ISSN 1975-4736 ⓒMITA2023

Developing Empathetic Virtual Humans:

Mimicking User's Facial and Eye Responses

Deaho Yoon
Department of Emotion Engineering, Sangmyung University
Seoul 03016, Republic of Korea
eoghrkwhr@gmail.com

Sung Park
Department of Emotion Engineering, Sangmyung University
Seoul 03016, Republic of Korea
sjpark@smu.ac.kr

***Abstract*— The objective of developing a virtual human that can identify human emotions and demonstrate empathy has been pursued with the aim of gaining user trust, promoting likability, and fostering intentions for long-term engagement. Central to achieving affective empathy is the concept of motor mimicry. This paper details our efforts to enable a virtual human to identify and mimic a user's facial and eye responses over time. We utilized Convolutional Neural Network (CNN) and Support Vector Machine (SVM) classifiers for facial and eye response recognition, respectively and fused the responses at the decision level.**

Keywords—virtual human; virtual agent; digital human; ECA; metaverse; empathy; affect; emotion; multimodal

I. Introduction

Empathy is a multifaceted construct, ranging from affective and cognitive empathy, with several factors contributing to its manifestation, including empathic capability, experience, personality, and relationship dynamics (for a comprehensive review, refer to [1]). Efforts to engineer empathy within virtual humans are still in its early stages, confronted with the challenge of recognizing human affect, a complex process involving multiple verbal and nonverbal cues from human users. In this study, we present our efforts toward enabling a virtual human to identify the user's emotions by integrating and mirroring the user's eye and facial responses. We took this first step due to the understanding that motor mimicry serves as the cornerstone of affective empathy [2].

II. Method

In this section, we describe how our virtual human recognize and express empathy to users. We attempted to design a virtual human empathy expression process based on the theory of "emotional contagion" among the definitions of empathy [3]. Therefore, our design goal was to develop virtual humans express similar emotions to the user's emotional state. The developed prototype consists of recognizing the emotional state through the user's non-verbal response (face and eyes) and expressing empathy based on the recognized user's emotional state. Considering that the user's emotional state is recognized through the responses of the face and eyes, empathy expression of virtual human is also designed to adjust the facial expression and pupil size similar to the user's emotional state (see Figure 1).

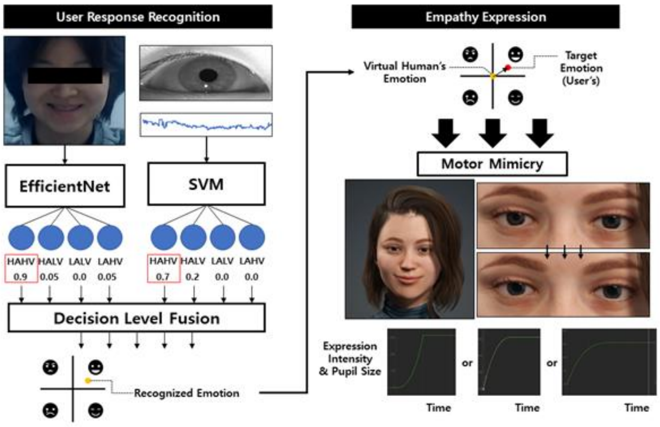

Fig. 1. Virtual Human Empathy Recognition and Expression Process

A. Data Acquisition

Datasets for training emotion recognition models were collected through the following experiment with 47 participants. As depicted in Figure 2, facial responses were collected by requesting the user to perform an imitation task that mimics the virtual human expression shown on the monitor, and then recording the user's face through a webcam during the task (30 fps, 1080p). As shown in Figure 3, eye responses were collected by presenting video stimuli designed to stimulate specific emotions and then recording pupil changes while the participant watched the video. Gazepoint's GP3 eye tracker was used to record the user's pupil changes (60 fps).

Virtual human facial expressions for imitation tasks and videos for emotion stimulation are designed based on Russell's dimensional emotion model [4]. Specifically, those were designed according to the four categories of the Valence-Arousal dimension by Russell (HAHV: High Arousal High Valence, HALV: High Arousal Low Valence, LALV: Low Arousal Low Valence, LAHV: Low Arousal High Valence).

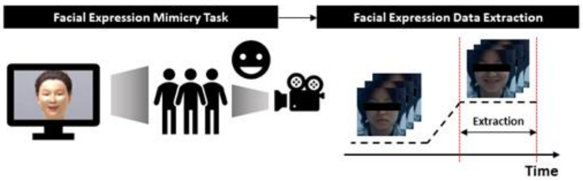
Facial Expression Mimicry Task
Facial Expression Data Extraction
Extraction
Time

Fig. 2. Process of Measuring Facial Response

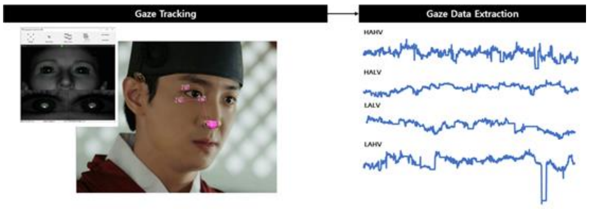

Fig. 3. Process of Measuring Eye Response

B. *User Response Recognition*

This section describes how we recognize emotional states by integrating the user's facial and eye responses. We developed a fusion model that classifies four emotional states. Each emotion classifier trained single response data using CNN or SVM.

