IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 6, 2017, accepted December 6, 2017, date of publication December 11, 2017, date of current version March 9, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2781699

Geometrically Consistent Projection-Based Tabletop Sharing for Remote Collaboration

DAISUKE IWAI, (Member, IEEE), RYO MATSUKAGE, SOTA AOYAMA, TSUYOSHI KIKUKAWA, AND KOSUKE SATO, (Member, IEEE)
Graduate School of Engineering Science, Osaka University, Toyonaka 560-8531, Japan

Corresponding author: Daisuke Iwai (daisuke.iwai@sys.es.osaka-u.ac.jp)

This work was supported by JSPS KAKENHI under Grant JP16H02859.

ABSTRACT Projection-based tabletop sharing (PBTS) systems allow a local user to share a pointing gesture or a handwritten note on a tabletop document with a remote user who places the same document on a tabletop by projecting the upper limb image of the local user onto the remote document and tabletop. A vertical display is used to share an image of the upper body, including the face of the remote user. However, in previous systems, the spatial layouts of the shared documents must be identical on both tabletops, and the projected upper limb is not extended from the upper body image of the remote user. This paper proposes a PBTS system to address such geometric consistency issues to improve remote collaboration. First, we propose to maintain the geometric consistency of a pointing gesture and handwritten note between each pair of the shared documents rather than between the entire tabletops. This allows users to freely change the document layouts on both tabletops. Second, we propose to overlay the upper limb image such that it is extended from the vertical display, where the upper body image is shown. This is achieved by rotating the upper limb image around the fingertip that performs the pointing gestures or around the tip of the pen used to write notes. We constructed a prototype to determine if the proposed system resolves the geometric consistency issues. Then, we evaluated how accurately a user can convey a pointing position to a distant partner when the document layouts differ between the remote tabletops. Finally, we evaluated how the user experience, particularly the social presence, is improved by the proposed geometrically consistent upper limb direction.

INDEX TERMS Spatial augmented reality, remote collaboration, tabletop sharing, projection mapping.

I. INTRODUCTION

Projection-based tabletop sharing (PBTS) technologies provide seamless collaboration between remote users as if they were colocated around a single shared tabletop [1]–[4]. The tabletop's appearance, including physical documents (e.g., printed papers and books) and the user's upper limb on and above the tabletop, is captured by an overhead camera, transmitted to a remote counterpart and projected onto the remote tabletop. Consequently, users can share various non-verbal contextual information and handwritten notes. We focus on a pointing gesture to indicate a specific area of a document, which is important non-verbal information. These features make PBTS systems useful for many remote collaboration scenarios, such as distance education.

An important advantage of PBTS is that it allows users to manipulate physical documents, which, in many cases, are preferable to digital media, such as tablet computers. Generally, physical documents provide richer haptic feedback, higher flexibility relative to form factors, and higher spatial resolution than digital displays. For desk work, an important property of physical documents is that the spatial layout can be adjusted freely, which significantly facilitates effective work. However, to maintain geometrically correct mutual overlay, previously proposed PBTS systems assumed that the spatial layouts on both distant tabletops were identical. Such geometric consistency must be maintained during a collaboration session; thus, users were restricted to a fixed layout (Fig. 1(a)).

In some PBTS systems, a vertical display is used to share the real-time image of a remote user's upper body, including their face. This video conferencing function enables face-to-face distant communication and provides users with more immersive collaborative experiences. However, in previous systems, the direction of the projected upper limb of a remote user was not geometrically correct, i.e., the upper limb was not extended from the upper body image on the vertical display (Fig. 1(a)). This occurs because (1) the spatial layouts of the documents are fixed and identical on both tabletops, (2) both users must be positioned on the same side of their tabletops to allow the text to be read easily, i.e., not upside down, and (3) the vertical display is generally placed on the opposite side of the tabletop from each user. Although previous studies did not consider geometric inconsistency as a serious problem, it restricts the number of remote users. In other words, when more than two remote users join a collaborative session, the remote upper limbs are projected onto the tabletop such that they extend from the same side of the tabletop. In this situation, it is particularly impossible to recognize the owner of each projected upper limb. Geometric inconsistency can also significantly degrade the sense of presence of a remote user, particularly social presence, which is a measure of the perceived presence of another intelligent being. Social presence has been considered one of the most important factors influencing the quality of distant communication [5], [6].

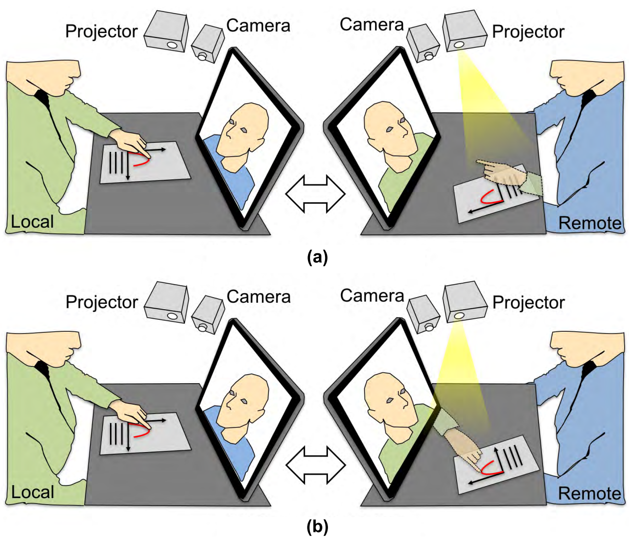

FIGURE 1. PBTS system: (a) pointing position of a local user is not correctly transferred to the remote tabletop when the layout of a shared document differs between the local and remote tabletops, and the projected upper limb image is not extended from a vertical display showing the upper body image in a conventional system, (b) the proposed system addresses these geometrical issues.

