Development of Video Chat System
Based on Space Sharing and Haptic Communication

Takahiro Hayashi[1][*] and Keisuke Suzuki[2]

[1]Faculty of Informatics, Kansai University, 2-1-1 Ryozenji-cho, Takatsuki-shi, Osaka 569-1095, Japan
[2]Graduate School of Science and Technology, Niigata University,
8050, Ikarashi-2-no-cho, Nishi-ku, Niigata-shi, Niigata 950-2181, Japan

Video chat is a communication tool widely used by people who live in distant locations. However, there are some differences between video chat communication and real world communication. To enhance the reality of video chat communication, we propose a novel video chat system that enables space sharing and haptic communication. To actualize these functions, the system extracts human regions from video camera images and synthesizes the regions onto a common background image. In addition, the system gives haptic feedback by activating the vibration of a smart watch the user wears when a collision occurs in the virtual shared space. We empirically evaluated the system to confirm the effectiveness and limitation of the functions of space sharing and haptic feedback. From experimental results, we confirmed that users were able to enjoy communication with the space sharing function. In addition, owing to the haptic feedback, users were able to naturally communicate with others in the virtual shared space.

1. Introduction

Video chat is a communication tool widely used by people who live in distant locations. There are some differences between video chat communication and real world communication. In real world communication, people are in the same space and sometimes use haptic communication such as touching. In contrast, in video chat communication, users are in different locations. Haptic communication cannot be used.

For the purpose of enhancing reality in video chat communication, in this study, we proposed a novel video chat system that virtually enables space sharing and haptic communication. Using the technique of image synthesis, the proposed system provides users with the feeling of being in the same space with the remote user. In addition, using a device having a vibrating interface such as a smart watch, the system provides haptic feedback.

The rest of the paper is organized as follows. In Sect. 2, we describe related work. In Sect. 3, we propose a video chat system that has the functions of space sharing and haptic feedback. In Sect. 4, we empirically evaluate the effectiveness and limitations of the proposed system. Finally, we conclude the paper in Sect. 5.

2. Related Work

In the field of human–computer interaction (HCI), some pilot studies for actualizing a shared space between remote users in video chat have been reported.[1–6] To provide the feeling of being in the same space to remote users, HyperMirror[5] uses an image processing technique for projecting the figures of remote users onto another background image. For further enhancing the reality in remote communication, OneSpace[6] generates a shared shape where objects in each room and users are overlaid by considering depth information measured by Kinect sensors. The main focus of these systems is to generate a virtual shared shape between remote users based on image synthesis techniques. These systems, however, do not consider haptic communication.

Haptic communication is a kind of nonverbal communication that plays an important role in facilitating natural communication.[7–9] HugMe[9] is a system that introduces haptic feedback to video chat. The system uses a touch screen and a jacket with embedded vibration devices. Remote users wear the jackets. When a user touches a particular part of the body of the remote user on the screen, the corresponding vibration device of the jacket of the remote user is activated, resulting in the remote user noticing that he/she has been touched.

Despite the importance of shared space and haptic feedback in remote communication, to the best of our knowledge, there is no video chat system incorporating these two functions at the same time. In this study, we combine the idea of shared space and the idea of haptic feedback to augment the reality of remote communication with video chat.

3. Proposed System

3.1 Hardware components

Figure 1 shows the components of the proposed system. As shown in the figure, the proposed system is composed of a PC, a web camera, a microphone, and a smart watch. The same components are also installed in the locations of remote users. In the system, the smart watch is used as a vibrating device. The smart watch and the PC are connected by Bluetooth. Between remote PCs, visual and audio data are exchanged in real time.

From the camera images, human regions are automatically extracted. The regions of users are projected onto a common background image to generate a shared space. The users can communicate with others by seeing the shared space via the PC screen as in HyperMirror. When a user touches the conversation partner in the shared space, haptic feedback is given to the two users via the smart watches.

3.2 Software modules

The proposed system is composed of two main software modules. One is a module for generating a shared space, and the other is a module for providing haptic feedback.
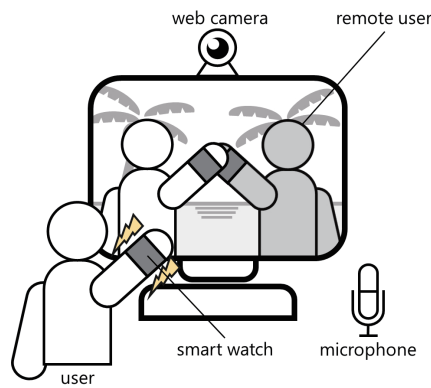
Fig. 1.   (Color online) Components of the proposed system.

