Pacific Graphics (2023)
R. Chaine, Z. Deng, and M. H. Kim (Editors)

Poster

Avatar Emotion Recognition using Non-verbal Communication

Jalal Safari Bazargani, Abolghasem Sadeghi-Niaraki and Soo-Mi Choi†

Department of Computer Science and Engineering and Convergence Engineering for Intelligent Drone, XR Research Center, Sejong University, Korea

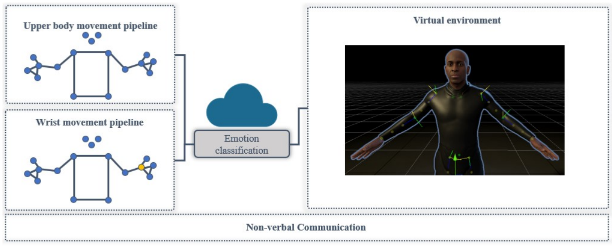

Figure 1: *Overall framework of avatar emotion recognition in the virtual environment.*

Abstract

Among the sources of information about emotions, body movements, recognized as "kinesics" in non-verbal communication, have received limited attention. This research gap suggests the need to investigate suitable body movement-based approaches for making communication in virtual environments more realistic. Therefore, this study proposes an automated emotion recognition approach suitable for use in virtual environments. This study consists of two pipelines for emotion recognition. For the first pipeline, i.e., upper-body keypoint-based recognition, the HEROES video dataset was employed to train a bidirectional long short-term memory model using upper-body keypoints capable of predicting four discrete emotions: boredom, disgust, happiness, and interest, achieving an accuracy of 84%. For the second pipeline, i.e., wrist-movement-based recognition, a random forest model was trained based on 17 features computed from acceleration data of wrist movements along each axis. The model achieved an accuracy of 63% in distinguishing three discrete emotions: sadness, neutrality, and happiness. The findings suggest that the proposed approach is a noticeable step toward automated emotion recognition, without using any additional sensors other than the head mounted display (HMD).

CCS Concepts

• ***Human-centered computing*** $\rightarrow$ ***Human computer interaction (HCI);***

1. Introduction

Traditionally, it was believed that working with computers efficiently required discarding emotions. However, in the human–computer interaction field, effective cooperation is achieved if the computer component is provided with some knowledge of the emotional state of the user; this makes emotion recognition paramount. With regard to body movements, it has been demonstrated that a person's gait or body expressions can influence how others perceive their feelings. In addition, body movements can be considered a better approach when emotion recognition is required from a distance. That said, this area of research has not been explored extensively; [SKPA19]. In terms of wrist movement, the analysis of physiological data does not rely on recordings in a laboratory setting with limited ecological validity [QGY18].

The research gap suggests a possible approach for obtaining information about avatar emotions that is not computationally costly and does not require additional sensors. Currently, most avatars are upper-body avatars, and facial expressions are not offered pervasively in all devices or networks in the metaverse. To the best of our knowledge, no similar study has been found targeting emotion recognition based on avatar upper-body movements in the metaverse, including upper-body keypoints and acceleration of wrist movement while walking. This study recognizes four discrete emotions based on an avatar's upper-body keypoints and three emotions based on wrist movements in a virtual environment. With the help of the meta-movement SDK and XR Interaction SDK, explicit fea-

† Coressponding author

DOI: 10.2312/pg.20231277

ture extraction to provide inputs was omitted which adds to the contribution of this paper.

2. Methodology

The overall framework of the study is indicated in Figure 1. For the upper-body keypoints, the HEROES video dataset was chosen for the experiment, presenting four emotions: boredom, disgust, happiness, and interest. As body language includes different indicators such as body posture, gestures, and eye movements [NCK*18], 17 upper-body keypoints were selected for each frame. The left eye, right eye, and nose were selected for eye movement and head orientation. After the face, arms are believed to be a great source of body language information [MGKK15]. In this regard, five keypoints including the elbow, wrist, thumb, index finger, and little finger were chosen for each arm. The remaining four keypoints are indicators of the shoulders and hips, i.e., the torso.