This paper proposes a PBTS system to solve the two geometric consistency issues mentioned above in order to provide better remote collaboration (Fig. 1(b)). We provide two main technical contributions. First, we propose to maintain the geometric consistency of a pointing gesture and handwritten note between each pair of shared documents rather than between the entire tabletops. From a captured top-view image of each tabletop, we extract an upper limb region and handwritten note. In addition, we recognize all physical documents on the tabletop. We apply a simple image feature matching computer vision technique to find document correspondences between shared tabletops. We then estimate the 2D (two dimensional) geometric transformation (i.e., homography) parameters for each document. Then, we translate the extracted upper limb image of a local user such that the fingertip performing a pointing gesture points at the same position in the shared document on the remote tabletop. We also transform the handwritten note geometrically using the homography transformation. We project these images onto the remote tableto, which allows users to freely change the layout of the documents on both tabletops. Second, we overlay the upper limb image such that it is extended from the vertical display where the upper body image is shown. This is achieved by rotating the upper limb image around the fingertip performing the gestures or the tip of a pen.

The reminder of this paper is organized as follows. Related work is discussed in Section II. We introduce the principle of the proposed method in Section III. In Section IV, we describe a prototype used to provide proof of concept demonstrations and confirm whether the geometric consistency issues are resolved. Then, in Section V, we evaluate how accurately a user can convey a pointing position to the distant partner when the document layouts differ between the remote tabletops. We evaluate how the user experience, particularly social presence, is influenced by the direction of a projected upper limb image relative to a vertical display showing an upper body image in Section VI. Evaluation results are discussed in Section VII, and conclusions and suggestions for future work are given in Section VIII.

II. RELATED WORK

Interactive surfaces, such as tabletops and walls, have been studied extensively to realize distant document sharing in various fields, such as human computer interaction and augmented reality (AR). Previously proposed systems roughly fall into two categories based on the properties of the shared documents.

In the first category, digital documents are shared in interactive surface systems. For example, digital users can share digital documents projected onto large touch sensitive tabletop displays. In such system, users can draw lines and insert text using stylus [7]. Another type of system employs virtual transparent boards. Here, one user draws visual information from one side of the board and a remote user does so from the other side, and the information is overlaid [8]–[12]. Because these systems emulate transparent boards, they also share the appearance of the users' bodies in front of the surfaces, which allows rich non-verbal communication through body gestures.

In the second category, interactive surfaces are used to share physical documents on tabletops. Basically, the appearance of a local tabletop on which physical documents are placed is captured by an overhead camera and projected onto a remote tabletop [1], [4] or overlaid digitally on standard computer monitors [13]. To increase the capability of non-verbal communication, such as gestures and facial

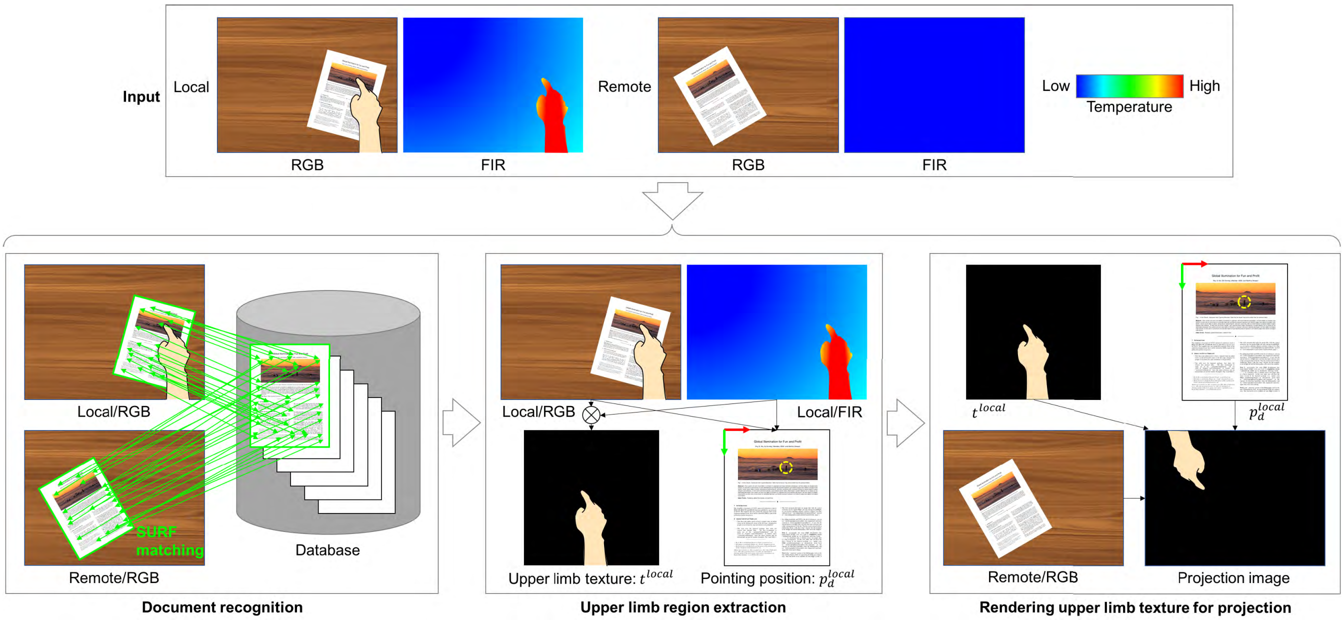

FIGURE 2. Server process flow for sharing upper limb image from a local tabletop to a remote counterpart.

expressions, some systems are equipped with an additional vertical display on which the upper body image of a distant user is shown [2], [3], [14]. Recent advancements in depth sensing technologies have enabled sharing tabletop objects in 3D by capturing the objects using a depth camera and a stereo projection system to display the objects on a distant tabletop [15]. However, as mentioned in Section I, geometric consistency issues have not been considered adequately in previous studies, i.e., the spatial layouts of the physical documents must be fixed on both tabletops. In the systems where a vertical display shows a distant user's upper body, the direction of the projected upper limb image is not extended from the vertical display. In this paper, we focus on distant surface sharing and develop a system to address these consistency issues.

Social presence is a key measure to assess the quality of distant communication [5], [6], [16]. A greater degree of social presence leads to more efficient collaboration. For example, in distance education scenarios, providing high degrees of social presence of a teacher is desirable to enhance the learning environment [17]–[19]. To increase social presence, a system must be able to transmit a variety of signals at high fidelity [20]. We propose overlaying a user's upper limb image on a distant tabletop such that its direction geometrically corresponds to the upper body image shown on a vertical display. We performed a user study to determine if this visualization technique improves the social presence of distant users.

