Shopping Together: A Remote Co-shopping System Utilizing Spatial Gesture Interaction

Minghao Cai[1(✉)], Soh Masuko[2], and Jiro Tanaka[1]

[1] Waseda University, Kitakyushu, Japan
mhcai@toki.waseda.jp, jiro@aoni.waseda.jp
[2] Rakuten Institute of Technology, Rakuten, Inc., Tokyo, Japan
so.masuko@rakuten.com

Abstract. In this paper, we introduce a remote co-shopping system– *Shopping Together* for two geographically separated people: an in-house user who remains in a house and an in-store user who goes to a shopping place. We support the two of users to achieve a real-time two-way spatial gestural interaction in the physical shopping world where the in-store user stays, with an attention awareness subsystem to enhance a common feeling. The in-house user accesses to the remote shopping venue with an immersive shopping feeling, while the in-store user experiences an augmented reality feeling. Through our system, users could accomplish a shopping task together and share a *Shopping together feeling* which means a sensation that they are going for purchase together in the same space.

Keywords: Immersive shopping · Remote communication
Gestural interaction

1 Introduction

With the rapid developments of reliable network service and telecommunication techniques, it is convenient to achieve a low-delay and high-quality video conferencing with light setups. This has provided an existence proof for the feasibility of a remote communication with people far apart. Having the possibility to reduce the perception of spatial separation and strengthen the connection of the participants to some extent [1], however, existing commercial video communication systems are still not satisfactory enough to support a feeling of being together. They mostly only provide a capture of users face. It helps little to directly reveal the other information like body language or the ambient.

People often intend to gesture with their hands while they speak. Those gestures might be particularly well-suited for the representation and transmission of information about spatial location or performing actions with objects [2]. With only verbal description, it might be challenging for users while they want

M. Kurosu (Ed.): HCI 2018, LNCS 10903, pp. 219–232, 2018.
https://doi.org/10.1007/978-3-319-91250-9_17

to access to a wide variety of information about the world, some of which does not readily lend itself to representation in language.

For example, imagine making a video call with your friends in distance, asking for a favor to buy a local specialty in the market. Rather than just staring on the screen and using some scanty expressions like "that one", "this one" or "over there", it might more intuitively and conveniently that you could point out your satisfactory selection directly to your friends, which might make the talk smoother and more meaningful. Although might possible with current technologies, an effective way for users to achieve gestural communication is offered by few communication platforms. Such constraints make it difficult for users to get a common perception or achieve smooth interaction.

In this paper, we introduce our remote communication prototype for a co-shopping scenario in which the communication always involves the environment and objects. What this research targeting is to offer a sensation shared by the two geographically separated users that they are co-located together side by side going for a shopping in the same world. We define it as a *Shopping together feeling*. Although it might require varieties of aspects to fully realize such sensation, in this prototype we intend to support users to accomplish a shopping task together like they are co-located side by side so as to make users feel a close connection and aware a certain extent of "togetherness".

2 Research Approach

Shopping Together is designed for two users in separated places: an in-store user goes to a physical shopping world, such as a store, to which an in-house user staying in a house locating far apart accesses with an immersive shopping experience. The in-house user might be someone elderly who has mobility problems or just have difficulties in reaching the remote places. He/she may ask the in-store user who might be one of his/her friends or family to buy something in the store.

To enhance the human-to-human interaction, we develop an effective two-way gestural communication approach which could be used in a mobile condition. This system allows the in-house user to perform a free hand gestural input, without limitation of hand postures, in the immersive shopping world with a first-person perspective. His/her precise gestures would present to the in-store user with a side-looking perspective. On the other hand, a capture of the in-store user's hand gestures is easily accessed by the in-house user. Additionally, we support the in-house user to use practical control functions with hand gestures in order to improve user's observation ability and enhance the immersive feeling.

Different from traditional video communication techniques, this system allows the users have free manipulations of independent viewpoint. We construct a virtual shopping environment, in which the in-house user gets a 360° panoramic view of the physical venue and simply control the viewpoint by head movement. It gives a feel of being personally in the scene. For the in-store user,

we introduce the augmented reality technique. With the use of a pair of smart glasses, our system presents the 3D air gestures superimposing in the physical world, which still allows a clear view of the surrounding.