In the module for generating a shared space, the following four steps are executed in turn:
(1) camera capturing,
(2) human region extraction,
(3) information exchange between remote PCs, and
(4) image synthesis.

Figure 2 illustrates how the video camera images are processed, exchanged, and used for generating a shared space by the four steps.

In the camera capturing step, images are obtained from the web camera. In the proposed system, the image size of the web camera is 640 × 480 pixels and the framerate is 30 frames per second (fps).

In the human region extraction step, the human regions in each frame of the camera images are extracted in real time. To extract human regions, the proposed system utilizes the technique of background subtraction[10] and combines the average filtering, median filtering, and image thresholding. We have experimentally confirmed that the human region extraction step can be completed without delay when using an Intel Core i7 6700 CPU. The details of this step are explained in Sect. 3.3.

In the information exchange step, the human regions are exchanged between remote PCs. To ensure real-time communication, the system requires a 100 megabits per second (Mbps) network speed. In addition, to avoid the delay in transmission, the system adopts the user datagram protocol (UDP) for the information exchange between PCs. UDP is suitable for real-time applications because of its lack of transmission delay. In the image synthesis step, the user regions are projected onto one common background image. The synthesized image is displayed on the PC screen.

In the module for providing haptic feedback, the smart watch vibrates when one user touches the conversation partner in the shared space. The purposes of this module are to detect the collision between users in the shared space, to send a trigger signal to smart watches for activating the vibration, and to provide haptic feedback to the users. The details of collision detection in the shared space are described in Sect. 3.4.
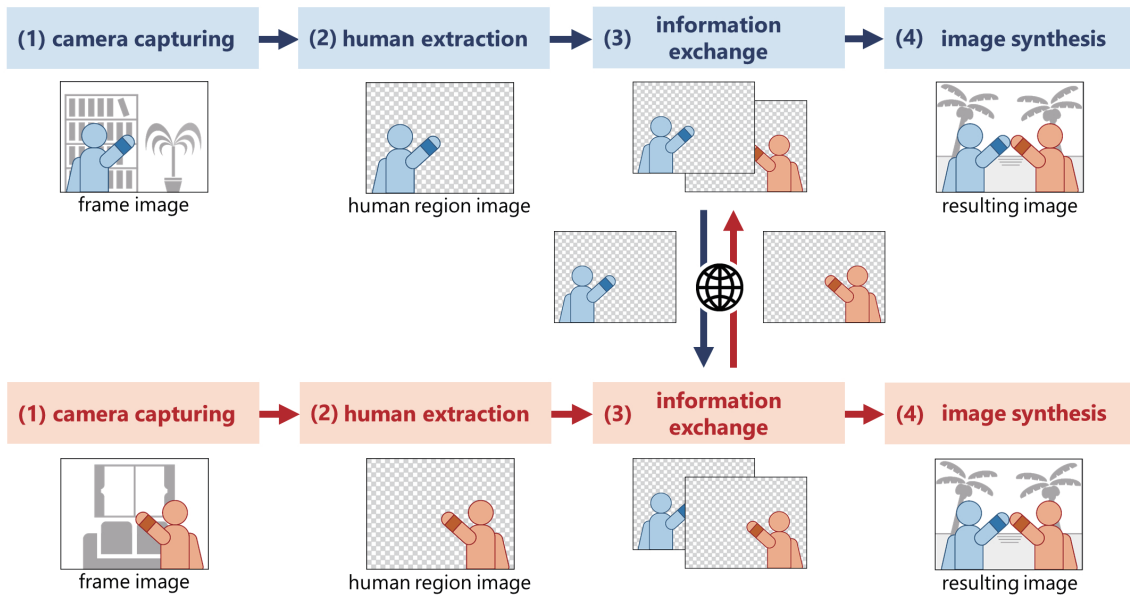
Fig. 2.　(Color online) Computational steps for generating a virtual shared space.

3.3　Extraction of human regions

The human region in a camera image is extracted by utilizing the technique of background subtraction.[10]　Background subtraction is an image processing method for detecting objects from the subtraction image between the input image (camera image) and a reference image (background image).

