

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/220233782>

# Basic Design of Video Communication System Enabling Users to Move Around in Shared Space

**Article** in *IEICE Transactions on Electronics* · November 2009

DOI: 10.1587/transele.E92.C.1387 · Source: DBLP

CITATIONS

2

READS

44

**9 authors**, including:



**Keiji Hirata**

Future University Hakodate

**114** PUBLICATIONS **539** CITATIONS

[SEE PROFILE](#)



**Toshihiro Takada**

Nippon Telegraph and Telephone

**31** PUBLICATIONS **231** CITATIONS

[SEE PROFILE](#)



**Naomi Yamashita**

NTT

**65** PUBLICATIONS **455** CITATIONS

[SEE PROFILE](#)



**Shigemi Aoyagi**

NTT Communication Science Laboratories

**22** PUBLICATIONS **110** CITATIONS

[SEE PROFILE](#)

**Some of the authors of this publication are also working on these related projects:**



gesture recognition [View project](#)



Masters thesis at MIT EECS [View project](#)

# Basic Design of Video Communication System Enabling Users to Move Around in Shared Space

Keiji HIRATA<sup>†a)</sup>, Yasunori HARADA<sup>†</sup>, Toshihiro TAKADA<sup>†</sup>, Naomi YAMASHITA<sup>†</sup>, *Nonmembers*, Shigemi AOYAGI<sup>†</sup>, *Member*, Yoshinari SHIRAI<sup>†</sup>, Katsuhiko KAJI<sup>†</sup>, *Nonmembers*, Junji YAMATO<sup>†</sup>, and Kenji NAKAZAWA<sup>†</sup>, *Members*

**SUMMARY** We propose a 2D display and camera arrangement for video communication systems that supports both spatial information between distant sites and user mobility. The implementation of this arrangement is called the “surrounding back screen method.” The method enables users to freely come from and go into other users’ spaces and provides every user with the direct pointing capability, since there is no apparent spatial barrier separating users, unlike the case of conventional video communication systems. In this paper, we introduce two properties (“sharedness” and “exclusiveness”) and three parameters (a distance and two angles) to represent the geometrical relationship between two users. These properties and parameters are used to classify the shared spaces created by a video communication system and to investigate the surrounding back screen method. Furthermore, to demonstrate and explore our surrounding back screen method, we have developed a prototype system, called t-Room. Taking into account practical situations, we studied a practical case where two t-Rooms with different layouts are connected.

**key words:** face-to-face interaction, spatial cue, surrounding back screen method, sharedness, exclusiveness, t-Room

## 1. Introduction

Our ultimate goal is to achieve high-telepresence communication where we can speak or work with people at remote sites as if we were all in the same room. A video communication system (VCS) providing life-sized images on large displays was first proposed by Bellcore [1], [2]. This system enabled users to experience a high-telepresence. A life-sized display technique and a sound localization technique were effective in providing this sensation [3], where the voice was positioned at the person displayed on the screen, and thus users could easily identify the speaker at the remote site. Moreover, that system used the same type of curtain as the background at all of the sites, and curved desks were arranged so that they all appeared to be part of a single desk.

However, under the restricted conditions imposed by the current 2D display and capture technologies, it is difficult to reproduce the various phenomena that people take for granted when they perform face-to-face (F2F) interaction in the same room. We believe there are three key issues to consider: (1) Spatial information, (2) User mobility, and (3) Pointing to an object’s image. As for (1), spatial information\* has been shown to be critical, especially in group activities [4]. For example, gaze direction is consid-

ered very important spatial information for the regulation of turn taking during communication in large groups [5]. However, in multipoint conferencing using a conventional VCS, for example, it is difficult to exchange spatial information accurately. In such a situation, all of the remote users are displayed on the screen typically in a split or tiled manner, but naturally the real remote users are not arranged as such in the real world. In other words, the visualized layout of all users does not reflect the actual physical relationships among them.

As for (2), most conventional VCSs support spatial information only when the users stay within a given position. While most VCSs do not allow the users to move around the space, the ability to move within a shared space (i.e., mobility) is regarded as one of the important affordances of face-to-face communications. Mobility actually gives users greater flexibility in adapting to each other’s perspectives; for example, a user can move closer to see what a remote colleague is looking at [6].