To reinforce the connection between users, *Shopping Together* has an attention awareness subsystem. By tracking and computing the head movements, users could easily share their attention conditions to improve a common feeling.

3 Shopping Together

3.1 Immersive Virtual Shopping

In traditional capture techniques, normal cameras have a limited capture angle which would restrict the user's view of vision. Even the wild-angle camera has some certain blind spots. In this case, always multiple combined cameras or an adjustable shooting direction of the camera is required if users try to access a panoramic viewing of the environment.

Fig. 1. Panoramic view of the shopping environment: the in-house user controls an independent viewpoint naturally by turning the head.

In this system, we introduce a spherical camera which could provide a high qualified 360° capture of the surrounding. The camera is carried by the in-store user and provides a real-time panoramic view to the in-house user. Wearing a head-mounted display (HMD), the in-house user accesses the real-time shopping venue with a 360° panoramic browsing, both vertical and horizontal without missed information. The in-house is provided an independent free viewpoint, manipulating the viewing direction naturally by head movements, just like one truly goes for a shopping (Fig. 1).

3.2 User Gestural Interaction

The two-way gestural interaction including a gestural input from the in-house user showing to the in-store user in an augmented-reality way, and a side-looking capture of the hand gestures of the latter in real time.

Gestures from In-House User. To extract the hand gestures of the in-house user and recreate them in the virtual environment, we use a depth recognition approach. Some previous research has shown that, to an extent, the depth-based hand gestures recognition has the advantage of accuracy and robust [3–5]. It allows no wearable or attached sensors on the hands while the in-house user making some gestural input freely, which extends the freedom and comfort. A compact depth camera is used to extract the real-time depth data of user's hands which includes not only the rotation and orientation of the user's fingers but also the subtle changes of their spatial positions.

We construct a pair of simulated hand models which have flexible joints and palms. By matching the real-time depth data to the models we built, the system reappears the free hand gestures of the in-house user in the virtual sightseeing with a certain extent of precision. The in-house user could see the own hand gestures with a first-person perspective in the shopping venue. Once the user changes the hand postures or moves the hands, the models change to match the exact same gestures almost simultaneously in the shopping venue (Fig. 2).

Fig. 2. The in-house user's view: user performs free hand gestures in the venue with a first-person perspective.

These gestures would be streamed to the remote partner. The in-store user uses an augmented-reality glasses to see the system information while still could see the surrounding clearly at the same time (Fig. 3). The gestures present on the left side of the field of vision superimposing on the physical world, showing with a side-looking perspective. We define such third-person perspective as Side-by-side Perspective which simulates watching the hand gestures of the partner from the side. It enhances the feeling of staying together while the in-store user still gets a good view of the physical world without being disturbed by the overlapping hand models.

Fig. 3. Visualization of the in-store user's view: user gets an AR experience and has a side-by-side perspective of the in-house user's gestures.

Gestures from In-Store User. As we have mentioned above, the in-house user uses the HMD to access a panorama of shopping environment both horizontally and vertically. Such panoramic view of the remote world also includes the view of the in-store user's hands and profile face. For example, as shown in Fig. 4, the in-house user simply turns the head, they could directly see the in-store user is making some guidance in the scenery (directing something on the shelf).

Fig. 4. The in-house user's view: seeing the real-time hand gestures and profile face of the in-store user in the physical world

3.3 Practical Gesture Control

By recognizing two designed gestures, *Shopping Together* supports the in-house user two functions to reinforce the immersive experience.

Observation. Shopping environment, such as a supermarket or a convenient store, usually is complex and contains a large number of products and objects. To assist the in-house user in observing things during the virtual shopping so as to enhance an immersive feeling, we design a 3D hand gesture recognition method for the in-house user–*Observation Gesture*. This spatial gesture could be used to magnify the field of vision in an adjustable scale (Fig. 5).