In the proposed system, a camera image of background alone is used as the reference image. To reduce the noise spikes occurring in camera images, a 3×3 averaging filter is applied to the reference image and the input image in advance.

By subtracting the reference image [Fig. 3(a)] and the input image [Fig. 3(b)], a subtraction image is generated [Fig. 3(c)].　By applying image thresholding to the subtraction image, a thresholded image [Fig. 3(d)] is generated.　Generally, the thresholded image has small isolated regions in the background and a small hole in the human region.　Next, to delete these small isolated regions and fill the small holes, a 35×35 median filter is applied to the thresholded image.　As a result, a mask image [Fig. 3(e)] can be obtained.　Finally, by applying the mask to the input image, the human region is extracted from the input image [Fig. 3(f)].

Background subtraction requires static background for extracting user regions accurately. Therefore, we assume that the system is used in an indoor scene where there is no moving object except for users.

As mentioned, in the human region extraction step, the four kinds of simple image processing algorithms, i.e., (1) average filtering, (2) image thresholding, (3) median filtering, and (4) masking, are executed.　The computational speeds of these algorithms are fast enough that the proposed system handles camera images at a speed of 30 fps.
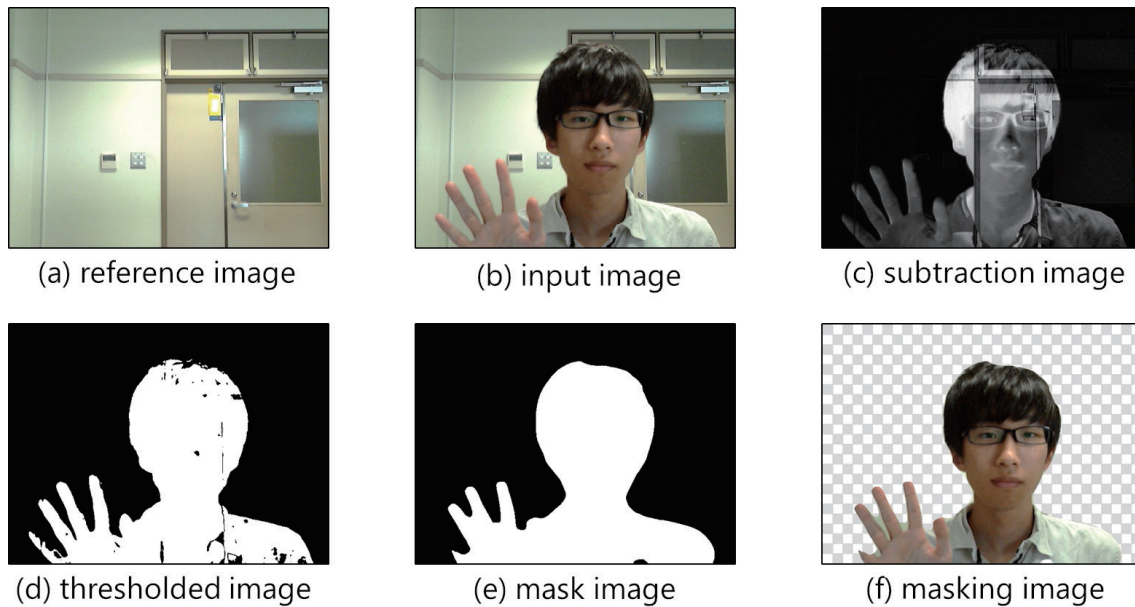
(a) reference image   (b) input image   (c) subtraction image

(d) thresholded image   (e) mask image   (f) masking image

Fig. 3.   (Color online) Human region extraction.

3.4   Collision detection in shared space

When a user touches a conversation partner in the shared space, a vibration is given to the two users. To detect a collision in the shared space, the proposed system calculates the logical conjunction between the mask images of two users. By the logical conjunction, the area of overlap between the two users in the shared space image can be obtained. Figure 4 shows the overview of collision detection. In the figure, S is the area of overlap between the two users in the shared space, and D is the duration over which overlaps occur. When the area of the overlapped region is wider than a particular threshold (S_T) and the overlap lasts for a particular duration (D_T), the system judges that a collision has occurred.