As for (3), if a user points an image displayed on a screen with his/her finger, a remote user may not be able to recognize what the local user is pointing at in the manner of F2F interaction, because the pointing finger is outside of the video camera’s view angle or the image of the pointing finger conveys wrong spatial information. These phenomena occur due to the inability to exchange the correct spatial information and the lack of user mobility.

We have proposed a 2D display and camera arrangement that supports both spatial information between distant sites and user mobility [7]. The virtual space shared by remote and local users is created by installing multiple screens and video cameras so that the screens are arranged surrounding the space and the video cameras capture users standing in front of the screens opposite to the cameras. As a result, remote and local users can share the same virtual space in which they can exchange spatial information and move to any position as long as they stay close to the surrounding displays. Our 2D display and camera arrangement should be considered an attempt to simulate, to the extent possible, a 3D communication space by using current 2D display and

Manuscript received March 6, 2009.

Manuscript revised June 16, 2009.

<sup>†</sup>The authors are with NTT Communication Science Laboratories, NTT Corporation, Kyoto-fu, 619-0237 Japan.

a) E-mail: hirata@btl.ntt.co.jp

DOI: 10.1587/transele.E92.C.1387

\* (i) The physical-level information of the spatial relationships among people and objects (e.g., distance and angle), and (ii) the phenomena invoked by the physical-level information that people take for granted (e.g., gaze and gestures, such as body/head movements/orientations). Spatial information has also been called “a nonverbal cue.”

capture technologies. Accordingly, we believe our method reflects the cross-fertilization of these technologies, thus allowing further development as they progress.

The organization of this paper is as follows. First, we describe the design method of the proposed shared space supporting user mobility and spatial information between distant sites. Next, we present a practical implementation of our method, called “t-Room.” Then, we discuss how to preserve spatial information under a practical condition when two t-Rooms with different screen layouts are connected.

## 2. Previous Work

Given that spatial information is of such importance, various systems are being developed to support the communication of spatial information across distant sites. Hydra was a multiparty VCS designed to simulate a 4-way round-table meeting, while exchanging spatial information as in F2F interaction [8]. Each Hydra unit consisted of a camera, display, and speaker, and its size was comparable to a PDA or a cellular phone. The Hydra unit took the place a remote user, showing the a remote user’s image on its display, and each user looked at 3 Hydra units. As long as the four users stayed at fixed separated places, gazing toward someone was effectively conveyed, looking away and gazing at someone else was also conveyed, and the direction of head turning indicated who was being looked at.

MAJIC was also a multiparty VCS (a single user at each site) that projected life-size parallax-free images of users onto a large curved screen as if the users at remote sites were participating in a meeting together and sitting around a table [9]. MAJIC supported both multi-person eye contact among users and the awareness of the user’s gaze directions. Furthermore, it provided life-size portraits of users by adopting a special thin transparent film having a large number of very small hexagons printed on both sides of the film. The reality of gaze was improved, but users still had to stay at fixed places.

GAZE-2 was a gaze-aware VCS that adopted an eye-tracking system with a video tunnel to capture the user’s front facial view, and it enabled users to convey eye contact as long as each user stood still and his/her face was correctly captured [10]. Since the users’ facial images were displayed in a virtual meeting room, their layout on the screen did not reflect that in the real world, as attempted in a conventional VCS. When user A wanted to talk to user B, for example, the facial image of A was rotated toward B at a certain angle to express A’s intention being oriented to B. However, this visual effect may mitigate the system’s effectiveness in exchanging the correct spatial information.

Private Display Method was another technique for achieving eye contact, and it allowed more than two users to see several images on a single screen simultaneously so that each user could see a unique image [11]. This method employed double-lenticular screens to convey three-dimensional stereoscopic vision.

