Seamless Avatar Expression Transfer: Merging Camera and Smart Eyewear Embedded with Photo-Reflective Sensor Technologies

Masai, Katsutoshi
Kyushu University

Seamless Avatar Expression Transfer: Merging Camera and Smart Eyewear Embedded with Photo-Reflective Sensor Technologies

Katsutoshi Masai*
Kyushu University

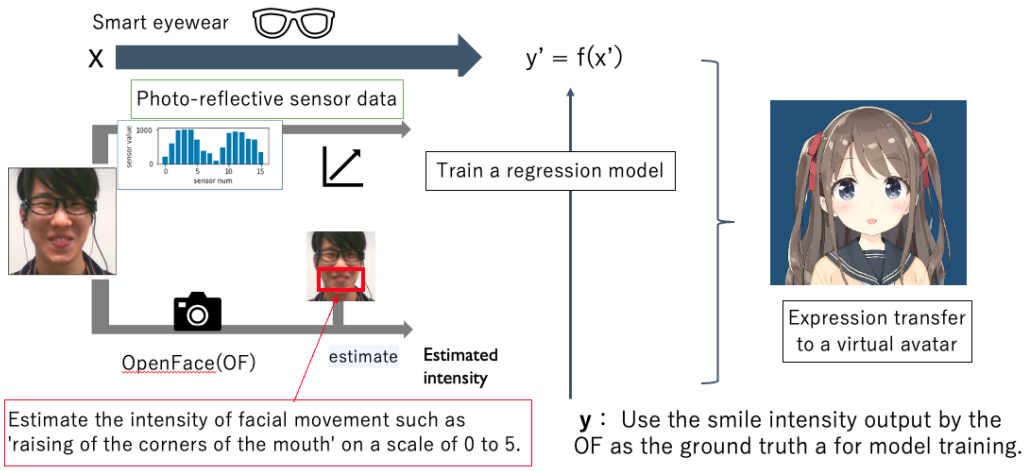

Figure 1: The method summary: Our method combines a camera-based method with smart eyewear embedded with photo-reflective sensors. This enables continuous and convenient expression transfer of a user's facial expression into a virtual avatar.

Abstract

Virtual Reality (VR) offers new ways to interact that are different from usual communication in the real world. In VR, avatars are key displays for showing users' feelings and personalities non-verbally. Addressing the need for more convenient and continuous expression representation, we introduce a novel method for facial expression transfer in VR. This method integrates camera systems with photo-reflective sensors embedded in eyewear, overcoming the limitations of traditional camera-based tracking. By offering smoother tracking and reducing manual calibration needs, this approach highlights the potential of multimodal technology to enhance non-verbal communication in virtual environments. Building on this, we demonstrated the example implementation of smile transfer and discussed future direction.

Index Terms: Human-centered computing—Interaction paradigms; Human-centered computing—Interaction devices;

1 Introduction

Communication between people varies depending on the method and environment. For example, in face-to-face situations, the ease of conversation can change with eye contact. Compared to a face-to-face conversation, some people may find it easier to talk naturally while engaged in a side activity. In virtual environments, users can choose their favorite avatars, and the communication style and perceived psychological pressure may differ from the real world. For example, the anonymity of virtual avatars may make communication easier. Virtual reality (VR) goes beyond traditional communication and offers new possibilities for interaction.

*e-mail:masai@ait.kyushu-u.ac.jp

Avatars in VR are essential for expressing users' personalities and emotions through their non-verbal information. They can reflect user actions such as head movement, blinking, and changing facial expressions. One of the applications is VTtubing, an emerging entertainment phenomenon in which individuals use VR avatars to create content and engage with audiences. Researchers are exploring ways to enrich expressions in VTubing. For example, AlterEcho, a VTubing animation system, automates dynamic avatar animations that synchronize with the streamer's movements for a more realistic experience [9]. This enhances VTubing visuals and VR non-verbal communication. In addition, virtual avatars can enhance nonverbal communication by using unique expressions and symbols, such as emoticons or cartoon-like visuals, that are not possible in real life. In such applications, the method of measuring facial expressions using facial tracking technology is critical [10].

