# Do Digital Human Facial Expressions Represent Real Humans?

Shiori Kikuchi, Oky Dicky Ardiansyah Prima, Hisayoshi Ito
Graduate School of Software and Information Science
Iwate Prefectural University
Takizawa, Iwate, Japan
e-mail: g231u017@s.iwate-pu.ac.jp, {prima, hito}@iwate-pu.ac.jp

***Abstract*— The recent development of advanced digital humans has the potential to faithfully represent human nonverbal information. There are many examples of the use of digital humans for interactive communication, such as customer support and digital healthcare. There are, however, issues to be addressed as communication tools because of the lack of evidence regarding the extent to which nonverbal communication is performed by digital humans. In this study, we evaluate the quality of facial expressions by the digital human and its actors objectively using deep learning-based facial expression recognition. For the experiment, facial images of an actor expressing six basic emotions (anger, fear, disgust, happiness, surprise, and sadness) and a digital human face resembling the actor were captured, and scores of both facial expressions were measured. The results showed that the expression of "happiness" by the digital human and its actor were significantly consistent, but there were no significant differences in other facial expressions. The facial expression recognition used in this experiment was not trained on digital humans, thus there were cases in which the facial expressions of digital humans could not be judged accurately.**

***Keywords-expression; digital human; facial expression; basic emotions; avatar.***

## I. INTRODUCTION

The spread of coronavirus (COVID-19) has changed the way people communicate to form interpersonal relationships, leading to the use of video calls for meetings and discussions. Video calls enable more effective communication by conveying visual and non-verbal information along with voice information. In Japan, the establishment of “Virtual Shibuya” [1] and the “Medical Metaverse Joint Research Chair” [2] have raised interest in virtual spaces and digital humans and are expected to promote communication using these technologies.

Facial expressions are a very important element in human communication, and in particular, human emotions play an especially important role in facial expressions. Consequently, the quality of reality of facial expressions of digital humans is also considered to be important for communication in virtual spaces. The advance computer vision has enabled facial expression recognition from facial images at a practical usage. Its application is widely available in areas, such as medicine, security, and marketing.

The representation of real buildings and people in virtual space has attracted much attention for a long time. Digital humans can be generated in high reality by capturing images of people with multiple cameras [3]. MetaHuman by Epic Games provides an elaborate representation of reading by an actor, enabling the actor's detailed changes in facial expression to be seen in the digital human [4]. These factors have led to the expectation of faithful representation of nonverbal information by sophisticated digital humans, and further research and development are currently in progress. However, it has not been fully verified to what extent the quality of facial expressions of digital humans is comparable to that of humans.

This study compares human and digital human facial expressions and examines the quality of digital human facial expressions. For this purpose, we created a system that reflects the actor's facial expressions to the digital human in real-time and automatically recognizes the six basic emotions based on their facial expressions.

The rest of this paper is organized as follows. Section II discusses the related works of facial expression analyses and digital human. Section III describes the generation of facial expression data and tools used for this purpose in this study. In Section IV, we describe our experiments to evaluate the quality of facial expressions for the digital human. Finally, Section V summarizes the results of this study and discusses future perspectives.

## II. RELATED WORKS

Facial expressions are important in communication both in virtual space and face-to-face. This section describes research on facial expressions and analysis of facial expressions in face-to-face communication and research on digital human reality.

### A. *Facial Expressions in Communication*

Understanding one's emotions is an important part of communication. Facial expressions in particular reveal human emotions. Ekman et al. proposed the Facial Action Coding System (FACS), which classifies the Action Units (AUs) of facial parts to identify emotions from facial expressions [5]. They pointed out that there are "display rules" based on such cultural norms thus the intensity of facial expressions differs depending on the culture. For example, Japanese people tend to suppress their facial expressions [6]. In addition to facial expression recognition using AUs, recently, facial expression recognition from facial images using deep learning has been widely used [7].

