情報処理学会研究報告
IPSJ SIG Technical Report

Vol.2010-EC-16 No.22
2010/5/15

HANDY：ソーシャルプレセンスを向上させるビデオ会話システム

IGOR DE SOUZA ALMEIDA† JORDI CARRES POLO†
MARINA ATSUMI OIKAWA† 平松達也† 天野敏之†
藤澤誠† MARK BILLINGHURST† 加藤博一†

既存のビデオ会話システムは，対面でのコミュニケーションを模倣しようとしている．その目的は，自分が遠方にいる話し相手と一緒にいるんだと感じれるようなソーシャルプレセンスの感覚を実現することにある．我々は，ビデオ会話システムにおいて，このソーシャルプレセンスを向上させるために，一方のユーザの存在感を他方のユーザの映像の中に感じさせる方法を提案する．実際にこのアイデアに基づく試作システムを構築し，試験的な評価実験を行ったので，それについて報告する．

HANDY: Enhancing Social Presence in Video-chat Systems

IGOR DE SOUZA ALMEIDA† JORDI CARRES POLO†
MARINA ATSUMI OIKAWA† TATSUYA HIRAMATSU†
TOSHIYUKI AMANO† MAKOTO FUJISAWA†
MARK BILLINGHURST† HIROKAZU KATO†

Current video-chat systems try to copy face-to-face communication. They aim to create a sense of social presence, which is the feeling that one person is present with another person in a mediated environment. We propose an approach to enhance social presence of video-mediated systems by simulating the presence of one user into the second user's video image. A prototype was implemented and a pilot study was conducted.

1. INTRODUCTION

Video chatting systems can be considered one of the most popular communication channels used between remote people, implemented by popular software such as Skype or Yahoo Messenger. In these systems, typically two windows are presented to the users: one to display the user's own video image and another one to show the remote person.

The affordable status of this kind of interface has attracted a large amount of users, allowing them to enjoy it with minimum settings. However, traditional video-chat systems lack the same interaction and sense of presence as face to face conversation, which presents a range of natural non-verbal communication cues including point in space, gaze direction or proximity behavior [1]. For instance, let's say during Mother's Day, a user is video chatting with their mother. Even though they can listen and see each other, the user may wish that they could touch her or give her a gift, just like in a real world meeting.

Such scenario shows the need of a space where both users could naturally go in and out and feel like they co-exist. Various researchers have explored ways of creating this kind of environment, a *shared space*, similar to collaborative virtual environments (CVE) in which both participants and their information shares a common display space [2]. It has been proved that using a dedicated shared space environment improves the results of certain tasks, such as layout design, over not having it [3, 4].

In HyperMirror [5], a shared scene is created with images of remote users. The final result displays the users as if they were in front of a mirror (front views captured by cameras) sharing a common world. However, users never face each other, since they keep the focus on the screen and even if they try to simulate looking at each other, they will lose their sense of presence.

Feedback is also an important factor, as implemented by Lazy Susan [6], a virtually-shared physical tool based on the rotation of the users' chairs to enhance the sense of connectedness among remote people. However, the body action used in this approach requires an unusual interaction from each user, since they need to perform "chair rotation" in order to give/receive feedback from other user as a sense of presence.

Like Lazy Susan, an updated version of HyperMirror included a feedback device [7] with a sensor located on the user's shoulder for sensing partners with vibration. This vibration can be produced intentionally, but it is usually created spontaneously, when the user's bodies cross on HyperMirror.

† 奈良先端科学技術大学院大学情報科学研究科
Graduate School of Information Science, Nara Institute of Science and Technology

We intend to combine the user's gesture image with other user's environment image into a common display window, creating a shared space environment naturally as the users interact. At the same time, this space is used to enhance the sense of social presence, defined as the salience of the partner in a mediated communication and the consequent salience of their interpersonal interactions [8].

2. HANDY SYSTEM

Figure 1 shows a comparison between the traditional video-chat system and our proposed system, referred to as HANDY. In the former case, the user has one camera capturing his face image which is showed beside the remote user's face image. Users are restricted to interact in their own environment or watch the remote user environment.

In our proposed system, the main idea is to minimize the feeling of being geographically separated by allowing interaction of user A inside the video image of user B. It is showed the users playing *rock-paper-scissor* using HANDY in which it is possible to see user A's hand appearing in user B's video image, as if they were in the same environment.