III. GEOMETRICALLY CONSISTENT TABLETOP SHARING

The proposed method employs a client-server system. A tabletop's appearance is captured by an overhead camera and transmitted from each client (i.e., tabletop) to the server. The server computes projection images and transmits them to the clients. This section describes the technical details of the proposed geometrically consistent tabletop sharing method. For simplicity, we assume only one remote user; however, the proposed technique can be applied to multiple remote users. Figure 2 shows the server process flow for sharing an upper limb image from a local tabletop to the remote counterpart.

A. RECOGNIZING AND LOCALIZING DOCUMENTS

We recognize physical documents on tabletops and estimate their poses using an image feature matching technique. We assume that the documents are digitized and stored in an online database in advance [21] or are the printed versions of these digital documents.

Considering the trade-off between the speed and robustness of the document recognition performance, we apply SURF (speeded-up robust features) [22] to detect feature points and compute their feature descriptors. SURF feature points and descriptors are computed offline for all documents in the database. SURF features in a tabletop image captured by the overhead camera are detected online. Then, we recognize documents in the image by applying a feature matching technique based on the Euclidean distance between the feature vectors of the detected points in the captured image and the database documents. After finding matched documents, we estimate homography transformation parameters as the poses of these documents. Note that we apply RANSAC (random sampling consensus) to ensure robust estimation. We perform this process in parallel for both the local and remote tabletops.

B. EXTRACTING UPPER LIMB REGION

Robust upper limb region extraction is crucial for the proposed geometrically consistent tabletop sharing system. We assume that tabletop appearances change dynamically due to projected imagery, environmental light, and document layout modification; thus, simple thresholding for RGB color information does not work robustly. For example, a previous study proposed calibrating the color spaces of a camera and projector to realize robust background appearance estimation for upper limb extraction [11]. However, this method only works correctly when there is no physical document on a tabletop and the environmental light does not change. Therefore, this method is insufficient for our target situation. To address this problem, a previous study used a near infrared (NIR)-based depth camera (i.e., an RGB-D camera) to extract a user's upper limbs, and this method was robust against dynamic changes to the tabletop appearance [15]. However, we assume that a user's upper limbs are normally very close to the tabletop surface or even touch it. In this situation, the depth difference between an upper limb and the tabletop is not distinguishable in a depth image captured by existing RGB-D cameras. Although other techniques using 2D NIR cameras can robustly detect touch gestures on a tabletop by analyzing the small image patch around a user's fingertip [23], [24], such techniques were not designed to extract the entier upper limb region. Following a pioneering work [25], there have been a lot of investigations on hand extraction based on color stereo matching. Researchers already overcame the problem of occlusion and high similarity between fingers in the stereo matching of hands [26], [27]. Such hand segmentation technology was already used in handling upper limb occlusion in mixed reality applications [28]. In contrast to the stereo matching approach, we propose a simpler solution to robustly extract upper limb region.

We propose using a far infrared (FIR) camera to detect heat information. Typically, a user's upper limb is warmer than other objects on a tabletop; thus, they can be extracted by applying a simple thresholding technique to FIR images [29]. In addition, we use an RGB camera to capture the appearance of the extracted upper limb region. Geometric registration between the FIR and RGB cameras is achieved by 2D homography transformation. Here the homography parameters are calibrated on the tabletop plane. Using this process, we can acquire the RGB texture of the extracted upper limb on the tabletop. We denote the upper limb texture in a tabletop i as $\boldsymbol{t}^i$.

From observations of co-located people communicating about a document positioned in front of them, we found that the most dominant hand gesture was a pointing gesture to indicate a specific position on the document. They used the fingertip that was the furthest from the wrist, which we refer to as the pointing fingertip, to perform the pointing gesture. Therefore, after extracting the upper limb region in the FIR image, we compute the position the pointing fingertip is pointing to in the document coordinate system of the document. The document coordinate system of each document is a 2D Euclidean coordinate system whose origin is at the top-left corner. The pointing fingertip position is detected as the furthest point of the extracted upper limb region from the side of the tabletop where the user is located. This method is simple, but works well for our target situations. The computed pointing fingertip position in the coordinate system of document d is denoted $\boldsymbol{p}_d^i$.

C. RENDERING UPPER LIMB IMAGE

Here, we explain how we render an upper limb image to be projected onto a remote tabletop from an upper limb texture extracted from a local tabletop. Note that the same process is applied when rendering a projection image for the local tabletop. First, the extracted upper limb region of local user $\boldsymbol{t}^{local}$ is translated such that the pointing fingertip is located at the same position in the corresponding document on the remote tabletop (i.e., $\boldsymbol{p}_d^{local} = \boldsymbol{p}_d^{remote}$). Second, the texture is rotated around the fingertip such that it extends from a vertical display that shows the upper body image of the local user. The rotational angle is the angle between the bottom side of the vertical display and the side of the tabletop where the remote user locates. For example, when the user faces the vertical display, the rotational angle is 180 degrees.

Note that there is a technical limitation in this method. Particularly, the arm of an upper limb image may possibly be too short to be connected to the vertical display on which the upper body image is shown. Such disconnection might weaken the impression that the projected upper limb belongs to the distant partner. In addition, with multiple distant users, the disconnected upper limbs make it difficult to recognize the owner of each limb. Therefore, we employ another rendering method for the upper limb images. We refer to the upper limb representation explained thus far as *original* and the new representation as *extended*. We render the *extended* upper limb image by connecting a quadrilateral to the *original* upper limb silhouette to extend the arm such that it connects to the bottom of the vertical display. Note that it is difficult to generate a natural texture for the quadrilateral area; therefore, we uniformly paint the *extended* upper limb image using a single color computed as the average color of the *original* upper limb image. Although the *extended* upper limb image sometimes has an unnaturally long arm, a previous study confirmed that such extended arm representation is acceptable and useful in many face-to-face collaboration scenarios [30]. Here, each representation has different advantages and disadvantages as described above; therefore, we allow users to select a representation that best fits their application.

