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

Poster

Emotion-based Interaction Technique Using User's Voice and Facial Expressions in Virtual and Augmented Reality

Beom-Seok Ko[1], Ho-San Kang[1], Kyuhong Lee[1], Manuel Braunschweiler[2], Fabio Zünd[2], Robert W. Sumner[2] and Soo-Mi Choi[1]†

[1]Department of Computer Science and Engineering and Convergence Engineering for Intelligent Drone, XR Research Center, Sejong University, Korea
[2]Game Technology Center, ETH Zürich, Switzerland

Abstract

This paper presents a novel interaction approach based on a user's emotions within augmented reality (AR) and virtual reality (VR) environments to achieve immersive interaction with virtual intelligent characters. To identify the user's emotions through voice, the Google Speech-to-Text API is used to transcribe speech and then the RoBERTa language processing model is utilized to classify emotions. In AR environment, the intelligent character can change the styles and properties of objects based on the recognized user's emotions during a dialog. On the other side, in VR environment, the movement of the user's eyes and lower face is tracked by VIVE Pro Eye and Facial Tracker, and EmotionNet is used for emotion recognition. Then, the virtual environment can be changed based on the recognized user's emotions. Our findings present an interesting idea for integrating emotionally intelligent characters in AR/VR using generative AI and facial expression recognition.

CCS Concepts

• ***Human-centered computing*** → ***Human computer interaction (HCI)***; • ***Hardware*** → *VIVE Pro Eye; Facial Tracker;*

1. Introduction

Recently, the metaverse has emerged as a shared virtual space where people can experience life and communicate with others in ways that cross the boundaries of the physical world. As the metaverse has gained attention, augmented reality (AR) and virtual reality (VR) technologies have received a lot of attention as important elements of the metaverse. The application of AR/VR technologies in the metaverse has been widely studied in various fields such as healthcare, medical treatment, education, performance, and industry. The applications in these fields can be implemented with AR or VR technologies, or a combination of AR and VR technologies in a single system to benefit from both [KYK*23]. As each technology advances, research on virtual characters that can serve as guides in AR and VR is also actively underway as a method to enhance the level of immersion and technological achievement [Car22]. In AR/VR, non-player characters (NPCs) act as guides who can enhance immersion and provide help for new environments. NPCs are hence essential, and intelligent characters are becoming increasingly important because they can be used in various content. Moreover, incorporating artificial intelligence (AI) technology into the NPCs enables them to respond in natural dialogues. For realistic interaction with characters, emotion is identified through the recognition of the character's facial expressions in VR [VQFCM*22] In recent years, with the advancement of generative AI technologies, ChatGPT, a language model that can communicate with humans, has emerged, and attempts to use this technology are increasing. The field of AR and VR, there is also a growing expectation of lifelike interactions with intelligent characters using ChatGPT. To this end, we aimed to study interaction that recognizes the users' conversation and facial expressions to enhance their immersive experience based on the emotion.

2. Methods

2.1. Overview

Figure 1 outlines the system pipeline for this study. The user uses the VIVE Pro Eye HMD and microphone to deliver his/her eye and facial movements and voice, respectively, to the system. Each input is used as an input to a deep learning model to perform interactions based on the user's voice and interactions based on the user's eye and lower face movements in the VR environment, i.e., facial expression. For the system, we used a desktop computer (Ryzen 9 5950X, RTX 3080 Ti, 64G RAM), VIVE Pro Eye HMD, VIVE Facial Tracker, and a rozet pin mic. To construct the AR and VR environments, we used the SRWorks v0.9.7.1 software development kit (SDK) provided by VIVE in the Unity3D engine (2019.4.40f1) environment.

† Corresponding author
(smchoi@sejong.ac.kr)

DOI: 10.2312/pg.20231286

2.2. AR/VR Integration Environment

In this study, we implemented a technique to change the AR or VR environment according to the user's emotion. In particular, we aimed to provide users with a higher sense of immersion by enabling them to experience both AR and VR on a single device. This system can recognize chairs and desks around the user in the AR environment to create an environment where the user can interact with the intelligent character. To recognize the objects, we used the SRWorks SDK's AI module and Mobilenet-v2 deep learning model to distinguish the features of real objects. The user in the real world is transitioned into the virtual environment according to his/her emotion. During the transition, a fade in-out method was used to provide an interesting experience to the user.