Figure 1 Traditional video chat system (left) and HANDY system (right) components.

2.1. Implementation

In the system prototype showed in Figure 2, each user has two cameras in his or her own environment: one placed towards their face, in the same position used in conventional video-chat system and another camera is beside the user capturing the area where the users can place their hand or objects they want to appear in the remote user video image. This area basically consists of a monochrome static background and will be referred to *shared space* along this text.

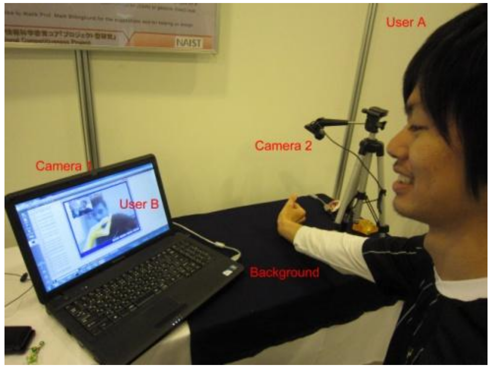

Figure 2 System Configuration: Camera 1 captures the face image of User A and Camera 2 points to the shared space (area with the monochrome background). Using HANDY, the face image of User B is showed in the computer's screen merged with the hand's image of User A, which is not possible in traditional video-chat systems.

In this system, first the connection between the computers of user A and B is established and the face camera video image is sent to both sides. A training phase is performed with the second camera using the first n initial frames containing the monochrome background.

The training step is necessary to allow the segmentation of foreground and background pixels from an incoming frame according to a threshold value t, distinguishing the pixels from the background to the non-static objects placed in the shared space, for example, the user's hands or objects (Figure 3).

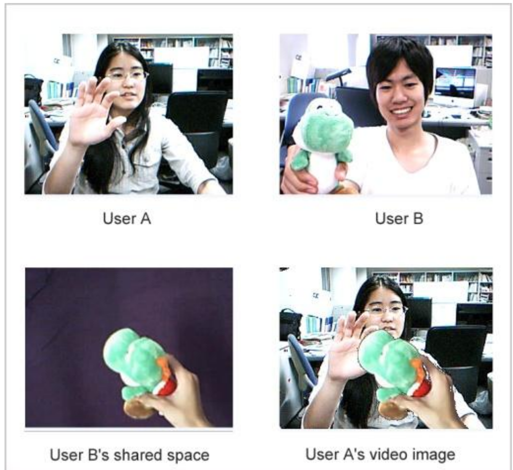

Figure 3 Placing an object in the remote environment.

After this segmentation, the background is subtracted and the image parts of interest are merged with the remote user's video image in real time. If nothing is placed in the shared space, the video image of the remote user appears as in a usual video chat system.

OpenCV library was chosen to deal with the capture and merging of local and remote video images and *Qt* library was used to develop the entire communication architecture for the dual video streams.

3. EVALUATION

The system was evaluated according to a summative usability approach [9] in which the goal is to evaluate how well the system's functionality meets its objective. A user study was designed in order to compare using a traditional video conferencing system to one that allows object placement in the remote environment. Initially, it was hypothesized the following:

- Interaction in traditional video-chat systems is limited and physical barriers between the local and remote users are still very noticeable;
- Human-human communication in video-chat systems can be enhanced by simulating physical presence of one user into a second user's real world;

The chosen metrics were defined to measure different degrees of satisfaction of the system in terms of the user's perception. In other words, the participants were asked to report about their experience while using the system in different conditions and tasks.

Self-reported metrics were used through after-scenario questionnaires, given to each participant after completing each experimental task. Answers to each question was given on a Likert Scale from 1 to 7, where 1 = easy/fun and 7 = not easy/not fun. The attributes assessed were:

- Q1: Easiness: *how easy was the system to use?*
- Q2: Enjoyment: *how fun was the experience with the system?*
- Q3: Ease of communication: *how easy was it to communicate with the other person?*
- Q4: Intuitiveness: *how natural was it to use the system?*
- Q5: Ease of understanding: *how easy was it to understand the other person?*
- Q6: Closeness: *how close to a face-to-face communication was the experience with the system?*

The user study was divided into two sections: communication and gesture sections. The former consists of tasks that intentionally stimulated the communication between participants. The latter focused on the gestural aspect of the collaboration between them.