**ISBN: 978-1-61208-982-9**

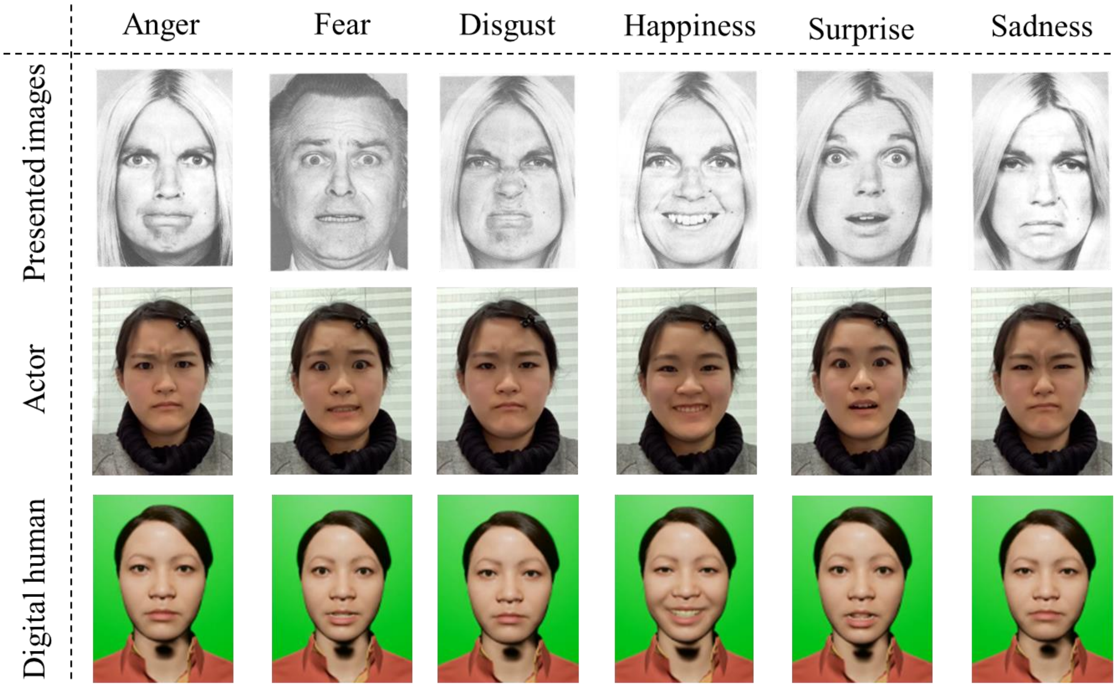


Figure 1. Expression of actors and digital humans on images of six basic emotions.

TABLE I. FACIAL EXPRESSION RECOGNITION RESULTS

| Facial image | Facial images with correctly estimated emotion | | | | | |
|---|---|---|---|---|---|---|
| | Anger | Fear | Disgust | Happiness | Surprise | Sadness |
| Actor | 0 | 28 | 0 | 80 | 24 | 29 |
| Digital human | 3 | 0 | 0 | 147 | 0 | 0 |

[frame]

### B. Reality of the Digital Human

Recently, digital humans have been developed that resemble humans in appearance and are capable of projecting their body movements [4]. Kang et al. [8] surveyed and simplified the research on digital human reality, introducing two types of reality: visual realism, which is the similarity between the rendering of visual information of a person, and behavioral realism, which is the similarity between human behavior and the reality of a person. Visual realism is higher as the digital human looks more like a person, and behavioral realism is higher as the digital human performs natural movements. They also stated the importance of the influence of digital human reality on communication. Grewe et al. [9] compared the reality of facial expression animations created by experts with the reality of facial expression animations created statistically from a database of images and found that the statistically created animations were perceived as a more digitally human reality.

## III. GENERATION OF FACIAL EXPRESSION DATA

This study compares actor and digital human facial expressions and examines the quality of digital human facial expressions. For this purpose, we create a system that reflects the actor's facial expressions to the digital human in real-time and automatically recognizes the six basic emotions based on their facial expressions. In this section, we first present the tools we used. Next, we describe the system that reflects the actor's facial expressions onto a highly realistic digital human and the collected facial images.