D. SHARING HAND DRAWN LINES

A digital pen can be used to draw a line on a document on the local tabletop. The drawn line is then projected onto the corresponding document on the remote tabletop. To render a line to be projected onto the remote tabletop, a line drawn by a digital pen whose trajectory on the local tabletop is measured is transformed geometrically by homography transformation using the parameters described in Section III-A.

In the pen drawing scenario, the same process is applied to render an upper limb image as described in Sections III-B and III-C. Note that there is only one exception, i.e., the upper limb image is rotated around the pen tip position rather than the pointing fingertip.

E. SEE-THROUGH DOCUMENTS

The proposed system allows users to freely change the spatial layout of tabletop documents; therefore, it is possible that shared documents may overlap partially. Our SURF-based document recognition and localization method works well in such a case. However, a problem may occur when a user points to an area on a shared document that is covered by another document on the distant tabletop. To address this issue, we apply a projection-based diminished reality technique [31]. Diminished reality is an AR framework that makes a real object see-through so that a user can see its background. The latest comprehensive survey summarizes different techniques in video see-through and projection-based AR environments [32]. Here, we project the texture of the hidden document onto the occluding document on the distant tabletop.

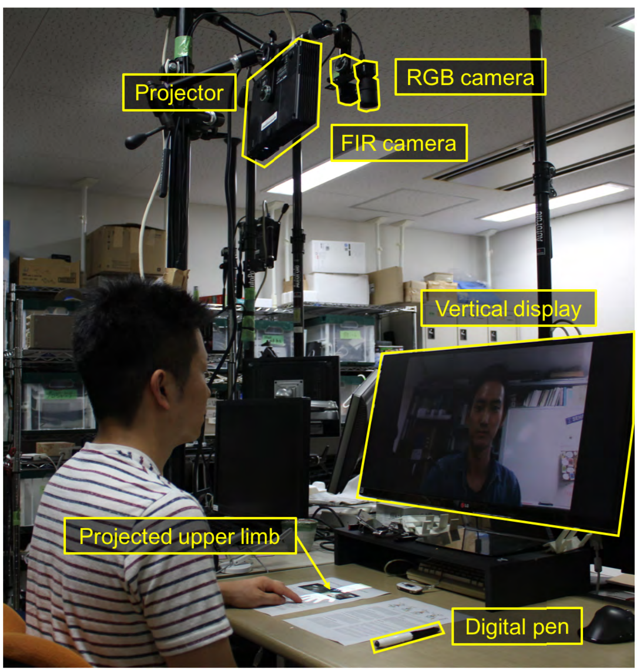

FIGURE 3. Prototype system.

IV. PROTOTYPE SYSTEM

Figure 3 shows our working prototype system, which consists of overhead RGB (PointGrey Flea3 FL3-U3-32S2C, 1920×1080 pixels) and FIR (Optris PI-450, 382×288 pixels) cameras, a projector (NEC NP-L50WJD, 1280×1080 pixels, 500 ANSI Lumen), and a digital pen (Pentel airpenPocket++). We prepared three of these systems, which were controlled using a single PC (CPU: Core i7-4770K 3.50 GHz, RAM: 16 GB) on which remote collaborations were emulated. We used several flat panel displays of different sizes as vertical displays to show the upper body images of the distant users. To reduce computational load on the main PC, upper body image communication was performed independently. We communicated upper body images using Skype by connecting low-power PCs with the displays or by using laptop PCs.

We conducted proof of concept demonstrations to show how the prototype works when remote users share pointing gestures and handwritten notes for the same documents placed on remote tabletops. Figure 4 shows scenes captured from these demonstrations. The *original* upper limb representation was applied in the demonstrations shown in Figures 4(a) and (c), while the *extended* representation was applied in the demonstrations shown in Figures 4(b) and (d).

In Figure 4(a), two participants sitting at remote tabletops are collaborating using the prototype. A printed document and a book are placed on the local tabletop, and the same documents are placed on the remote tabletop in a different spatial layout. Both participants point at the same position in the book using their fingers. Each projected upper limb image also points at the same position on both the local and remote tabletops. From this result, we confirm that pointing locations can be transferred between distant tabletops even when the document layouts differ. Furthermore, we confirm that the arms of the projected upper limbs are extended from the vertical displays that show the upper body images of the distant users.

Figure 4(b) shows a distant collaboration using a digital pen. The local participant drew lines on a paper document. The drawn lines were projected on the remote document as red lines. The hand with the pen was also projected such that it was extended from the vertical display and the pen tip corresponded to the most recently drawn line. From this result, we confirm that our geometrically consistent tabletop sharing system works for pen drawing scenarios.

Figure 4(c) shows a situation where the local participant pointed at a place in the book while the corresponding place was occluded by a paper document on the remote tabletop. In this situation, the texture of the occluded (i.e., pointed at) page is projected to provide the see-through effect explained in Section III-E (Fig. 4(c-1)). Then, the remote participant removed the overlaid document to see the pointed part of the occluded page (Fig. 4(c-2)). From this result, we confirm that the see-through projection works effectively for occluded documents.

Figure 4(d) shows s collaboration with three remote people using pointing gestures. We confirm that the proposed method can support collaborations with more than two remote users. The projected upper limbs of the remote participants were extended from the vertical displays. Obviously, this makes it easy for each local user to recognize the owner of each projected hand.