This paper presents a novel method for transferring users' facial expressions to virtual avatars by integrating camera systems with 16 photo-reflective sensor-equipped smart eyewear [5] (see Fig. 1. Camera-based systems, which use a large database of facial images annotated with action units (AUs), enable expression recognition without the need for individual user calibration. However, they face limitations such as occlusion. Photo-reflective sensor-based methods, while requiring manual calibration to obtain annotated data, provide an alternative for facial expression tracking that addresses occlusion issues [5, 8]. Our approach combines these technologies. The smart eyewear captures subtle expressions [4], which enhances continuous expression tracking. This approach also avoids the discomfort associated with EMG facial tracking technologies because it does not require electrodes to be attached to the face. In addition, it is less resource-intensive by using cameras to capture facial expressions at reduced data acquisition intervals and using sensors to fill in the gaps compared to camera-only systems, allowing for more efficient use of computational resources. By integrating these modalities, we explore the potential of multimodal facial tracking

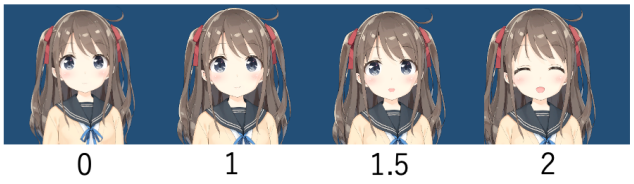

Figure 2: Designed expression intensity parameter of an avatar, ranging from neutral (0) to full smile (2).

technology to achieve accurate and continuous expression mapping on VR avatars without the limitations of traditional camera-based methods, thereby enhancing avatar expressiveness.

2 Recognizing Facial Expressions using Camera and Smart Eyewear

There are several methods for mapping expressions to avatars, such as using facial geometry, basing them on emotional states, or using Action Units (AUs) for more detailed representation [3]. AUs refer to specific movements of facial muscles that are fundamental to understanding human expressions. Each AU represents the movement of a different part of the face, and the combination of these movements creates different expressions. However, Geometry-based methods are less flexible. They link each model part directly to a facial movement. On the other hand, emotion-based methods mix many AUs. This mixing can make it difficult to show complex emotions. In contrast, AU-based methods are easier to use and modify. They allow users to customize parts for different facial expressions. OpenFace provides the state-of-the-art technology to estimate these AU movements [1, 2]. This tool helps to collect data to annotate sensor readings and to provide initial data for facial expression transfer. This prepares the transition to using annotation data for smart eyewear-based expression transfer. Smart eyewear [5], equipped with 16 photo-reflective sensors consisting of infrared LEDs and phototransistors, measures facial expression activity. When facial muscles move, they cause skin deformation, which changes the distance between these sensors and the skin surface. For example, when smiling, the skin around the cheeks contracts, bringing it closer to the sensors and changing the intensity of the reflected signal. This change in intensity, captured by the sensors, is used to transfer the user's smile to a virtual avatar.

3 The Case Study and Observation of Smile Transfer Possibility

In our study, we focused on implementing the reflection of a smile on avatars by concentrating only on AU12 (lip corner puller). We first used the AU12, derived from OpenFace, to estimate smile intensity [1, 2]. By using this data to train our system to annotate sensor readings, we developed a linear model with photo-reflective sensors to estimate the intensity of a smile in a range of intensities from 0 to 5. To prevent outputs below 0, we configured the system to output 0 whenever the value fell below this threshold. This approach follows the method established by Masai et al. [4].

To accommodate the continuous output of smile intensity, which ranges from 0 to 5, we designed the avatar parameters to correspond to these levels. We used the Hiyori Momose character model from Live2D Inc. and set parameters for four different levels: 0, 1, 1.5, 2, and above (Fig. 2). These parameters should be designed based on how the user wants to express their avatar's personality. In this case, the mapping was created with a happy personality in mind, one that smiles frequently. We then achieved a seamless transition between these levels by using linear interpolation, ensuring that the parameters were adjusted to represent the varying intensities of the smile. This approach resulted in a smooth and dynamic expression of the smile on the avatar.

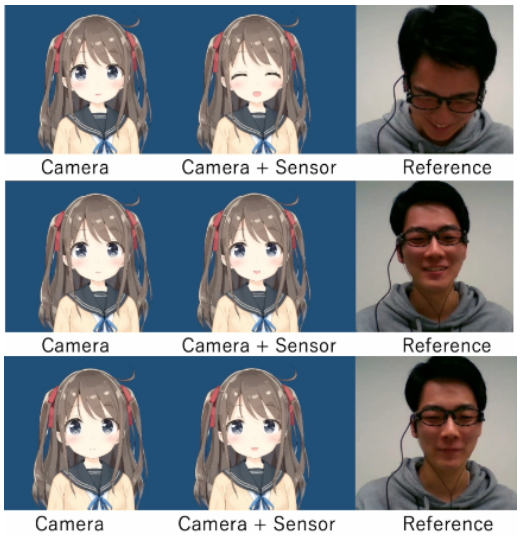

Figure 3: Examples of Avatar Expressions Compared.