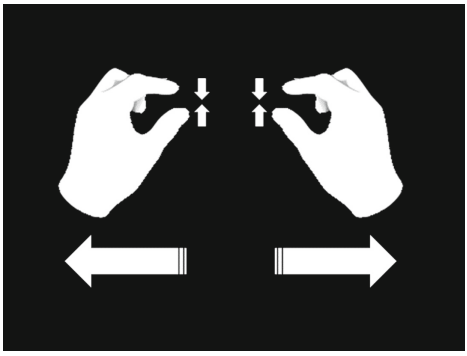


Fig. 5. Observation Gesture

As shown in the Fig. 6, the recognition method is: Once the user's both hands rise and keep pinching gestures, the gesture is activated. The system computes the change of the distance between two hands to adjust the scale of magnification. When the in-house user spread out two hands keeping pinches, the field of vision zooms in. When the user puts two hands close gradually, the field of vision zooms out accordingly.

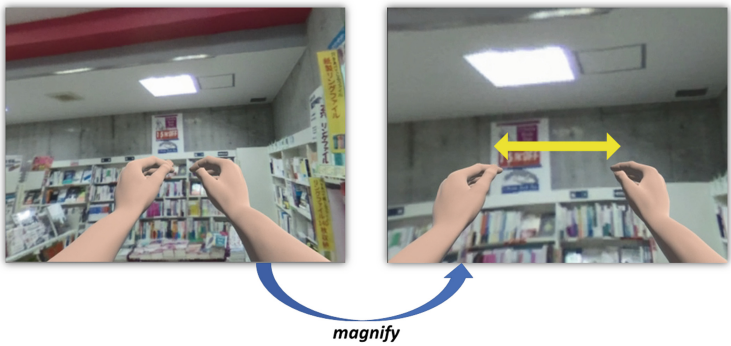


Fig. 6. The in-house user magnifies the field of vision.

Shooting. In this function, the in-house user simply makes a gesture to take a photo record of the current field of vision. For example, as shown in Fig. 7, a product interests the in-house user, and the user records by using the shooting gesture.

(a) Gesture method (b) Photo record

Fig. 7. Recognition method: Once the system recognizes hand the posture, the shooting gesture is activated. When two hands are close enough, a photo record would be taken.

3.4 Awareness Cues

Because we aim to enhance a feeling of being together, it is important that each user could get a common feeling and achieve a smooth communication. *Shopping Together* supports two awareness cues to aware the user's attention and enhance the communication especially with the context of space information: an avatar and a pointing arrow.

Avatar. The *Avatar* is used to assist users to know the partner's current focusing direction so as to join in some interesting points and grasp the potential conversation topics. We define such interaction as a joint attention moment. As we introduced above, the in-house user could see the in-store partner's profile face directly in the panoramic view. We also try to reveal the in-house user's viewing direction to the in-store user. We create an avatar representing the in-house user. It tracks and follows his/her current head movements (Fig. 8). It presents on the left side of the vision, showing the in-house user's precise facing direction of the shopping venue (see Fig. 3).

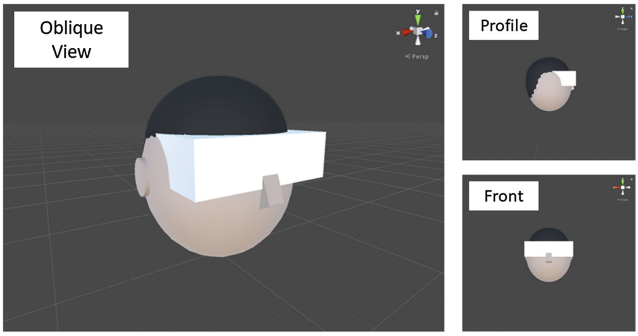


Fig. 8. Avatar representing the in-house user

Directing Arrow. The *Directing Arrow* is a 3D virtual arrow that could be manipulated by the in-house user and used to highlight the pointing direction. It appears at the tip of the user's index finger with a red arrow tip showing the precise directing. When the user changes the position and orientation of the index finger, this virtual arrow changes to match the current pointing direction of the finger. The in-house user uses it to transmit a selecting direction to the partner. For example, it could be used to direct the in-store user to pick up a specific product in the store (Fig. 9).