The dataset provided by [QGY18] was used for the wrist-movement pipeline. Three discrete emotions: happy, sad, and neutral, were offered in this dataset. For this study, only the accelerometer data were employed because after the initial tests, retrieving the gyroscope data from the controllers seemed unreliable. Seventeen features were extracted from the accelerometer data along each axis, resulting in 51 features per sample. The processed data was then regarded as input features for training a random forest model. The remaining procedure is similar to that used in upper-body pipeline, wherein the LSTM model is connected to the virtual scene with the help of a web socket. Inside the virtual scene, the acceleration was calculated based on the velocity of the left controller, assuming that the data came from a smartwatch. This approach holds some benefits, including eliminating the need for a smartwatch and an interface to connect the data obtained from the smartwatch to the engine.

3. Results and Discussion

For the upper-body pipeline, real-time tests were conducted without any noticeable delay, primarily because the feature extraction part was omitted and SDK was utilized to retrieve keypoints. This approach also prevents challenges arising from the distance between the point responsible for visual feature extraction and avatar location. In terms of emotions, for boredom, the associated body language normally involves having the hands parallel to the torso or palms on top of each other, as shown in Figure 2. These movements might not be tracked when relying only on the HMD due to the fact that hand tracking depends on the line of sight, and the head orientation does not allow a line of sight to the hands, in some configurations. In this case, the hands turn into standby mode, and the body configures a boredom posture, resulting in an incorrect prediction by the trained model. This is also somewhat true for happiness, wherein users choose to open and move their arms. Losing the line of sight between the headset and hands in the upper-body pipeline is a major problem that prevents feature extraction. Having said that, feature extraction in the wrist-movement pipeline was more reliable owing to the presence of controllers for retrieving acceleration data.

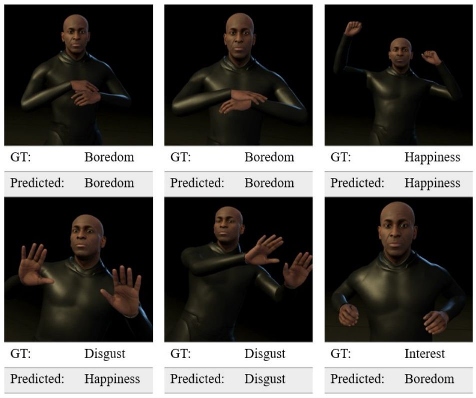

Figure 2: *Real-time tests of the proposed solution.*

4. Conclusion

Non-verbal communication is an essential component of human interaction and plays a key role in the design of user-friendly interfaces in HCI. In this regard, an automated emotion recognition solution based on non-verbal cues was proposed and evaluated after performing real-time tests. The keypoint and acceleration data extracted from the avatar were obtained using SDKs, which significantly lowered the computational costs, compared with feature extraction using trained AI models, without requiring additional sensors.

5. Acknowledgment

This work was supported in part by the Ministry of Science and ICT (MSIT), South Korea, under the Information Technology Research Center (ITRC), Supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP), under Grant IITP-2023-RS-2022-00156354; and in part by the Ministry of Trade, Industry and Energy (MOTIE) and Korea Institute for Advancement of Technology (KIAT) through the International Cooperative Research and Development Program under Project P0016038.

References

[MGKK15] Molchanov P., Gupta S., Kim K., Kautz J.: Hand gesture recognition with 3d convolutional neural networks. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 2015-October* (10 2015), 1–7. doi:10.1109/CVPRW.2015.7301342. 2

[NCK*18] Noroozi F., Corneanu C. A., Kaminska D., Sapinski T., Escalera S., Anbarjafari G.: Survey on emotional body gesture recognition. *IEEE Transactions on Affective Computing 12* (1 2018), 505–523. URL: https://arxiv.org/abs/1801.07481v1, doi:10.48550/arxiv.1801.07481. 2

[QGY18] Quiroz J. C., Geangu E., Yong M. H.: Emotion recognition using smart watch sensor data: Mixed-design study. *JMIR Ment Health 2018;5(3):e10153 https://mental.jmir.org/2018/3/e10153 5* (8 2018), e10153. URL: https://mental.jmir.org/2018/3/e10153, doi:10.2196/10153. 1, 2

[SKPA19] Sapiński T., Kamińska D., Pelikant A., Anbarjafari G.: Emotion recognition from skeletal movements. *Entropy 2019, Vol. 21, Page 646 21* (6 2019), 646. doi:10.3390/E21070646. 1