2023 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (RISP NCSP 2023), Feb. 28 - Mar. 3, 2023

Supporting Participants' Synchronized Hand Gestures on a Video Conferencing System

Kan Oyama[1] and Hidekazu Shiozawa[1]

[1] Tamagawa University
6-1-1 Tamagawa-gakuen, Machida, Tokyo 194-8610, Japan
E-mail: trl621232@stu.tamagawa.ac.jp, shiozawa@eng.tamagawa.ac.jp

Abstract

Online video conferencing is becoming dramatically popular in these recent years and is used in various situations such as business meetings and remote education. However, it significantly decreases non-verbal information compared to face-to-face discussions. Usually, this causes problems such as insufficient quality of communication and participants' mental stress, in particular when they are not regularly acquainted with one another in person. To deal with this problem, we propose a video conferencing system that visually supports interactive synchronized gestures in which participants mutually synchronize their actions remotely. We think that such a system can be useful for icebreakers, group work in education, and casual meetings. This paper reports the implementation of an interactive high five gesture, and the evaluation of the usability and realistic feeling of the implemented functions.

1. Introduction

Online video conferencing is becoming dramatically popular in these recent years and is used in various situations such as business meetings and remote education. Even highly interactive collaboration such as group work in elementary education is demanded and actually performed with online video conferencing software.

However, it significantly decreases non-verbal information compared to face-to-face communication. This makes it very difficult for participants to share a sense of togetherness through interactive actions such as high fives, hugs, and handshakes as they are in person. For example, the way to share the sense of accomplishing a group task is almost only to say something and show small gestures applauding in front of their cameras individually.

In the recent COVID-19 pandemic situation, Japan Football Association offered an online program titled "Remote High Five with SAMURAI BLUE!" [1] at international friendly matches in the Netherlands for the players' families and supporters in Japan. In this program, the players performed high five gestures online with the participants on a large screen via Zoom before the matches.

There is an interesting study [2] that compared interactive gestures such as handshaking and high five, and non-interactive gestures that do not strictly require the receiver's response, such as the OK sign, thumbs up, and peace sign. The results suggest that interactive gestures make people actively participate in social interactions.

In this research, we propose a video conferencing system that visually supports interactive synchronized gestures in which participants synchronize their actions remotely. We think that such a system can be useful for icebreakers, group work in education, and casual meetings.

2. Related Work

HyperMirror [3] developed by Morikawa is a distributed system providing an environment in which participants can feel as if they were in the same room during online video meetings. The system composites video images of remote participants received from remote sites into the same scene with chroma key compositing and sends them back to remote sites. Each site has a set of a projector and a screen that projects the composited scene in front of a participant so that all the participants appear to be in the same place on the screen even though they are far apart. The author reported that synchronized gestures by participants such as handshakes, hugs, and touches were frequently observed.

NTT's Remote High-Five [4] is a research project, which transmits the physical feel of high fiving between users even at a distance. In this system, each user has a transparent board attached to a screen showing a video image of a remote user. When a user touches the board in front of the displayed images of the remote user like a real high five action, the remote user's board also vibrates to convey the tactile sensation of a high five.

MirrorBlender [5] developed by Grønbæk et al. supports hybrid meetings connecting real places by the network. In the system, camera images taken at connected sites are made translucent and superimposed on a shared document such as a map image. This allows participants to point with a gesture directly at the same shared document so that even remote participants in a hybrid conference feel as if they were participating in the meeting in the same way.

There are several software tools that recognize specific gestures from a camera image and show a user's intention or emotion as a picture superimposed on the image such as [6]. These software tools recognize specific hand gestures performed by users in the video images by using machine learning with feature matching. This technology allows people to react without speaking and express their intentions even in a large group video conference.

3. Proposal

We propose a video conferencing system that visually supports synchronized interactive gestures [7] [8], in which participants can perform interactive gestures such as high five by synchronizing the timings of their actions in front of their cameras. It is based on the idea that functions intentionally supporting visual interactive gestures increase non-verbal communication and can have a positive effect on participants' communication activity even online.