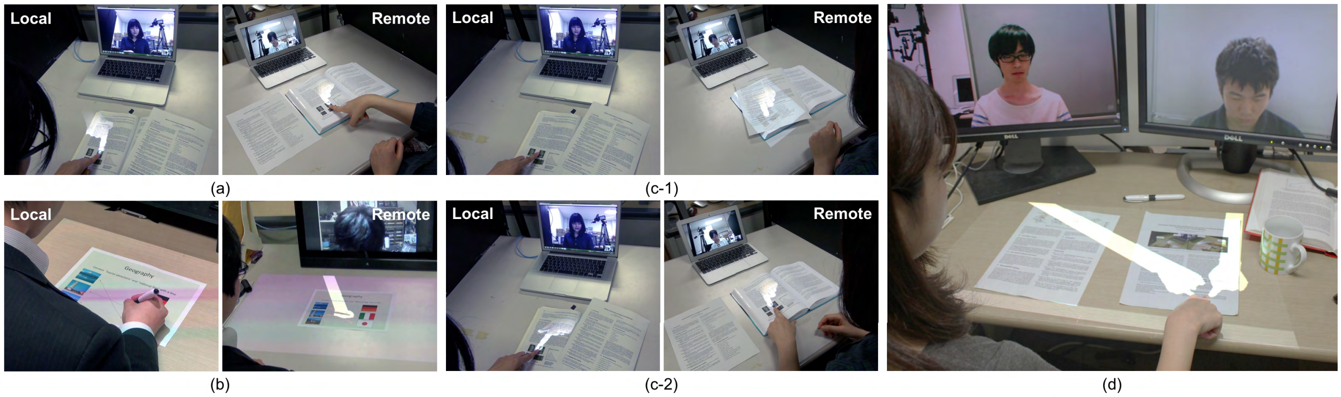

FIGURE 4. Proof of concept demonstrations: (a) pointing gesture shared by two remote users; (b) sharing digital pen drawing; (c-1) see-through effect revealing lower document information; (c-2) sharing pointing gesture after removing the upper document; and (d) pointing gesture shared among three remote users. The *original* upper limb representation was used in (a) and (c), and the *extended* representation was used in the other demonstrations.

V. EVALUATION OF POINTING GESTURE ACCURACY

We investigated how accurately a remote pointing gesture conveys a pointing position. Because the proposed system applies several vision-based geometric transformations, projected pointing positions may shift from correct positions due to calibration and measurement errors. The transformations, particularly rotation around the fingertip, may also confuse a user's perception and make it difficult to recognize which part of the projected upper limb image corresponds to the pointing fingertip. Through a user study, we evaluated the effect of the geometric transformations on pointing position communication. The effect is generally more significant in the *extended* upper limb representation because its appearance is a silhouette without the upper limb texture. Therefore, the user study was conducted using the *extended* representation to investigate the lower bound of the pointing position communication capability.

In each task, a participant (*pointer*) pointed at a position on a document using a finger, and a remote participant (*marker*) observed the projected finger and marked the pointed position on the corresponding remote document using a pen. Then, we measured the distance between the pointed and marked positions, which we considered the pointing gesture error. An experimental session consisted of multiple tasks with different pointing positions, and the average error was evaluated.

We prepared two types of documents for this experiment, i.e., one text document and one photographic print (both A4 size). Five pointing positions were determined randomly for each document before each session began such that they were spread across the given document. In each session, the document was placed randomly in the projection area, which was 630×360 mm^2 on a tabletop. Ten participants were recruited from a local university (eight males, two females; 22–25 years). The participants were divided into five groups of two. For each document, each group performed two experimental sessions in which a participant acted as either the *pointer* or *marker* in the first session. The roles were reversed in the second session. Thus, each group performed four experimental sessions (=two documents × two roles).

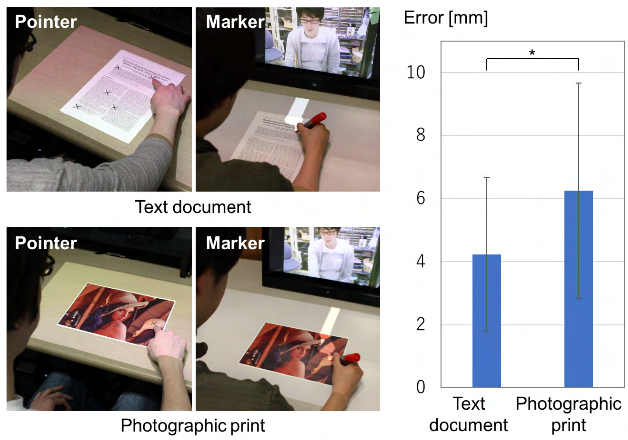

FIGURE 5. Pointing gesture accuracy: (left) captured experimental scenes of *pointer* and *marker* participants; (right) average errors with standard deviations of errors between pointed and marked positions. The printed crosses indicate where the *pointer* participant is required to point (*: $p < 0.05$).

Figure 5 shows the experimental scenes and the average errors with standard deviations for each of the documents. Here, the participants could convey pointing positions at an accuracy of 4.2 mm with the text document and 6.3 mm with the photographic print on average. These accuracies are less than 2% of the length of the diagonal of the A4 document.

Note that the error was greater with the photographic print. A two-tailed paired t-test showed a statistically significant difference between the results ($p < 0.05$). The main reason for this difference is that the visibility of the projected hand was worse with the photographic print due to the lower reflectance and spatially more complex pattern of the surface texture. A solution to improve visibility would be to apply dynamic radiometric compensation to visually cancel the surface textures [33], [34].

VI. EVALUATION OF SOCIAL PRESENCE

We conducted a user study to investigate if the collaboration experience is improved by our geometrically consistent representation of the projected upper limb direction.

A. EXPERIMENTAL CONDITIONS

We compared the perceived social presence of a remote user between the following naïve and proposed conditions. In the naïve condition, which has been applied to previous systems [2], the direction of a projected upper limb is nearly identical to that of a local user's physical upper limb. With the proposed condition, the direction of the projected upper limb is consistent with the upper body image shown on a vertical display facing the local user. Therefore, the difference in direction between the projected and physical upper limbs is approximately 180 degrees. The goal of this experiment was to evaluate the effect of the direction of the upper limb image; thus, we applied the *original* upper limb representation in the proposed condition to ensure that the other factors were the same between conditions.

When the size of the vertical display is too small to display a life-size upper body image, the rotated upper limb may appear somewhat unnatural and thus fall into the uncanny valley. Such small displays, including tablet computers (e.g., the iPad and Microsoft Surface) and laptop PCs, are often used in TV conferencing and PBTS systems [2]; therefore, we decided to conduct our investigation in consideration of such small displays. Here, a large display (31 inches) and a small display (14 inches) were used in the experiment. As a result, we prepared four experimental conditions (=two directions×two display sizes).