Two conditions were defined for each task:

Condition 1: HandyOff - without the proposed shared space. Essentially, it simulated a traditional video-chat system.

Condition 2: HandyOn - the proposed shared space became available.

Using as reference the tasks proposed by [1], one communication task and two gesture tasks were defined as follow:

a. Communication Section:

a.1. Survival game: It is a classic team building activity in which participants were asked to discuss their choices and come to a consensus. Two scenarios were used, one for each condition: *Plane crash* scenario for the condition HandyOff and *Lost At Sea* for the condition HandyOn [10].

b. Gesture Section:

b.1. Paper-rock-scissor game: This is a gesture based game played by two or more people where pre-defined ranked gestures are used to determine who wins or loses. Participants were asked to play three games of three rounds each for each defined condition.

b.2. Giving a gift: Participants were given a real object and told to describe it and find the best way for the person on the other side to understand how it works. Descriptions included the size, weight, color, shape and how to use the object.

The pilot study was conducted with eight Japanese graduate students, collaborating in pairs. Each session took about 30-40 minutes, in which each pair spent about 25 minutes performing communication and gesture tasks. Additional time was spent filling out after-scenario questionnaires and with subject interviews.

4. RESULTS FROM USER EVALUATION

Following completion of the each task users filled out a subjective survey asking them to respond to Likert Scale questions on a range from 1 to 7, 1 = easy, 7 = not easy. A two factor (task, Handy On/Off) ANOVA test was performed on the results.

We found no significant differences between conditions, except for Q2: Enjoyment ($F(2,42) = 4.072$, $P < 0.05$) and an almost significant result for Q3: Ease of Communication ($F(2,42) = 4.072$, $P = 0.058$). The average scores for these questions are shown in Figures 4 and 5. They show that there is a difference in user scores for the gesture tasks between the HandOff and HandyOn conditions. For these tasks, users thought the HandyOn condition was more fun and made it easier to communicate.

In addition to conducting a user survey we also asked users for their comments about the system. The comments reflected the results obtained in the questionnaires: most of the users agree that when playing the paper-rock-scissor game it was more fun and easy when using HANDY system.

Other comments referred to the position of the second camera. Ideally, this second camera should be placed on the user's back and in a height close to the user's viewpoint. However, due to limitations when using the background subtraction (training stage in the beginning) and since it was not a comfortable setup to the user, the camera was placed beside the user.

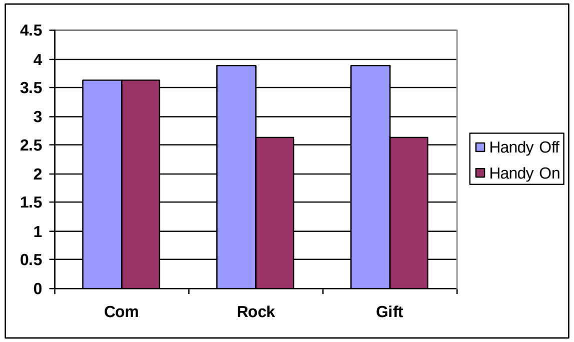

Figure 4 Q2 Enjoyment Results (1-7).

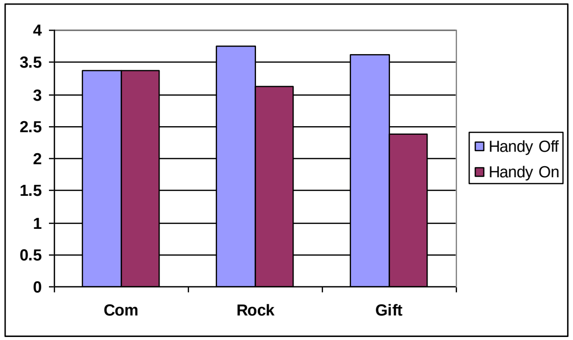

Figure 5 Q3 Ease of Communication Results (1-7).

5. DISCUSSION

During the communication section of our experiments, when the condition HandyOn was given to the participants, they would place the hand inside the shared space because they were required to do it. However, they would not use it or even move the hand whatsoever what may have been one factor for the lower score of our system.

