the future, the proposed technique will be enhanced to handle more levels of facial expressions on online learning engagement.

The next paper entitled “Facial Mimicry Training Based on 3D Morphable Face Models” by Prima et al. [9] proposed a self-learning-based expression training system by evaluating the similarity of facial images based on the shape derived from the 3D Morphable Face Model (3DMM). The proposed

system automatically annotated 68 locations of 2D facial landmarks from facial images using the Dlib library[10]. These landmarks were associated with 3D facial landmarks to compute the pose of the head. Experiment results showed that the facial appearance between subjects and their corresponding mimic targets were highly correlated.

The third paper entitled "A Perspective-Corrected Stylus Pen for 3D Interaction" by Takahashi et al. [11] proposed a novel perspective-corrected stylus pen that can be used for 3D interaction with a 3D spherical display. The camera is set up at the bottom of the display to detect the contact point between the pointing device and the display surface, whereas the pointing device gets its 3D orientation from the smartphone's built-in Inertial Measurement Unit (IMU). The proposed pointing device enables the user to see the auxiliary line from the device tip from different angles. Some 3D pointing applications, such as selecting objects inside the display were demonstrated to show the usability of the proposed pointing device.

The last paper entitled "Toward Automated Analysis of Communication Mirroring" by Hosogoe et al. [12] proposed a framework to perform a time series analysis based on Dynamic Time Warping (DTW) to determine whether communication mirroring has been established. The framework uses human pose estimation techniques to track hand gestures of two persons during a conversation. Experiment results found a difference between similar behaviors perceived by humans and similar behaviors seen from sensor information. To successfully identified distinct gestures using DTW without being too obsessed with small movements, the information of the hand gestures was compressed into one-dimensional using the L2 norm.

III. Conclusion & Future Perspectives

The special track "BAVC: Human Computer Interaction and Behavior Analysis Using Vision Cameras" provided a forum for discussing research topics on human-computer interaction and human behavioral analysis using computer vision, as well as issues related to each topic and their solutions. The four papers presented here dealt with the analysis of facial features, gesture analysis, and sensing of position and rotation information; however, we expect to receive more submissions of applied research in the future. Future perspectives regarding the topic include questions concerning whether an integrated analysis of facial image information and detailed facial landmarks is necessary in the observation of changes in certain facial expressions, and whether all the geometric information (skeletal information) of the body is required in the evaluation of resembled body movements.

The upcoming BAVC plans to call for case studies using eye movement information in addition to face and body information to discuss a wide range of topics related to human behavior analysis.

Acknowledgment

We would like to thank the organizers of ACHI2020 for accepting BAVC as a special track and for their hard work and support during the preparation. We would also like to thank the authors for their efforts on the special track.

References

[1] A. Adouani, W. M. Ben Henia, and Z. Lachiri, "Comparison of Haar-like, HOG and LBP Approaches for Face Detection in Video Sequences," 2019 16th International Multi-Conference on Systems, Signals & Devices (SSD), Istanbul, Turkey, 2019, pp. 266-271, doi: 10.1109/SSD.2019.8893214.

[2] A. Ponnusamy, "cvlib - High Level Computer Vision Library for Python," 2018. https://github.com/arunponnusamy/cvlib [retrieved: October 31, 2020]

[3] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models - Their Training and Application," Computer Vision and Image Understanding 61(1), pp. 38–59, 1995.

[4] V. Kazemi and J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 1867-1874, doi: 10.1109/CVPR.2014.241.

[5] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-stage Dense Face Localisation in the Wild," .arXiv:1905.00641, 2019.

[6] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," Computer Vision and Pattern Recognition, pp.7291-7299, 2017.

[7] J. Martinez, R. Hossain, J. Romero, and J. J. Little, "A Simple Yet Effective Baseline for 3d Human Pose Estimation," In IEEE International Conference on Computer Vision, ICCV, 2017.

[8] W. S. S. Khine, S. Hasegawa, and K. Kotani, "Engagement Estimation for an E-Learning Environment Application," The 13th International Conference on Advances in Computer-Human Interactions (ACHI2020), Special Track on Human Computer Interaction and Behavior Analysis Using Vision Cameras, Valencia, Spain, 2020.

[9] O. D. A. Prima, H. Ito, T. Tomizawa, and T. Imabuchi, "Facial Mimicry Training Based on 3D Morphable Face Models," The 13th International Conference on Advances in Computer-Human Interactions (ACHI2020), Special Track on Human Computer Interaction and Behavior Analysis Using Vision Cameras, Valencia, Spain, 2020.

[10] Dlib C++ Library, http://dlib.net. [retrieved: October 31, 2020]

[11] R. Takahashi, K. Hotta, O. D. A. Prima, and H. Ito, "A Perspective-Corrected Stylus Pen for 3D Interaction," The 13th International Conference on Advances in Computer-Human Interactions (ACHI2020), Special Track on Human Computer Interaction and Behavior Analysis Using Vision Cameras, Valencia, Spain, 2020.

[12] K. Hosogoe, M. Nakano, O. D. A. Prima, and Y. Ono, "Toward Automated Analysis of Communication Mirroring," The 13th International Conference on Advances in Computer-Human Interactions (ACHI2020), Special Track on Human Computer Interaction and Behavior Analysis Using Vision Cameras, Valencia, Spain, 2020.