In this research, we adopted the high five. High five is considered effective and feasible because it is an action using hands and arms that can be easily captured on video. At the same time, it can be used versatilely in successful situations of such as collaborative work.

When we high-five in person, each participant visually predicts where the other's hands will be in the next moment, then moves the hands toward the predicted positions. In video conferencing, however, it is difficult for a user to predict where the other's hands will move on the two-dimensional video image. Therefore, to support the high five function, we propose that the system displays visual targets on the video image to guide where users move their hands to. When their hands are on the displayed visual targets, which implies a successful high five, the system emits a kind of visual and sound effects such as touch sound and small particles to indicate a successful high five.

We implemented the following three types of online high five methods to compare and evaluate their usability and realistic feeling.

(a) High Five at Free Positions

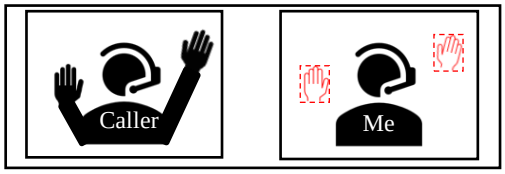
In this method, just as in a face-to-face situation, real-time positions of the remote user's hands are used as the target positions for the local user. As shown in Fig 1(a), the moving red frames as target positions are displayed on the video image of the local user according to the real-time automatically recognized positions of the remote user's hands.

(b) High Five at Fixed Positions

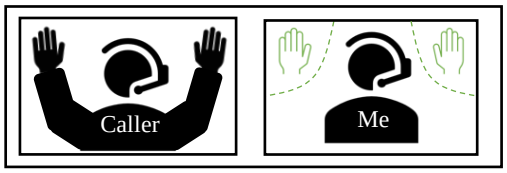
In this method, the touch areas are displayed as the target positions at the top corners of the local user's own video image, which are shown as green areas in Fig 1(b). It is expected that the user can smoothly and quickly move their hands to the touch areas, even if there is a communication delay in video transmission.

(c) Hybrid High Five Method

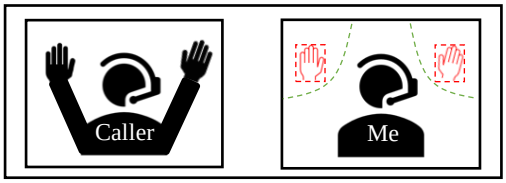
This method compromises the former two methods. As shown in Fig 1(c), the red frames as target positions are displayed only when the local user's hand positions are within the green touch areas on the top corners.



(a) High five at free positions



(b) High five at fixed positions



(c) Hybrid high five method

Fig 1: The three types of online high five methods

4. Implementation

Fig 2 shows the overview of the system developed in this research. We constructed a small video conferencing system on top of the web server Nginx. To realize video communication, we used SkyWay provided by NTT Communications. It is built with WebRTC technology, and its SDK and API are publicly available.