B. PROCEDURE

Each participant acted as a local user, and an experimenter acted as the remote user. Social presence is strongly affected by the intimacy between distant users; thus, each participant had not met the experimenter before, and performed the following experimental session under only one of the experimental conditions. Each session consisted of two tasks. First, the participant was asked to solve a Japanese crossword puzzle with the remote collaborator. Figure 6 shows scenes of the crossword puzzle task under the four experimental conditions. Then, the participant was asked to verbally explain how they came to the room where the experiment was conducted from the nearest train station.

After finishing the second task, the participant answered a questionnaire developed by Hams and Biocca [16] to assess the social presence of communication media. The questionnaire evaluates six social presence dimensions, i.e., co-presence (CoP), attentional allocation (Att), perceived message understanding (Msg), perceived affective interdependence (Aff), perceived emotion interdependence (Emo), and perceived behavioral interdependence (Bhv). Here, each dimension was evaluated by six questions, each of which was based on a nine-point Likert scale (1: disagree, . . . , 9: agree).

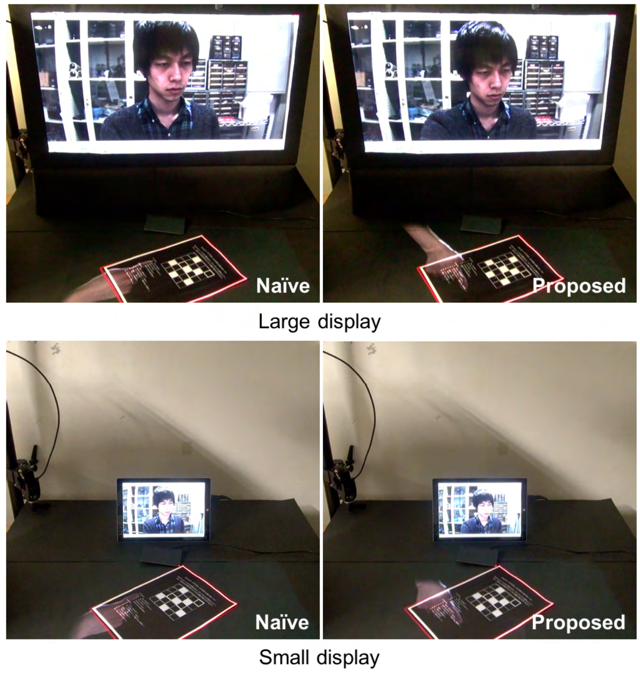

FIGURE 6. **Crossword puzzle and displayed upper limb and upper body images under four conditions.**

Thus, each participant scored 36 different questions in total. The score of each dimension was computed by averaging the scores of the six corresponding questions.

C. RESULT

Forty-four participants (33 males and 11 females; 19-25 years; native Japanese speakers) were recruited from a local university. They were divided into four groups (corresponding to the experimental conditions). Thus, 11 participants in the same group performed the experimental session under the same experimental condition.

Figure 7 shows the average scores with the standard deviations. To investigate the effect of the direction of a projected upper limb, we analyzed the data by performing a two-tailed unpaired t-test to compare the scores of the naïve and proposed conditions for each display size. As a result, there were statistically significant differences relative to the CoP ($p < 0.01$), Att ($p < 0.05$), Aff ($p < 0.01$), and Bhv ($p < 0.01$) dimensions between the naïve and proposed conditions with the large display. There were also statistically significant differences relative to the Att ($p < 0.05$), Aff ($p < 0.01$), Emo ($p < 0.05$), and Bhv ($p < 0.01$) dimensions between the two conditions with the small display. From this analysis, social presence relative to the Att, Aff, and Bhv dimensions were improved by the proposed hand direction regardless of display size. We also confirmed there was no social presence dimension where the score was significantly reduced by the proposed condition compared to the naïve condition.

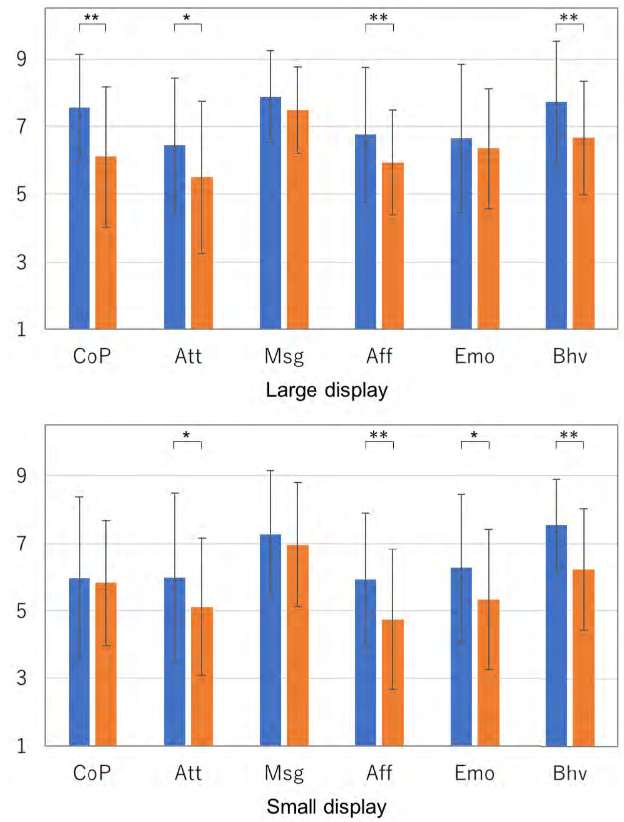

FIGURE 7. Average scores with standard deviations of six dimensions to evaluate social presence (: $p < 0.01$, *: $p < 0.05$).**

VII. DISCUSSION

From the proof of concept demonstrations described in Section IV, we confirmed that a geometrically consistent PBTS system is possible with the proposed method. When a user points to a position in a local document, the projected upper limb image also points to the same position in the remote counterpart. In addition, the upper limb is rotated appropriately such that it is extended from the vertical display showing the upper body image of the local user. We also confirmed that the system supports sharing handwritten drawings using a digital pen and enables seeing through occluded documents. In all cases, the poses of shared documents differed among the distant tabletops. Therefore, we have confirmed that the proposed system allows users to freely position shared documents.