Nguyen and Canny (2005) proposed the novel concept

of “spatial faithfulness,” which means the system’s ability to preserve spatial relationships among people and objects, and built a VCS based on this concept. Their system, called MultiView, supported collaboration between remote groups of people (multiple users at each site) [4], [12]. MultiView’s screen reflected the incoming light so that its angle was equal to the angle of the outgoing light, and it enabled users to move around within the area of their site. As a result, MultiView allowed remote groups of people to exchange various kinds of visual cues typically present in F2F interaction, such as stereo vision, motion parallax (individual views depending on user’s position), and life-size images.

Considering the previous work described above, we find that VCS technology has already been developed with the aim of supporting both spatial information and user mobility, which we consider significant factors as discussed in the previous section.

## 3. Basic Design of Shared Spaces with 2D Displays and Cameras

We start by considering what geometrical relationships occur among users during F2F interaction and how they change dynamically.

### 3.1 Geometrical Relationships in Face-to-Face Interaction

Consider this example: If person A sees person B from a diagonal perspective, then person B also sees person A from a diagonal perspective (Fig. 1). The figure depicts the top view of the geometry among users, and so the dimension of height is ignored here. The apparent distance of A from B,  $r$ , is assumed to match the apparent distance of B from A. When A sees B at angle  $\theta$  with regard to the center of the room, B necessarily sees A at the rotation angle  $\rho$ , which is the same as  $\theta$ . If A moves, then the apparent direction and distance of A from B change correspondingly, and vice versa. Moreover, their positional relationship is also immediately apparent to a third person C who is looking at them from the side. These phenomena, which people take for granted during F2F interaction in the same room, are rarely conveyed by conventional VCSs. We argue that preserving the same values for these parameters,  $r$ ,  $\theta$ , and  $\rho$ , as those

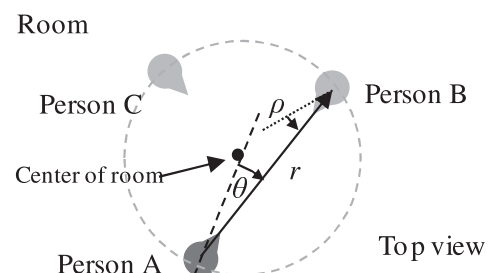
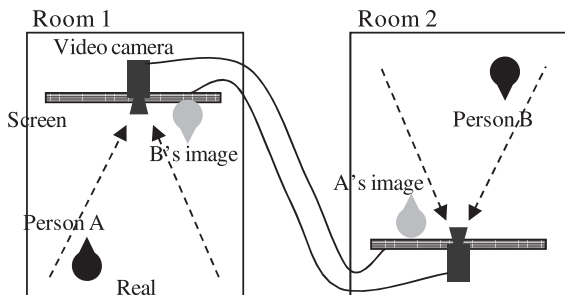
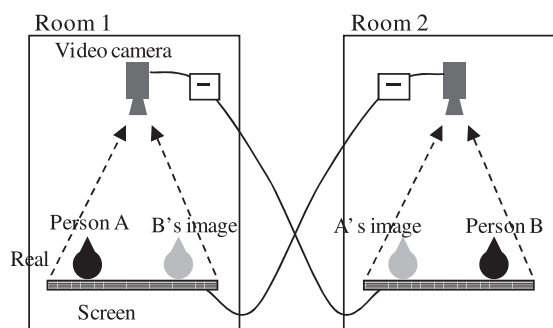


Fig. 1 The geometrical relationships among people carrying out face-to-face interaction are represented by parameters  $r$ ,  $\theta$ , and  $\rho$ .



**Fig. 2** Separating a F2F interaction space into subspaces for each user (location) and projecting remote users onto front screens, where a user faces the front of a display.



**Fig. 3** Duplicating a F2F interaction space and projecting remote users onto back screens, where a user stands with his back close to a display.

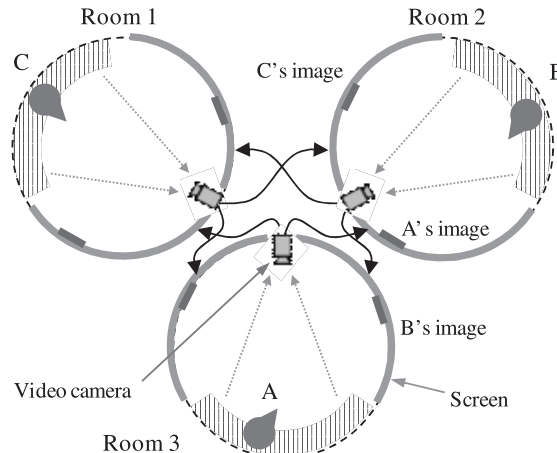
that pertain in F2F interaction would help support the spatial information of mobile users located in different rooms.