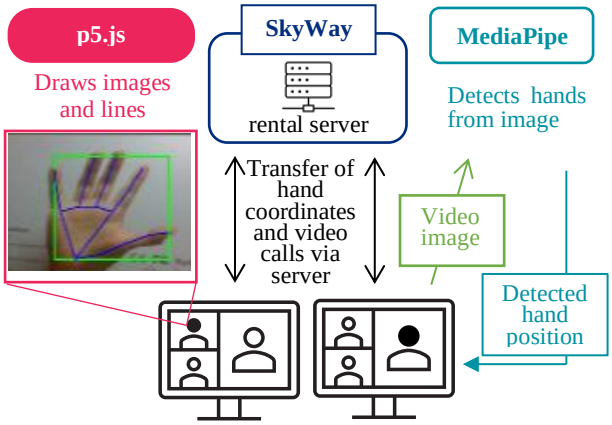


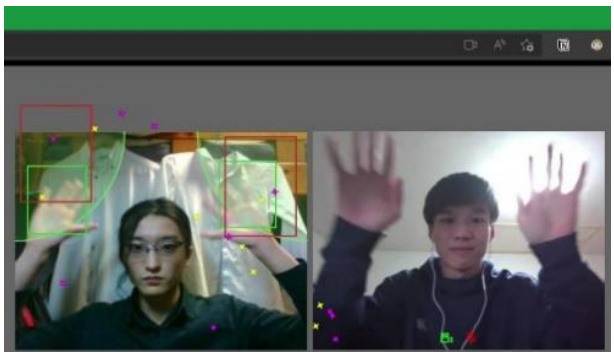
Fig 2: Overall view of the system architecture

We implemented the functions that the system displays participants' video images and target positions (red and green frames) in JavaScript and p5.js on a web browser as a client. To recognize users' hand gestures, we used Google MediaPipe Hands, which is a high-fidelity hand and finger tracking solution and provides a service that sends back the joint positions of hands from a received image. The system detects hand position coordinates and transmits them to all the participants to synchronize target positions.

Three types of high fives as shown in Fig 3 are supported. In addition, to investigate the effective presentation, we implemented a mode in which the closer both the participants' hands are, the more particles are emitted, for the free position method and the hybrid method.

(a) High five at free positions

(b) High five at fixed positions



(c) Hybrid high five method

Fig 3:Screen images of online high fives

(a) High Five at Free Positions

For the free position method, according to the remote user's hand positions, the system draws red frames on the local user's image (Fig 3(a)). The positions of the local user's hands are also detected and drawn as green frames. When the green frames overlap the corresponding red frames, visual and sound effects are generated.

(b) High Five at Fixed Positions

In the fixed position method, the system draws the green touch areas on the top corners as shown in Fig 3(b). When the user moves their hand into the areas, yellow particles are generated. If the other user's hands are also in the touch areas at the same time, a touch sound and red particles are generated as a successful high five.

(c) Hybrid High Five Method

The hybrid high five method is a compromise between free and fixed positions. The red and green frames are displayed only when hands are in the touch areas on the top corners, and visual and sound effects are generated when the red frames and the green frames overlap (Fig 3(c)).

5. User Study

To evaluate and compare the usability and realistic feeling of the proposed high five methods, we conducted a user study on 13 participants (9 males and 4 females between the ages of 21 and 23). The participants were asked to work in pairs to solve riddles and give a high five to each other when they successfully answered a riddle.

For comparison, we used six types of high five functions including four types of high five methods: just pretending high five, high five at free positions, high five at fixed positions, and the hybrid method, as well as two types of high fives at free positions and the hybrid method with different quantity of visual effects.

After experiencing all types of functions, participants answered the questionnaire shown in Table 2. Most of the questionnaire items were rated on a 5-point Likert scale and additional free comment sections.

Table 2 shows the excerpted results of the user study on the high five function.

The result of Q2 shows the responses about the easiness of matching hands. The fixed position method and the hybrid method received high evaluations. On the other hand, the free position method showed a large scattering.

The result of Q4 shows the responses about the suitability for video communication. While the fixed position method received better average points, the hybrid method remarkably received no negative evaluations. This suggests that the hybrid method mitigates the issues of the free and the fixed position methods.

Table 2 Questionnaire for evaluation (excerpt)

No.	Question
Q2	How easy was it to match your hands in the high five function? (1 very difficult to match ~ 5 very easy to match)
Q4	Please tell us your score (1~5) of high five in terms of its suitability for video meetings.
Q7	What do you think about the function that changes the number of visual effects? (1 not necessary ~ 5 necessary)
Q9	Please tell us your score (1~5) for the high five function in terms of similarity to face-to-face.
Q11	I want to ask you about your level of relaxation before participating in the experience. (1 very nervous ~ 5 very relaxed)
Q13	Did you enjoy online high fives with or without the function? (1 not enjoyable at all ~ 5 very enjoyable)
Q16	What is your score (1~5) for the high five in terms of enjoyment?