2.3. Intelligent Character Interaction

To implement various interactions with the user in AR/VR environments based on his/her emotions, we implemented an intelligent character that can recognize emotions. It recognizes user's emotions through the user's dialog in the AR/VR environments, and by using a part of the user's facial expression in the VR environment. The user's voice input through the microphone is converted into text using the Speech-to-Text (STT) API of Google Cloud, as shown in Figure 2 (Left). The converted text is used as an input to the RoBERTa model to extract specific words related to emotions, based on which the user's emotion is identified. Interaction with the user can be performed by changing the styles and properties of certain objects according to the identified emotion. Figure 2 (Right) shows an example of interaction in the VR environment, where, VIVE Pro and Facial Tracker were used to track the user's eyes and the lower part of his/her face, including the mouth and cheeks. Facial Tracker measures the degree of movement of each recognized facial part as a weight and stores it as an input to the next deep learning network. These weights are then used as inputs to EmotionNet to classify the emotion. The classified emotion can then be used to change the virtual environment around the user to provide a new interaction for the user. Our emotion-based interaction approach using user's voice and facial expressions in AR/VR helps the user become more immersed in AR/VR environments, enabling more realistic interactions with intelligent characters.

3. Conclusion and Futurework

We proposed a system that utilizes the user's voice and facial expressions in AR/VR to create a more immersive experience. In this study, the user experiences interactions in AR/VR through a single device. In AR, the user can experience interaction in the form of change in specific objects. This change is performed by AI using keywords extracted from the conversation with the user. In VR, we studied interactions where the system accepts the user's facial expressions as input and extracts the underlying emotion by tracking the user's eyes and lower face. Then, it changes the environment based on the extracted emotion. In the future, we plan to research and develop natural and realistic interaction technologies using intelligent NPCs and improve the deep learning models currently utilized for these applications.

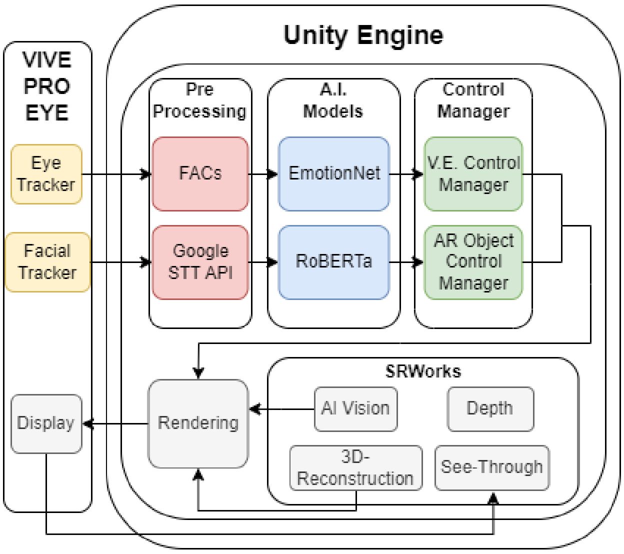

Figure 1: *System architecture*

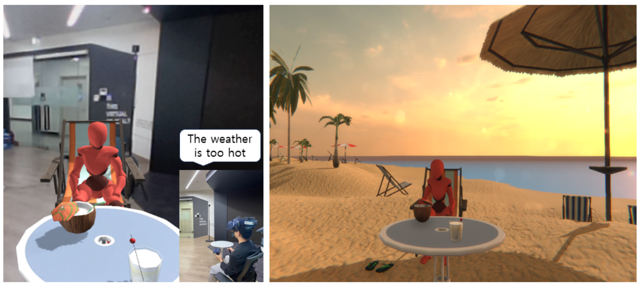

Figure 2: *(Left) AR interaction using voice, (Right) VR environment based on facial expressions.*

Acknowledgements

This research was supported in part by the MOTIE and KIAT through the International Cooperative R&D program (No. P0016038) and in part by the MSIT and IITP (IITP-2023-RS-2022-00156354).

References

[Car22] Carroll D.: Attention and communication in virtual worlds: Interacting with non-player characters in virtual reality. 1

[KYK*23] Kang H., Yang J., Ko B.-S., Kim B.-S., Song O.-Y., Choi S.-M.: Integrated augmented and virtual reality technologies for realistic fire drill training. *IEEE computer graphics and applications* (2023). 1

[VQFCM*22] Vicente-Querol M. A., Fernández-Caballero A., Molina J. P., González P., González-Gualda L. M., Fernández-Sotos P., García A. S.: Influence of the level of immersion in emotion recognition using virtual humans. In *International Work-Conference on the Interplay Between Natural and Artificial Computation* (2022), Springer, pp. 464–474. 1