Furthermore, the *Survival Game* may not have been the best choice of task since it changed the focus of the participant from the computer screen to the paper in which the list of items was printed out whether looking or pointing at it.

During the gesture section, while performing the paper-rock-scissor task, the orientation of the window caused the user to adapt gestures to avoid occlusion of hands.

In the *Giving Gift* task, users were asked to play with the same items in both conditions. We think that a new set of items for every new condition could refresh the mindset of the participants and make it less tiresome.

Some modifications were identified as needed, for example, the position of the remote user's window which was placed inside the local user's window, sometimes occluding gestures or the face. In addition, the size of the window was a constant complaint from participants who would often try to resize it. A bigger window (local and remote) could increase the fun and enjoyment.

6. FUTURE WORK

There are at least two main future development directions for this project. The first one is to improve the current software implementation and user experience.

It was noticed through the user evaluation that some details like placement or size of windows are important to the user. Furthermore, in order to get additional real-world testing, it would be important to minimize the amount of data that is being sent for each frame. For instance, this could be reduced by sending just the differences between frames, instead of always sending individual images.

The communication infrastructure also needs support for advanced networking management, like proxies, two ways signaling system and better flow control. Additional functionalities could improve communication and interaction between users, such as possibility of taking snapshots or sharing desktop windows.

The second development direction would be creation of new experiences. One idea worth exploring would be creating a shared space for the second camera only. By merging the second camera streams, a space equivalent to looking down to one's own hands can be created. This space can be purely virtual or it can be created by merging the video image of both HANDY's camera stream.

In a purely virtual shared-space, the background and interactive objects can be generated and they can react to movements of the hands of both sides of the conversation. When one stream is merged onto another, one hand can work on a real-world problem or game while the partner can give vocal and gestural instructions on the task.

Gaze detection can be used to further control how to merge the images. Gaze detection software could be used to know in which screen the user is currently focusing their attention, merging the hand to the screen the user is focusing on. This would give to the user a further feeling of interaction in a 3D space and control of their presence in the video.

Also, object detection can be used, depending on the object and its movements events can be created, plugins would hook on those events for creating new experiences.

As mentioned before, merging the secondary camera stream with the remote primary image stream creates a different experience compared to merging streams from both sides of the image from the second camera.

After implementing these modifications, new user studies would be necessary, not only focusing on communication and gestural tasks but also in task-space collaboration.

REFERENCES

1) Hauber, J.: Understanding Remote Collaboration in Video Collaborative Virtual Environments. PhD thesis. University of Canterbury (2008).
2) Benford,S.et al.: Shared Spaces: Transportation, Artificiality, and Spatiality". In: Proceedings of CSCW'96, pp.77-86 (1996).
3) Olson, J.S. and Olson, G.M. What Mix of Video and Audio is Useful for Small Groups Doing Remote Real-time Design Work? Proceedings of the SIGCHI conference on Human factors in computing systems table of contents, pp. 362 – 368 (1995).
4) Whittaker, S.; Geelhoed, E. and Robinson E. Shared workspaces: how do they work and when are they useful? Proceedings of International Journal of Man-Machine Studies archive, pp. 813 – 842 (1993).
5) Morikawa, O. et al.: Hyper Mirror: a video conversation system which can point at object in partner's space. The 50th Technical Committee on Pattern Measurement Conference, pp382-387 (2000).
6) Wesugi, S. and Miwa, Y.: "Lazy Susan" Communication System for Remote, Spatial and Physical Collaborative Works. In Proceedings of the First IEEE International Workshop on Horizontal Interactive Human Computer Systems (2006).
7) Morikawa, O.: HyperMirror-III: sensing the movements of partners with vibration. ACM Conference on Computer Supported Cooperative Work (2002).
8) Sallnas, E.-V. Effects of Communication Mode on Social Presence, Virtual Presence and Performance in Collaborative Virtual Environments. In Presence, vol. 14, n.4, pp. 434-449 (2005).
9) Tullis, T. and Albert, B.: Measuring the user experience: collecting, analyzing and presenting usability metrics, Interactive Technology Series, Morgan Kaufmann (2008).
10) Team Building Activities, Initiative Games, & Problem Solving Exercises. Available at: http://wilderdom.com/games/InitiativeGames.html.