### A. Tools Used in This Study

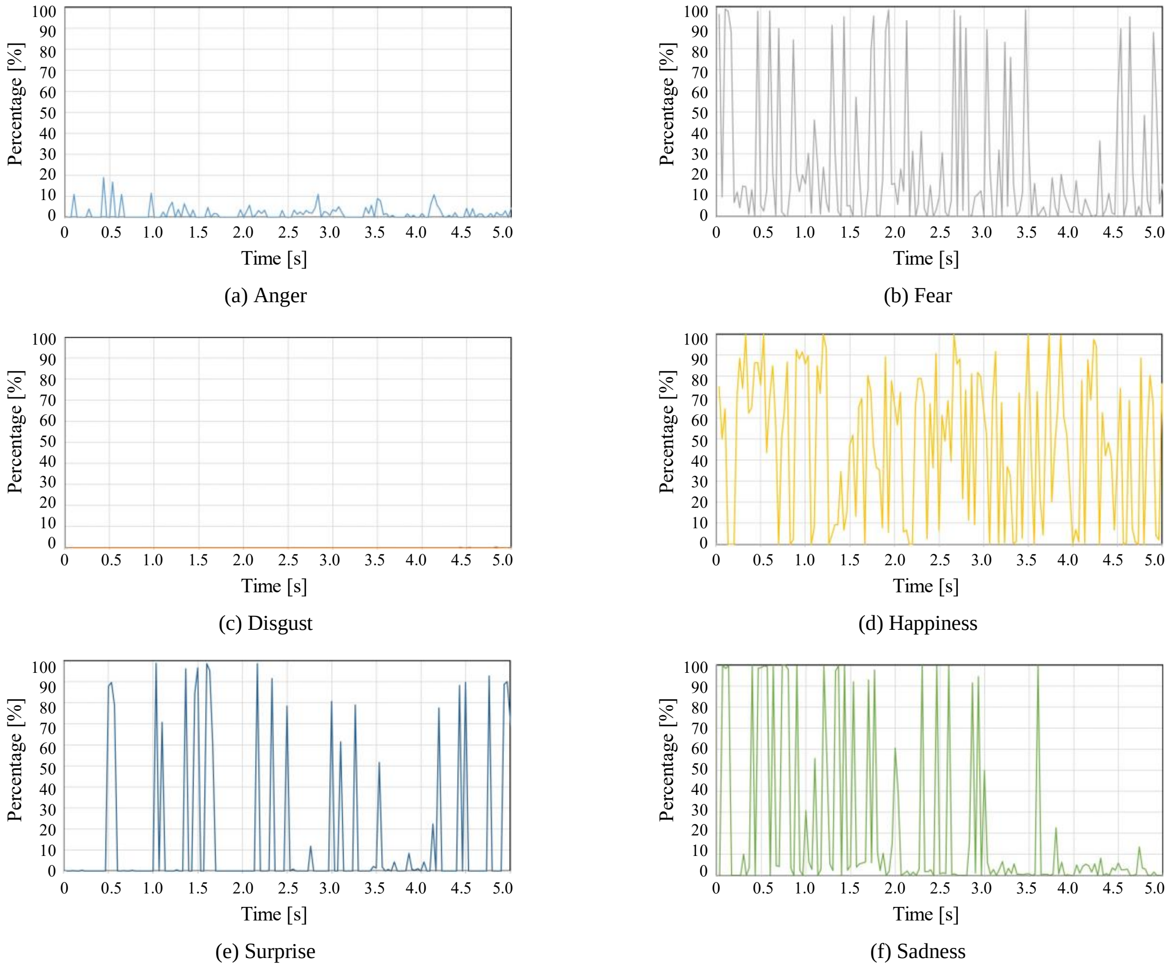
In this study, we use Epic Games' MetaHuman Creator (MHC), Unreal Engine (UE), and Live Link Face (LLF) to create a digital human and reflect the actor's facial expressions onto the digital human [10]. The MHC is a tool that facilitates the creation of photorealistic digital humans in the browser and can be used in conjunction with the 3D object rendering engine, UE. LLF, an iOS app from Epic Games, Inc. that transcribes actors' facial art to a superhuman in real-time, utilizes the device's true depth camera.

The tool for facial expression recognition used in this study is DeepFace [7], developed by Serengil and Ozpinar. It includes expression recognition trained using a convolutional neural network, which can estimate the percentages of six basic emotions from a single face image.

### B. Facial Expression Data

For the generation of digital human facial expression data, a set of six facial images were presented in sequence to a 22-

ISBN: 978-1-61208-982-9



(a) Anger

(b) Fear

(c) Disgust

(d) Happiness

(e) Surprise

(f) Sadness

Figure 2. DeepFace estimation of the actor's emotion for each frame.

year-old Japanese female actor, who imitated each facial expression for five seconds. The facial videos were acquired, and their features were recorded using LLF. UE was used to acquire videos of the corresponding digital human facial images. Facial images and their features were recorded at 30 fps. To minimize the influence of cultural differences and the actor's experience in expressing facial expressions, a set of facial images representing Ekman's six basic emotions [5] was presented to the actor, and the facial muscles characteristic of each emotion were explained to enable the actor to imitate them properly. Finally, 150 frames of facial images were collected from each five-second video of each emotion, bringing the total number of facial images of actors and digital humans imitating the six basic emotions to 900 frames respectively.