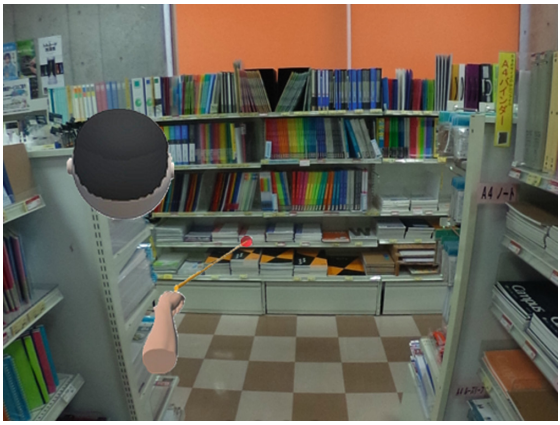


Fig. 9. Visualization sample of the in-store user's view. (Color figure online)

4 Implementation

Figure 10 shows the overview of our framework's setup. It mainly consists of two parts: the in-house user's part and the in-store user's part.

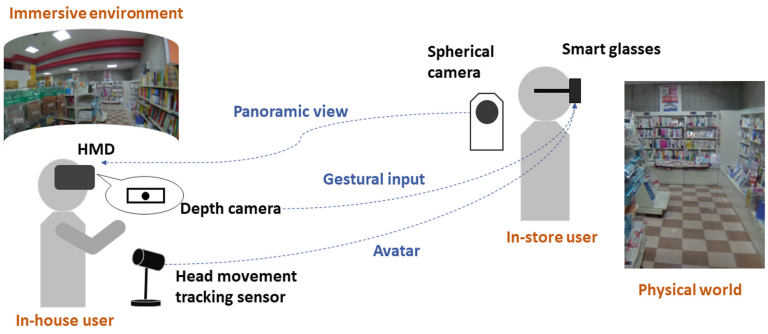
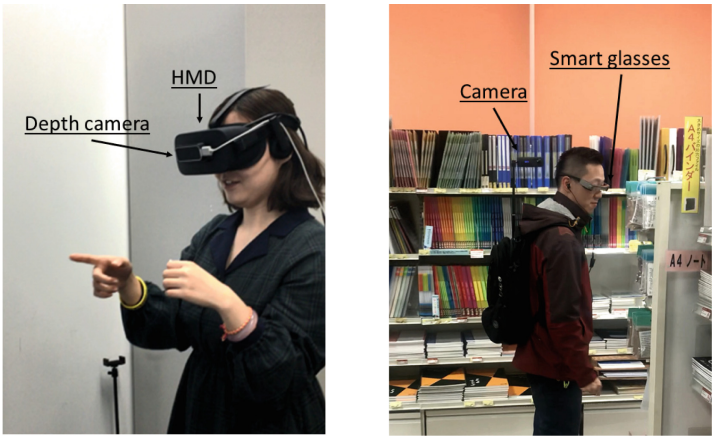


Fig. 10. System overview

In the in-house part, the main physical devices include a head-mounted display (HMD), a depth camera attached on the HMD and a head movement tracking sensor (Fig. 11(a)). In the in-store part, the main physical devices include an augmented-reality smart glasses and a spherical camera (Fig. 11(b)).

4.1 Panorama Stream of Shopping Environment

To provide the immersive shopping environment for the in-house user, we construct a real-time stream of the panorama from a spherical camera. The camera



(a) The in-house user (b) The in-store user

Fig. 11. Physical setups

is set on the top of a fixed metal mount on in-store user's back. It provides a continuous high-quality capture of the environment and is connected to a mobile computer over USB to generate a live stream to in-house user side with Real Time Messaging Protocol (RTMP). The in-house user uses the HMD as the GUI display to view the shopping venue.

4.2 Gestures Capture

The depth data of in-house user's hand including fingers and joints is captured by the depth camera attached to the front side of the HMD, being worn by the in-house user. It is light enough (only about 45 g) to make sure it is comfortable for users to wear, and works with a sub-millimeter tracking accuracy (an about 0.7 mm overall average accuracy with 8 cubic feet interactive range [6]).

4.3 Head Movements

To get the current focusing direction of the in-store user, we extract the head motion data from a 9-axis motion-tracking sensor compacted in the augmented reality smart glasses. A wireless module is used to exchange data via the Internet.

To reveal the focusing direction of the in-house user, a point tracking sensor is used to detect the current facing direction of the user. It measures a full 6 degree of freedom rotational and positional tracking of the head movement which is precise, low-latency, and sub-millimeter accurate. After calculating the angular deviation between the head movement of two users, the relative facing direction is matched to the performance of the avatar.