4.   Experimental Results

4.1   Method

To confirm the effectiveness and usability and clarify the problems and limitations of the proposed system, we conducted an experiment. In the experiment, we compared three kinds of video chat systems:
(S1) a video chat system having the functions of shape sharing and haptic feedback,
(S2) a video chat system having the function of space sharing, and
(S3) a video chat system having neither the space sharing nor haptic feedback functions.

Here, S1 is the proposed system, S2 is a conventional video chat system such as HyperMirror, and S3 is an ordinary video chat system such as Skype.
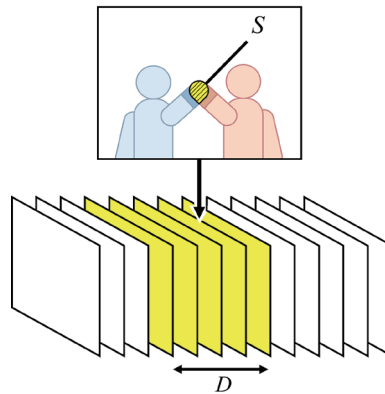
Fig. 4.   (Color online) Detection of a collision.

These systems are implemented on PCs having Intel Core i7 6700 CPUs. The PCs in the two rooms were connected with a gigabit Ethernet. The image size of the web cameras is 640 × 480 pixels and the framerate of the cameras is 30 fps.

Twelve subjects (six pairs of users) participated in the experiment. The two users in each pair were in different rooms in the same building. In the experiment, the subjects played a card game such as "Old Maid" using the provided video chat system. The game proceeds according to the following steps:

1. One Joker, 5 numbered cards (10, J, Q, K, and A) were dealt to each player.
2. The first player discards the Joker.
3. The second player spreads his/her cards face down and offers them to the first player. The first player selects a card from the second user's hand. After the first player selects a card, the second player shows the card face to the first player via the chat screen. The second player removes the card selected by the first player. The same card is removed from the first player's hand if the selected card is a number card. If a Joker is selected, a Joker is added to the first user's hand.
4. Exchanging the roles of the first and second players, Step 3 is repeated.
5. Steps 3 and 4 are repeated until no cards remain in one player's hand.

Figure 5 shows screen shots of the proposed video chat system (S1) where users are playing the card game. When using the proposed system, haptic feedback is given to the users when one user picks up a card from the conversation partner on the screen.

To evaluate the effectiveness and limitations of space sharing and haptic feedback in the proposed system, we conducted surveys by questionnaire after the users finished the card game. The questionnaire composed of nine questions is shown in Fig. 6. The first four questions (Q1–Q4) are the common questions for the three systems (S1–S3). The next two questions (Q5, Q6) evaluate the function of space sharing. The next two questions (Q7, Q8) evaluate the function of haptic feedback. The last question (Q9) is an evaluation of the combination of the space sharing and haptic feedback functions.

Fig. 5.   (Color online) Two users playing a card game using the proposed video chat. When one user picks up a card from the other user on the screen, haptic feedback is given to the users.

Please circle the number that represents how you feel about the video chat system.

		strongly disagree	disagree	neither disagree nor agree	agree	strongly agree
Q1	You did not feel noticeable latency in video transfer.	1	2	3	4	5
Q2	You did not feel noticeable latency in audio transfer.	1	2	3	4	5
Q3	You had a good communication.	1	2	3	4	5
Q4	You felt the conversation partner was in the same space.	1	2	3	4	5
Q5	The results of image syntheses were natural.	1	2	3	4	5
Q6	You felt the space sharing function was useful for stimulating conversation.	1	2	3	4	5
Q7	The timing of haptic feedback was appropriate.	1	2	3	4	5
Q8	You felt the haptic feedback function was useful for stimulating conversation.	1	2	3	4	5
Q9	You felt the combination of the space sharing and haptic feedback functions was needed for better communication.	1	2	3	4	5

Fig. 6.   Survey questions.

4.2   Results

Table 1 shows the average scores of the survey results. Comparing the results of Q1 and Q2 between the three systems, we can confirm that noticeable delay did not occur in the three systems. The performances of S1 and S2 are comparable to the performance of S3 (the ordinary

Table 1
Results of the survey.