### 3.2 Splitting Space versus Reproducing Space

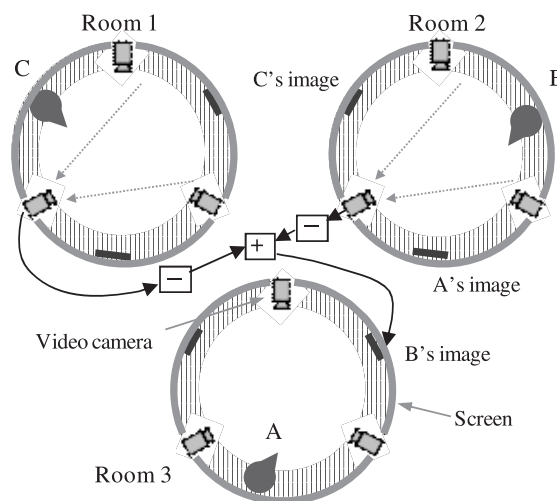
There are two basic models of layouts for reproducing F2F interaction: (1) Splitting the space of F2F interaction in Fig. 1 into subspaces for each user (location), and then projecting remote users onto front screens (Fig. 2); (2) Reproducing the space in Fig. 1 and then projecting remote users onto back screens (Fig. 3).

A conventional VCS employs model (1) in Fig. 1. Person A in Room 1 is captured by the video camera and displayed on the front screen in Room 2. Then, person B sees the image of person A. The same procedure is used for person B in Room 2. Here, a screen is constructed from various types of electronic displays (e.g., LCD, PD, and CRT), or projector screens.

In contrast, we propose the new models in Fig. 3 for achieving symmetric reproduction of images, and reproducing F2F interaction space. A feature of this model is that by having users stand around the surfaces of the screens, persons A and B are able to recognize each other's gestures, notice peripheral cues and point to real objects in the remote rooms. Here, module  $\boxed{-}$  is introduced to avoid visual echoes; the function of  $\boxed{-}$  is to extract only the light from real objects in front of the opposite screen and to cancel out the light from the screen.



**Fig. 4** Front screen method: splitting a space based on the model in Fig. 2 and projecting remote users onto front screens.



**Fig. 5** Surrounding back screen method: reproducing a space based on the model in Fig. 3 and projecting remote users onto surrounding back screens.

### 3.3 Front Screen versus Surrounding Back Screen

Following the above two basic models we consider two methods for reproducing a space of F2F interaction in VCSs: (1) Splitting a space based on the model in Fig. 2, and projecting remote users onto front screens (Fig. 4, called the *front screen method*); (2) Reproducing a space based on the model in Fig. 3, and projecting remote users onto surrounding back screens (Fig. 5, called the *surrounding back screen method*). The shaded areas in the figures depict the areas where a video camera can correctly capture the users and objects in front of the camera. With both methods, in order to project the images of users and objects so that they appear where they should appear, the physical relationships characterized by the values of  $r$ ,  $\theta$ , and  $\rho$  between local and remote users and objects must be the same as in F2F interaction. This requirement leads to the necessity of projecting

life-sized images onto the screen.

Conventional VCSs, and their variations, employ the front screen method. This arrangement is characterized by having no screens behind the user in the shaded area and the fact that the user can move around freely only within the shaded area (Fig. 4). For example, B's image is delivered to A and C through the right-hand front screen of A in Room 3 and the left-hand screen of C in Room 1. The shaded area should not overlap onto any screen.