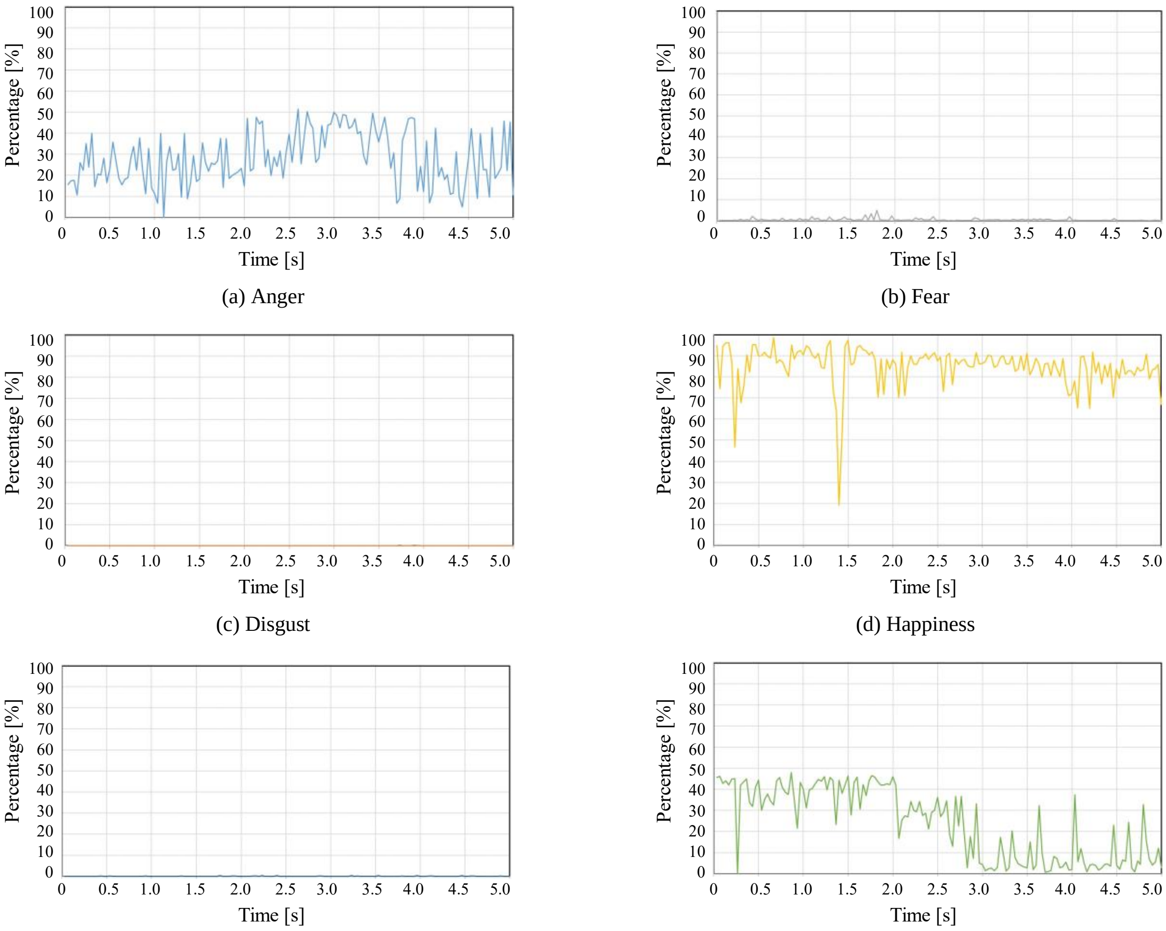
## IV. Evaluation of the Quality of Facial Expression

The actor's facial image and the facial image of the generated digital human are evaluated using DeepFace. The results of facial expression recognition for both are considered equivalent when the digital human can faithfully reproduce the actor's facial expression.