Table 1 Result of the questionnaire (excerpt)

No.	Method	1	2	3	4	5	Ave	SD
Q2	Free position	0	2	4	4	3	3.62	1.00
	Fixed position	0	0	0	6	7	4.54	0.49
	Hybrid	0	1	1	5	6	4.23	0.89
Q4	Pretending	5	2	4	1	1	2.31	1.26
	Free position	0	2	3	3	4	3.75	1.04
	Fixed position	0	1	2	2	5	4.10	0.91
	Hybrid position	0	0	4	4	4	4.00	0.82
Q7	-	0	1	3	6	3	3.85	0.86
Q9	Pretending	5	2	4	2	0	2.23	1.12
	Free position	0	0	4	5	4	4.00	0.78
	Fixed position	2	2	1	5	3	3.38	1.38
	Hybrid	0	0	2	7	4	4.15	0.66
Q11	-	0	4	2	4	3	3.46	1.15
Q13	-	0	0	2	7	4	4.15	0.66
Q16	Pretending	5	2	5	0	1	2.23	1.18
	Free position	0	0	2	3	8	4.46	0.74
	Fixed position	0	0	2	4	7	4.38	0.73
	Hybrid	0	0	2	3	8	4.46	0.74

The result of Q7 shows the responses about the function of varying the number of visual effects. The function does not seem to work effectively.

The result of Q9 shows the responses about the similarity to real face-to-face high fives. The hybrid method received the highest evaluation and just pretending high five is by far the worst.

The result of Q16 shows the responses about the enjoyment of the high five function. This result suggests that our online high five function was much more enjoyable than just pretending high five and there was not much difference among the three high five methods in terms of enjoyment.

6. Summary

We proposed and developed a system that enables participants' interactive synchronized gestures during video conferencing to increase non-verbal communication. The developed system supports the high five gesture since it is considered effective and feasible.

The user study of the three methods of high five functions was conducted. The free position method was evaluated better for its flexibility and easiness of use, while the fixed position method was evaluated better for its easiness of hand matching. The hybrid method compromising the former two methods had no notable disadvantages and is consequently applicable in real situations.

Since there were several comments that this system was "closer to a face-to-face encounter through a sense of togetherness," we think that it is preferable to aim for a system that enables a sense of togetherness even at a distance.

References

[1] Japan Football Association: Implementation of "Remote high five with Samurai Blue!" https://www.jfa.jp/samuraiblue/news/00025513 (accessed Jan. 2023).

[2] A. Curioni, G. K. Knoblich, N. Sebanz, L. M. Sacheli: The engaging nature of interactive gestures, PLoS ONE, Vol. 15, No. 4, e0232128. https://doi.org/10.1371/journal.pone.0232128 (2022).

[3] O. Morikawa, T. Maesako: HyperMirror: toward pleasant-to-use video mediated communication system, Proceedings of ACM CSCW '98, pp. 149-158 (1998).

[4] K. Komazaki, J. Watanabe: "Remote high-five" achieved by vibrotactile transmission: case studies for family communication of athletes and sports viewing with hearing-impaired person, Transactions of Virtual Reality Society of Japan, Vol. 27, No. 1 pp. 2-5 (2022). (In Japanese)

[5] J. E. Grønbæk, B. Saatci, C. Griggio, C. N. Klokmose: MirrorBlender: supporting hybrid meetings with a malleable video-conferencing system, Proceedings of ACM CHI 2021, pp.1-13 (2021).

[6] A. Chidambaram, A. He, A. Dattamajumdar: Video gesture recognition and overlay, https://github.com/ashwatc/Video_Gesture_Overlay (accessed Jan. 2023).

[7] K. Oyama, R. Morimoto, H. Shiozawa: Consideration on a video conferencing system supporting participants' synchronized gestures, IPSJ Groupware and Network Services Workshop 2021, pp.35-36 (2021). (In Japanese)

[8] K. Oyama, H. Shiozawa: Video conferencing system supporting participants' synchronized gestures, IPSJ DICOMO 2022, pp.1041-1049 (2022). (In Japanese).