From the accuracy evaluation results (Section V), we confirmed that the error relative to pointing position communication is less than 2% of the diagonal length of the documents on average. We believe that this level of error is acceptable for many distant tabletop sharing applications. We also consider that the pointing fingertip is recognizable even when the *extended* upper limb representation, which is simply a silhouette, is applied.

From the user study (Section VI), we confirmed that the geometrical consistency of a projected upper limb direction has significantly positive effect on the social presence of a remote user even when the vertical display size is too small to display a life-size upper body image. Therefore, we have confirmed that the proposed method improves social presence regardless of potential size inconsistency of the projected upper limb and displayed upper body.

A. LIMITATIONS

A projected upper limb moves along an unnatural trajectory when it crosses over two documents that are positioned differently, as shown in Fig. 8. This occurs because the upper limb image position is determined independently based on each document coordinate system, and the relationship of the locations between two documents is not considered. In future, we plan to address this issue by designing the trajectory and speed of the projected upper limb when the pointing fingertip is outside the shared document areas such that users perceive natural movement.

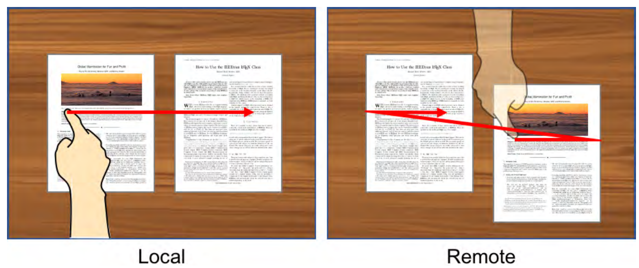

FIGURE 8. Trajectory (red line) of a user's pointing fingertip on a local tabletop and that of a projected fingertip on a remote tabletop.

In addition, a user cannot share two upper limbs simultaneously in the current system. We assume that people generally use only one upper limb for communication when using a pointing gesture or drawing a line; thus, we believe that this limitation will not degrade the quality of the distant communication provided by the proposed system. If there is sufficient demand for sharing two upper limbs, we will extend the proposed system to support this functionality, which will require the system to recognize two upper limb regions independently.

VIII. CONCLUSION

In this paper, we have proposed a geometrically consistent PBTS system, wherein an upper limb image of a remote user can point at a geometrically correct position in a shared document on a local tabletop, and this upper limb image is extended from a vertical display showing the upper body image of the remote user. Through proof of concept demonstrations using prototype systems, we confirmed that these concepts can be achieved even when the layouts of the shared documents differ between the remote tabletops. Using a prototype system, an experiment demonstrated that the pointing position can be conveyed accurately with an error of less than 2% of the diagonal length of an A4 document. We also evaluated how user experience is improved by our geometrically consistent representation of the projected upper limb direction. In addition, the results of a user study showed that multiple social presence dimensions are improved by the proposed method even when a small display is used as the vertical display. Interesting future work could include an investigation into how user behaviors changed with the proposed PBTS system compared to conventional systems.

REFERENCES

[1] J. C. Tang and S. L. Minneman, ''Videodraw: A video interface for collaborative drawing,'' in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, New York, NY, USA, 1990, pp. 313–320.

[2] S. Junuzovic, K. Inkpen, T. Blank, and A. Gupta, ''Illumishare: Sharing any surface,'' in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, New York, NY, USA, 2012, pp. 1919–1928.

[3] S. Yarosh, S. Cuzzort, H. Müler, and G. D. Abowd, ''Developing a media space for remote synchronous parent-child interaction,'' in *Proc. 8th Int. Conf. Interaction Design Children*, New York, NY, USA, 2009, pp. 97–105.

[4] C. J. Cornelius, M. A. Nguyen, C. C. Hayes, and R. Makena, ''Supporting virtual collaboration in spatial design tasks: Are surrogate or natural gestures more effective?'' *IEEE Trans. Human-Mach. Syst.*, vol. 43, no. 1, pp. 92–101, Jan. 2013.

[5] J. Short, E. Williams, and B. Christie, *The Social Psychology of Telecommunications*. Hoboken, NJ, USA: Wiley, 1976.

[6] F. Biocca, C. Harms, and J. K. Burgoon, ''Toward a more robust theory and measure of social presence: Review and suggested criteria,'' *Presence, Teleoper. Virtual Environ.*, vol. 12, no. 5, pp. 456–480, Oct. 2003.

[7] P. Tuddenham and P. Robinson, ''T3: Rapid prototyping of high-resolution and mixed-presence tabletop applications,'' in *Proc. 2nd Annu. IEEE Int. Workshop Horizontal Interact. Human-Comput. Syst. (TABLETOP)*, Oct. 2007, pp. 11–18.

[8] J. Zillner, C. Rhemann, S. Izadi, and M. Haller, ''3D-board: A whole-body remote collaborative whiteboard,'' in *Proc. 27th Annu. ACM Symp. User Interface Softw. Technol.*, New York, NY, USA, 2014, pp. 471–479.

[9] H. Ishii and M. Kobayashi, ''Clearboard: A seamless medium for shared drawing and conversation with eye contact,'' in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, New York, NY, USA, 1992, pp. 525–532.

[10] K. Higuchi, Y. Chen, P. A. Chou, Z. Zhang, and Z. Liu, ''ImmerseBoard: Immersive telepresence experience using a digital whiteboard,'' in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, New York, NY, USA, 2015, pp. 2383–2392.

[11] F. Coldefy and S. L.-dit-Picard, ''DigiTable: an interactive multiuser table for collocated and remote collaboration enabling remote gesture visualization,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[12] A. Tang, C. Neustaedter, and S. Greenberg, *VideoArms: Embodiments for Mixed Presence Groupware*. London, U.K.: Springer, 2007, pp. 85–102.

[13] H. Ishii, ''TeamWorkStation: Towards a seamless shared workspace,'' in *Proc. ACM Conf. Comput.-Support. Cooperat. Work*, New York, NY, USA, 1990, pp. 13–26.

[14] S. Izadi, A. Agarwal, A. Criminisi, J. Winn, A. Blake, and A. Fitzgibbon, ''C-slate: A multi-touch and object recognition system for remote collaboration using horizontal surfaces,'' in *Proc. 2nd Annu. IEEE Int. Workshop Horizontal Interact. Human-Comput. Syst. (TABLETOP*, Oct. 2007, pp. 3–10.