	(S1) space sharing + haptic feedback		(S2) space sharing		(S3) ordinary video chat	
	mean	SD	mean	SD	mean	SD
Q1	4.17	0.83	4.42	0.51	4.17	0.94
Q2	4.33	0.89	4.58	0.51	4.50	0.52
Q3	3.83	1.03	3.92	1.00	3.67	0.98
Q4	4.75	0.62	4.42	0.67	3.83	0.72
Q5	3.92	0.79	4.00	0.85	2.08	1.38
Q6	3.50	1.00	3.67	0.65	—	—
Q7	4.42	0.67	4.25	0.62	—	—
Q8	2.42	1.00	—	—	—	—
Q9	3.08	1.08	—	—	—	—

video chat system). These results indicate that the computational cost of the space sharing is not an issue for real-time communication.

The results of Q3 show that the systems having space sharing function, i.e., S1 and S2, are clearly superior to the ordinary video chat system, i.e., S3. These results indicate that the space sharing function is effective in assisting users to have better communication than ordinary video chat. The results of Q4 also show that, with the space sharing function in systems S1 and S2, the users feel as if their conversation partners are in the same space. The results of Q5 and Q6 indicate that the feeling of being in the same space influences communication positively.

The result of Q7 shows that the timing of haptic feedback is not always appropriate for the users. From interviews conducted after the questionnaire surveys, we found that overdetection of collision in the shared space frequently occurred. Owing to the overdetection of collision, unnecessary haptic feedback was given to the users. As a result, Q7 was negatively evaluated. The main reason for the overdetection of collision is the existence of noise in the mask images generated in the human region extraction step (Sect. 3.3). In this system, various image processing algorithms were used for human region extraction. However, noise cannot be perfectly deleted. To reduce unnecessary haptic feedback, the sensitivity of collision detection should be controlled. As explained in Sect. 3.4, the sensitivity of collision detection is determined by two threshold values, S_T and D_T. By raising these values, collision detection can be controlled. Finding the optimal setting of these values remains for future work. However, although the parameter tuning of haptic feedback in the experiments is a problem, the results of Q8 show that haptic feedback is a promising approach for stimulating conversation.

The result of Q9 shows the necessity of the space sharing and haptic feedback functions for enhancing the reality of video chat. This result indicates that the proposed system can enhance reality in video chat communication.

5. Conclusions

For the purpose of enhancing reality in video chat communication, we proposed a video chat system enabling space sharing and haptic communications. From the experiments, we confirmed that users were able to enjoy communication with the space sharing function.

In addition, owing to the haptic feedback, users were able to naturally communicate with others in the virtual shared space. Controlling the sensitivity of collision detection to prevent unnecessary haptic feedback remains for future work.

There are many possible scenarios in which the proposed system can be effectively used. For example, in an aging society like Japan where the percentage of nuclear families has been increasing, it is necessary to encourage the elderly to communicate to prevent social isolation. The proposed video chat system can contribute to the purpose of helping communication between the elderly and their children or grandchildren who live in distant locations. As future work, we would like to confirm the effectiveness of the proposed system from the practical point of view.

References

1   S. E. Hunter, P. Maes, A. Tang, K. M. Inkpen, and S. M. Hessey: Proc. 32nd Annu. ACM Conf. Human Factors in Computing Systems (2014) 1197.
2   J. Zillner, C. Rhemann, S. Izadi, and M. Haller: Proc. 27th Annu. ACM Symp. User Interface Software and Technology (2014) 471.
3   A. Kunz, T. Nescher, and M. Kuchler: Proc. Int. Conf. Cyberworlds (2010) 430.
4   R. Gumienny, L. Gericke, M. Quasthoff, C. Willems, and C. Meinel: Proc. Int. Conf. Computer-Supported Cooperative Work in Design (2011) 47.
5   O. Morikawa and T. Maekawa: Proc. ACM Conf. Computer-Supported Cooperative Work (1998) 149.
6   D. Ledo, B. A. Aseniero, S. Greenberg, S. Boring, and A. Tang: Extended Abstracts Human Factors in Computing Systems (2013) 997.
7   H. Nakanishi, K. Tanaka, and Y. Wada: Proc. ACM Conf. Human Factors in Computing Systems (2014) 2143.
8   I. Ahmed, V. Harjunen, G. Jacucci, E. Hoggan, N. Ravaja, and M. M. Spapé: Proc. 18th ACM Int. Conf. Multimodal Interaction (2016) 341.
9   L. Zhang, S. Jamal, and E. S. Abdulmotaleb: Multimedia Tools Appl. **74** (2015) 5489.
10   M. Piccardi: Proc. IEEE Int. Conf. Systems, Man and Cybernetics **4** (2004) 3099.