The facial response-based emotion classifier was trained using CNN developed based on EfficientNet [5] (see Figure 4). For better learning performance, the area where the face was located from each image was extracted as an area of interest using the face detection method, and then faces were aligned based on the distance between the two eyes. The aligned faces were resized to 224x224x3, the standard input size of EfficientNet, and then used as a training dataset. The dataset was divided into training and validation, using 80% as training and 20% as validation. The trained classification model achieved 80% validation accuracy.

The eye response-based emotion classifier was trained using SVM and used pupil diameter and saccade movement components (magnitude and direction) as features to classify emotion (see Figure 5). In addition, our eye response data were acquired for a very long time due to the long video running time (i.e., 6 minutes). Consequently, the data collected while one participant was watching one video was averaged. SVM classification models have been trained under various kernels (linear and rbf), C, and gamma parameter (from 0.001 to 100) conditions. The trained models were evaluated through 5-fold cross-validation, achieving a maximum accuracy of 45%.

Finally, we incorporate emotions from the two classifiers. Our fusion technique is a kind of decision-level fusion, and we chose the one with the highest probability among the results of the two emotion classifiers. Based on the definition of emotion contagion, our virtual human mimicked the user's response mapped on the dimensional model, varying pupil size and expressive intensity (see Figure 1).

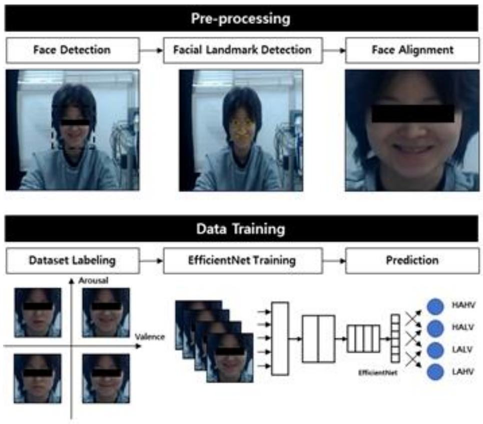

Fig. 4. Training Process of Eye Response-based Emotion Recognition Model

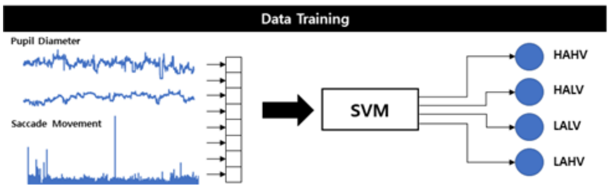

Fig. 5. Training Process of Eye Response-based emotion recognition model

III. Conclusion and Discussion

In this paper, we discuss our efforts to enable virtual humans to identify users' emotional states and exhibit empathy in response to these detected emotions. For emotion recognition, we gathered emotional response data from facial expressions and eye movements, subsequently training this data via CNN and SVM classifiers. Finally, we developed a prototype that integrates emotions detected by the two classifiers to enhance the accuracy of emotion recognition. Furthermore, it was designed to simulate empathetic expressions in virtual humans through facial expressions and eye movements, given that these modalities are the primary means for emotion recognition. In the future, we aim to refine the classification models by investigating various parameter configurations. Moreover, we will consider adopting alternative methodologies to improve the accuracy of emotion recognition derived from eye responses.

Acknowledgment

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2022R1I1A1A01066657).

References

[1] S. Park and M. Whang, “Empathy in human–robot interaction: Designing for social robots,” *Int. J. Environ. Res. Public Health*, vol. 19, no. 3, p. 1889, 2022.

[2] J. B. Bavelas, A. Black, C. R. Lemery, and J. Mullett, “‘ I show how you feel’: Motor mimicry as a communicative act.,” *J. Pers. Soc. Psychol.*, vol. 50, no. 2, p. 322, 1986.

[3] E. Hatfield, J. T. Cacioppo, and R. L. Rapson, “Emotional contagion: Studies in emotion and social interaction,” *Ed. la Maison des Sci. l'homme*, 1994.

[4] J. A. Russell, “A circumplex model of affect.,” *J. Pers. Soc. Psychol.*, vol. 39, no. 6, p. 1161, 1980.

[5] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *International conference on machine learning*, 2019, pp. 6105–6114.