[15] H. Benko, R. Jota, and A. Wilson, ''Miragetable: Freehand interaction on a projected augmented reality tabletop,'' in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, New York, NY, USA, 2012, pp. 199–208.

[16] P. C. Harms and P. F. Biocca, ''Internal consistency and reliability of the networked minds measure of social presence,'' in *Proc. 7th Annu. Int. Workshop, Presence*, 2004, pp. 246–251.

[17] B. Newberry, ''Raising student social presence in online classes,'' in *Proc.World Conf. WWW Internet*, 2001, pp. 905–910

[18] S. R. Aragon, ''Creating social presence in online environments,'' *New Directions Adult Continuing Edu.*, vol. 2003, no. 100, pp. 57–68, 2003. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/ace.119/abstract

[19] D. Garrison, T. Anderson, and W. Archer, ''Critical inquiry in a text-based environment: Computer conferencing in higher education,'' *Internet Higher Edu.*, vol. 2, nos. 2–3, pp. 87–105, 1999.

[20] K. Tanaka, H. Nakanishi, and H. Ishiguro, *Comparing Video, Avatar, and Robot Mediated Communication: Pros and Cons of Embodiment*. Berlin, Germany: Springer, 2014, pp. 96–110.

[21] Y. Watanabe, M. Tamei, M. Yamada, and M. Ishikawa, ''Automatic page turner machine for high-speed book digitization,'' in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 272–279.

[22] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool, ''Speeded-up robust features (SURF),'' *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[23] A. Wilson, ''PlayAnywhere: A compact interactive tabletop projection-vision system,'' in *Proc. 18th Annu. ACM Symp. User Interface Softw. Technol.*, New York, NY, USA, 2005, pp. 83–92.

[24] H. Xu, D. Iwai, S. Hiura, and K. Sato, ''User interface by virtual shadow projection,'' in *Proc. SICE-ICASE Int. Joint Conf.*, Oct. 2006, pp. 4814–4817.

[25] A. Utsumi and J. Ohya, ''Multiple-hand-gesture tracking using multiple cameras,'' in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Jun. 1999, p. 478.

[26] J. Sun, Y. Li, S. B. Kang, and H.-Y. Shum, ''Symmetric stereo matching for occlusion handling,'' in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2. Jun. 2005, pp. 399–406.

[27] Y. Ren, X. Xie, G. Li, and Z. Wang, ''A scan-line forest growing-based hand segmentation framework with multipriority vertex stereo matching for wearable devices,'' *IEEE Trans. Cybern.*, to be published. [Online]. Available: http://ieeexplore.ieee.org/document/7819455/

[28] F. Narducci, S. Ricciardi, and R. Vertucci, ''Enabling consistent hand-based interaction in mixed reality by occlusions handling,'' *Multimedia Tools Appl.*, vol. 75, no. 16, pp. 9549–9562, Aug. 2016.

[29] M. Kobayashi and H. Koike, ''EnhancedDesk: Integrating paper documents and digital documents,'' in *Proc. 3rd Asia–Pacific Comput. Human Interaction*, Jul. 1998, pp. 57–62.

[30] Y. Ueda, Y. Asai, R. Enomoto, K. Wang, D. Iwai, and K. Sato, ''Body cyberization by spatial augmented reality for reaching unreachable world,'' in *Proc. 8th Augmented Human Int. Conf.*, New York, NY, USA, 2017, pp. 19:1–19:9. [Online]. Available: http://doi.acm.org/10.1145/3041164.3041188

[31] D. Iwai and K. Sato, ''Document search support by making physical documents transparent in projection-based mixed reality,'' *Virtual Reality*, vol. 15, no. 2, pp. 147–160, 2011.

[32] S. Mori, S. Ikeda, and H. Saito, ''A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects,'' *IPSJ Trans. Comput. Vis. Appl.*, vol. 9, no. 1, p. 17, Jun. 2017.

[33] H. Asayama, D. Iwai, and K. Sato, ''Diminishable visual markers on fabricated projection object for dynamic spatial augmented reality,'' in *Proc. SIGGRAPH Asia Emerg. Technol.*, New York, NY, USA, 2015, pp. 7:1–7:2. [Online]. Available: http://doi.acm.org/10.1145/2818466.2818477

[34] H. Asayama, D. Iwai, and K. Sato, ''Fabricating diminishable visual markers for geometric registration in projection mapping,'' *IEEE Trans. Vis. Comput. Graphics*, to be published. [Online]. Available: http://ieeexplore.ieee.org/document/7831400/

DAISUKE IWAI (M'16) received the B.S., M.S., and Ph.D. degrees from Osaka University, Japan, in 2003, 2005, and 2007, respectively. He was a Visiting Scientist at Bauhaus-University Weimar, Germany, from 2007 to 2008, and a Visiting Associate Professor at ETH, Switzerland, in 2011. He is currently an Associate Professor with the Graduate School of Engineering Science, Osaka University. His research interests include spatial augmented reality and projector-camera systems.

RYO MATSUKAGE received the B.S. and M.S. degrees from Osaka University, Japan, in 2014 and 2016, respectively. He is currently with the Keyence Corporation. His research interests include spatial augmented reality and distant communication.

SOTA AOYAMA received the B.S. degree from Osaka University, Japan, in 2016. He is currently pursuing the master's degree with the Graduate School of Engineering Science, Osaka University, Japan. His research interests include spatial augmented reality and pseudo-haptics.

TSUYOSHI KIKUKAWA received the B.S. degree from Osaka University, Japan, in 2016. He is currently pursuing the master's degree with the Graduate School of Engineering Science, Osaka University, Japan. His research interests include spatial augmented reality and distant communication.

KOSUKE SATO (M'88) received the B.S., M.S., and Ph.D. degrees from Osaka University, Japan, in 1983, 1985, and 1988, respectively. He was a Visiting Scientist with the Robotics Institute, Carnegie Mellon University, from 1988 to 1990. He is currently a Professor with the Graduate School of Engineering Science, Osaka University. His research interests include image sensing, 3-D image processing, digital archiving, and virtual reality. He is a member of ACM.