We applied this method to a dataset containing both video data of a user's face and corresponding sensor data while the user was watching a comedy video [7]. Fig. 3 are screenshots of the user's video and the corresponding avatar's behavior, captured by the camera alone and in combination with the sensors. As expected, we confirmed that more continuous and enriched expressions were achieved, especially when changing the direction of the face or when trying to suppress a smile by opening the mouth.

4 Discussion and Future Work

We have created avatar expressions using only AU12. In the future, we plan to develop a mixture of models that can control individual parameters for emotion-related AUs, such as surprise (AU1+2) or anger (AU4). Using data from photo-reflective sensors and cameras and such models, we can animate avatars with more nuanced and expressive movements compared to the previous method [8].

VR technology transforms and manipulates users' real-life expressions onto their avatars, thus influencing social dynamics in virtual environments by controlling nonverbal parameters. We can use data collected from multiple individuals to replay reactions on avatars to enhance user engagement in virtual environments. How to aggregate information and how to reflect this information from different users can create new opportunities for social interaction using avatar expressions. One possible consideration is whether to base an avatar's expressions on a single individual or a blend of multiple people, as this influences the emotional experience of social interactions. This approach, while incorporating human responses, can explore the dynamics of communication in virtual environments. Avatar expressions can be manipulated to range from subtle to exaggerated. Oh et al. highlighted that controlled manipulation of expressions, such as enhancing smiles, can enhance positive communication experiences [6]. Our technology, which provides nuanced control over avatar expressions, expands the scope for studying psychological aspects in VR, making avatars valuable tools for analyzing social interaction in virtual environments and advancing our understanding of VR psychology.

Acknowledgments

This project is partially supported by JST ASPIRE Program Grant Number JPMJAP2327.

REFERENCES

[1] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pp. 59–66, 2018. doi: 10.1109/FG.2018.00019

[2] T. Baltrušaitis, M. Mahmoud, and P. Robinson. Cross-dataset learning and person-specific normalisation for automatic action unit detection. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 06, pp. 1–6, 2015. doi: 10.1109/FG.2015.7284869

[3] P. Ekman and W. V. Friesen. Facial action coding system. *Environmental Psychology & Nonverbal Behavior*, 1978.

[4] K. Masai, M. Perusquía-Hernández, M. Sugimoto, S. Kumano, and T. Kimura. Consistent smile intensity estimation from wearable optical sensors. In *2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 1–8, 2022. doi: 10.1109/ACII55700.2022.9953867

[5] K. Masai, Y. Sugiura, M. Ogata, K. Kunze, M. Inami, and M. Sugimoto. Facial expression recognition in daily life by embedded photo reflective sensors on smart eyewear. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, IUI '16, p. 317–326. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2856767.2856770

[6] S. Y. Oh, J. Bailenson, N. Krämer, and B. Li. Let the avatar brighten your smile: Effects of enhancing facial expressions in virtual environments. *PLOS ONE*, 11(9):1–18, 09 2016. doi: 10.1371/journal.pone.0161794

[7] C. Saito, K. Masai, and M. Sugimoto. Classification of spontaneous and posed smiles by photo-reflective sensors embedded with smart eyewear. In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '20, p. 45–52. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3374920.3374936

[8] K. Suzuki, F. Nakamura, J. Otsuka, K. Masai, Y. Itoh, Y. Sugiura, and M. Sugimoto. Recognition and mapping of facial expressions to avatar by embedded photo reflective sensors in head mounted display. In *2017 IEEE Virtual Reality (VR)*, pp. 177–185, 2017. doi: 10.1109/VR.2017.7892245

[9] M. T. Tang, V. L. Zhu, and V. Popescu. Alterecho: Loose avatar-streamer coupling for expressive vtubing. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 128–137, 2021. doi: 10.1109/ISMAR52148.2021.00027

[10] L. Wen, J. Zhou, W. Huang, and F. Chen. A survey of facial capture for virtual reality. *IEEE Access*, 10:6042–6052, 2022. doi: 10.1109/ACCESS.2021.3138200