Based on the facial expression recognition results, each emotion that was estimated to be more than 50% was considered the dominant emotion expressed by the facial image. Therefore, the facial images imitated by an actor are considered similar if the percentage of emotions estimated in both the actor's and the digital human's facial images is over 50%.

The actors and the generated digital human facial images for each emotion are shown in Figure 1. The results of DeepFace of the respective facial expressions of the actor and the corresponding digital human when basic emotions were expressed in 5 seconds (150 frames of facial images) are shown in Table 1. When the actor mimicked the presented image, DeepFace correctly estimated the actor's facial expressions except for “anger” and “disgust,” whereas the digital human correctly estimated only “happiness.”

Figures 2 and 3 show the raw results of emotion estimation for actors and digital humans by DeepFace for each frame, respectively. As shown in the facial images in Figure 1, the actor lowered his eyebrows, glared into the eyes, and narrowed the lips to express anger. The digital human corresponding to the actor is also shown to have similar facial



(a) Anger

(b) Fear

(c) Disgust

(d) Happiness

(e) Surprise

(f) Sadness

Figure 3. DeepFace estimation of the digital human's emotion for each frame.

features. However, the estimations for both emotions were low. Similarly, the results were low in disgust, despite the actor's wrinkling of the nose and raising of the upper lip. We also found that expressions of surprise and fear in digital humans were not recognized by DeepFace due to the lack of these important elements, such as pronounced raised eyebrows and widening of the eyes. The estimated emotional intensity varies from frame to frame. This suggests that the resulting emotion by DeepFace should be averaged over a certain frame length, rather than done in a single frame. Another reason why DeepFace cannot correctly estimate facial emotion in digital humans is that its neural network has not been trained on digital humans.

## V. Conclusion

This study examined the quality of facial expressions of digital humans by objectively evaluating the facial expressions of an actor and its digital human through facial expression recognition, assuming the use of digital humans in communication. A visually realistic digital human was created to represent the actor, and facial expressions corresponding to behavioral reality were evaluated.

Facial expressions of six basic emotions were captured for five seconds by an actor and a digital human, and then facial expression recognition DeepFace was performed on each facial image for each frame. The results showed that the emotion of "happiness" was the most similar between the actor and the digital human, indicating that the quality of "happiness" by the digital human was high. However, since the emotion estimation for each frame by DeepFace varied, it seems necessary to consider multiple frames in the estimation of basic emotions.

Future work includes learning facial expression recognition using digital humans, improving facial expression estimation from continuous frames, and further validating the expression representation of digital humans.

## References

[1] Virtual Shibuya, https://new s.kddi.com/kddi/corporate/newsrelease/2020/05/15/4437.html [retrieved: May, 2022]

[2] Juntendo Virtual Hospital, https://jp.newsroom.ibm.com/2022-04-13-Juntendo-Virtual-Hospital [retrieved: May, 2022]

ISBN: 978-1-61208-982-9

[3] Y. Iwayama, “Real Avatar Production - Raspberry Pi Zero W Based Low-Cost Full Body 3D Scan System Kit for VRM Format,” 10th International Conference and Exhibition on 3D Body Scanning and Processing Technologies, pp. 22-23, 2019.

[4] Digital Andy Serkis, https://www.unrealengine.com/en-US/blog/epic-games-and-3lateral-introduce-digital-andy-serkis [retrieved: May, 2022]

[5] P. Ekman and W. V. Friesen, “Unmasking the Face: A Guide to Recognizing Emotions From Facial Expressions,” Malor Books, 2003.

[6] T. Kudoh and D. Matsumoto, “The Emotional World of the Japanese - Uncovering the Mysteries of their Mysterious Culture,” Seishinshobo, 1996.

[7] S. I. Serengil and A. Ozpinar, "LightFace: A Hybrid Deep Face Recognition Framework," 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 1-5, 2020.

[8] S. H. Kang and J. H. Watt, “The impact of avatar realism and anonymity on effective communication via mobile devices,” Computers in Human Behavior, 29(3), pp. 1169-1181, 2013.

[9] M. Grewe et al., “Statistical Learning of Facial Expressions Improves Realism of Animated Avatar Faces,” Frontiers in Virtual Reality, 2, pp. 1-13, 2021.

[10] MetaHuman Creator, https://www.unrealengine.com/en-US/metahuman-creator [retrieved: May, 2022]