4.4 Perspective Calibration

To support in-store user getting the side-by-side perspective of the in-house user's hand gestures, we construct a continuous automatic calibration. The system computes the current facing directions of both users to get the angular deviation between both users' viewpoint and adjusts the presenting perspective of the hand gestures in the in-store user's field of vision.

5 System Evaluation

In this section, we introduce our user study and analysis of the results. Participants were asked to accomplish a shopping task. The major purpose of this study is to test whether our system could provide users an effective interaction assist the remote collaboration. We also obtained some feedback from a questionnaire.

5.1 Participants

We recruited eight participants, ranging in age from 20 to 28. All of them have regular computer skills. They were divided into 4 pairs. Each pair had two roles: an in-house user and an in-store user.

5.2 Task and Procedure

In each pair, one participant stayed in the laboratory (the in-house user), the other one went to a store (the in-store user). Participants were allowed to practice using the system for 20 min before starting the experiment.

The study task was going to a stationery store to purchase a little gift which could interest both participants (such as a plastic craft). In each pair, both participants were allowed a fully free viewpoint control. The in-store participant walked around and communicate with the in-house partner, and the latter might request the former to move or do some operations like pick up and show some objects in hands. The in-house user's subsystem was connected to the cabled Internet, while the in-store user's subsystem used a wireless connection. During the experiment, participants were allowed a speech communication via the voice call. The time limitation of each experiment was 30 min.

After each experiment, all four pairs of participants were asked to fill out a questionnaire including 5 questions to get the user feedback. The participants graded each question from 1 to 5 (1 = very negative, 5 = very positive).

5.3 Results and Discussion

In our user study, all four pairs of participants completed the task within the stipulated time.

To investigate the user performance, we record and analysis several duration data of each pair. We define the duration of entire experiment time of one

pair as T_t. It could be classified to two categories: T_c–the duration when participants conducted a user-to-user communication and T_n–the duration when participants only browsed the environment independently without conducting a communication. T_c would also be classified to two types: T_g–the duration when participants performed gestural collaboration and T_o–the duration when participants communicated with other approaches like speech except using gestures. The following formula shows the relation of these data.

$$T_t = T_n + T_c = T_n + (T_g + T_o)$$

We calculate a *Gesture Rate*–R using following formula.

$$R = T_g/T_c = T_g/(T_g + T_o)$$

This rate is measured to show the statistical proportion of gesturing in the user communication. It could reveal whether the user could achieve an effective gestural interaction and how important to support such two-way gestural communication to a certain extent. Table 1 shows the rate of each pairs.

Table 1. Rate of gesturing in user communication

Pair number	1	2	3	4
R	55%	67%	56%	64%

As shown in Table 1, for all pairs, the Gesture Rates are over 50%. In another word, in general conditions, users performed a gestural interaction in more than the half duration of the collaboration, which means that gesture plays an irreplaceable role in such remote communication. It might reflect that our system could truly assist the human-to-human communication by supporting users the two-way gestural interaction.

Table 2 shows the results of our questionnaires. We divided the results into two groups–the in-house users and the in-store users. We calculate the average score of each question in each group.

Table 2. Questionnaire results

	Question	In-house user	In-store user
1	I could easily transmit instructions by gestures	3.5	4.5
2	I could quickly understand the intentions of my partner	4.5	3.25
3	I felt the system operation was easy enough	3.75	4.5
4	I could quickly know my partner's focusing direction	4.5	3.25
5	I felt being with my partner together in same place	3.5	3.75

Question 1 and Question 2 are used to judge the practicability and effectiveness of our two-way gestural interaction design. It indicates that both in-house and in-store users could perform gestures to transmit their intentions and achieve a smooth communication through our system. We also observed that the information transmission from the in-store user to the in-house user was graded a little higher than that from the in-house user. After some post-task interviews with the participant, we found the reason might be that the in-store user performed gestures with physical hands and could touch a object and get some visual feedback such as depressing an object's surface.

Question 4 is used to test the ease and usability of this system. The results suggest that users generally found it is effortless to achieve a communication.

Question 5 indicates that the users could be aware of the partner's attention condition easily which provides the possibility to join in the same scenery and continue to communicate as well as to keep a close connection.

In Question 6, we intend to investigate the overall performance and user experience. It demonstrates that, by using our system, users could get a close relation and receive common perceptions, as well as might feel co-located.

In the post-task interviews, all the participants commented that they would found the designs of *Shopping Together* to be reasonable and useful in the co-shopping. The in-house participants could experience an immersive shopping feeling and felt personally on the scene to some extent. When asked about the feedbacks of the spatial gestural interaction, most in-house participants considered that it was helpful and convenient to perform gestures directly in the shopping venue, especially when making some pointing instructions or showing some specific hand operations. The in-store participants also agreed that presenting the in-house participant's 3D gestures in the physical world was intuitive and understandable enough to reduce the miscommunication.

In general, the average scores of all questions were graded higher than 3 points (3 = medium) by the in-house participants as well as the in-store participants, which meant our user study got a positive overall result. This might signify that our system designs are reasonable and practical. It demonstrates that our system could construct an intuitive communication approach for users to get a certain degree of *Shopping together feeling.*

6 Related Works

One of the related works is a previous research of the remote communication called "WithYou" [7]. WithYou defines three elements of remote communication system which could provide users a feeling of going out together: (1) Enabling both users to freely control the viewing direction onto the outside environment. (2) Users could know the viewing direction of the other one. (3) Gesture communication could support a smooth communication without audio. Withyou focuses on distinguishing the focus status of users and assisting users to gain the opportunities of joint attention moments. This work has inspired our design of sharing attention situation to help users get a common feeling.

Another related work is our previous prototype called "Trip Together", a remote pair sightseeing system supporting gestural communication [8]. It is designed to bridge a gestural communication between a user remaining indoor and a remote user going outside. It investigated providing users an intuitive approach to realize a spatial navigation and direction guidance for a mobile sightseeing. The positive feedback of this work has inspired our research of supporting spatial gestural interaction to assist a remote collaboration so as to enhance users' connection.

7 Conclusion

In this study, we introduce our design of a remote co-shopping system–*Shopping Together* for two geographically separated users, an in-house user and an in-store user. The in-house user remaining in a house gets an immersive shopping experience with the in-store user who goes into a physical shopping world. It simulates a scenario that users conduct the shopping together. We got positive results after carrying out a user study. It demonstrates that being supported by the two-way gestural interaction and attention awareness mechanism, both users could effectively transmit instructions which relate to the physical world and could achieve a smooth remote collaboration. Users could get a close connection when accomplishing a shopping task together and share a certain degree of *Shopping together feeling*.

References

1. Kegel, I., Cesar, P., Jansen, J., Bulterman, D.C., Stevens, T., Kort, J., Färber, N.: Enabling 'togetherness' in high-quality domestic video. In: Proceedings of the 20th ACM International Conference on Multimedia, pp. 159–168. ACM (2012)
2. Cook, S.W., Tanenhaus, M.K.: Embodied communication: speakers gestures affect listeners actions. Cognition **113**(1), 98–104 (2009)
3. Sodhi, R.S., Jones, B.R., Forsyth, D., Bailey, B.P., Maciocci, G.: Bethere: 3D mobile collaboration with spatial input. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 179–188. ACM (2013)
4. Karam, H., Tanaka, J.: Finger click detection using a depth camera. Procedia Manuf. **3**, 5381–5388 (2015)
5. Karam, H., Tanaka, J.: Two-handed interactive menu: an application of asymmetric bimanual gestures and depth based selection techniques. In: Yamamoto, S. (ed.) HCI 2014. LNCS, vol. 8521, pp. 187–198. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07731-4_19
6. Weichert, F., Bachmann, D., Rudak, B., Fisseler, D.: Analysis of the accuracy and robustness of the leap motion controller. Sensors **13**(5), 6380–6393 (2013)
7. Chang, C.T., Takahashi, S., Tanaka, J.: WithYou - a communication system to provide out together feeling. In: Proceedings of the International Working Conference on Advanced Visual Interfaces, pp. 320–323. ACM (2012)
8. Cai, M., Tanaka, J.: Trip together: a remote pair sightseeing system supporting gestural communication. In: Proceedings of the 5th International Conference on Human Agent Interaction, pp. 317–324. ACM (2017)