As an alternate method, to reproduce a space, we arrange cameras and screens so that they surround the users with users standing in front of the screens (Fig. 5). As with the first method, the video camera images of B are distributed to Rooms 1 and 3, but here we need the pre-processing denoted by  $\boxed{+}$  and  $\boxed{-}$  in the figure. Module  $\boxed{-}$  is introduced to eliminate visual echoes, and the function of this module is described in the previous section. On the other hand, the added function of  $\boxed{+}$  is to overlay or superimpose the two images captured in Rooms 1 and 2 and to project them onto the correct positions. Module  $\boxed{+}$  enables us to connect more than two rooms and to show any background image that we choose. In the figure, only the necessary part of the wiring has been depicted for simplicity.

### 3.4 Sharedness and Exclusiveness

When we construct a shared space using a particular method, the method can be characterized by how it establishes the space and by the parts that each distributed room shares and does not share with the other room(s). Accordingly, we introduce two properties to express the relative characteristics of local and remote rooms: *sharedness* and *exclusiveness*.

We define sharedness as the total angle of a camera view, namely the part of a scene or perspective in a room that is projected onto the screens in the other rooms. For example, the video camera in Room 3 shown in Fig. 4 obtains a 120-degree view from the center of the circle, but does not capture the remaining part of the room. Imagine a case where an object is placed just in front of the screen to the immediate right of A. Although A can see it, neither B nor C can see it because no camera can capture an object at such a position in Room 3. Therefore, in Fig. 4, 1/3 of the scene in each space is projected to the others, and thus this ratio of 1/3 can be used as an index of sharedness.

To define exclusiveness, we first consider the area a user occupies that cannot be shared with another user. With the front screen method shown in Fig. 4, for every room there is an area excluded from the other rooms, designated by the shaded area. Although a user can freely move around within this area, the user cannot leave the area; in the figure, the area for each user spans 1/3 of the circumference. This ratio of 1/3 can be used as an index of space exclusiveness.

Next, let us consider the sharedness and exclusiveness of the surrounding back screen method shown in Fig. 5. Since "what you see is what I see" (WYSIWIS) ideally

holds for the entire field of vision of every user, the surrounding back screen method achieves full sharedness and minimum exclusiveness. Therefore, wherever a user is, spatial information is correctly exchanged between moving users.

Sharedness and exclusiveness are closely related to the capability of direct pointing; direct pointing means that a user makes the gestures of pointing to an image on a display by using his/her own fingers, and it plays a crucial role in computer-supported cooperative work as well as in F2F interaction. If a user stands near an LCD panel and points toward it, this user's body may occlude his/her own direct pointing. Therefore, for direct pointing to work correctly, all of the users must be able to see the target object at its correct position, to move to the location of the target object, and to see the body parts related to users pointing to the target object, such as their fingers. That is, full sharedness is required, and exclusiveness must be avoided. Consequently, it is difficult for users in the space employing the front screen method to engage in direct pointing, whereas users employing the surrounding back screen method can easily do this.

Some VCSs implement a shared plane using a method similar to the surrounding back screen method, which makes direct pointing possible [13]–[15]. However, because the video cameras and screens used in these systems do not surround the users, full sharedness is not achieved, and users cannot obtain an appropriate workspace awareness. Some systems, such as ClearBoard [16] and VideoArms [17], effectively recover a rich awareness of a real workspace by presenting only the parts of a body (e.g., arms and faces) within the workspace. This leads to an improved sense of presence [19] and resolves the presence disparity problem [17].

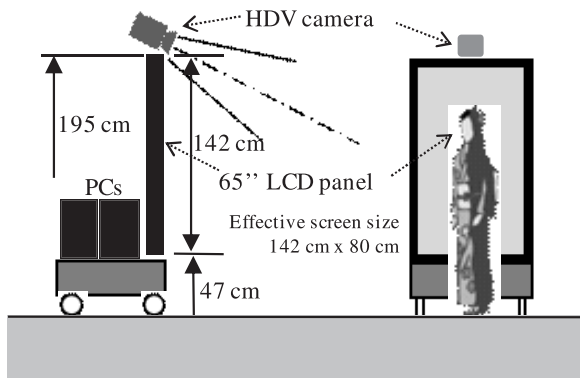
## 4. t-Room System

To demonstrate and explore the surrounding back screen method, we have been developing a prototype system, called t-Room [7], [18]. The t-Room system is designed to be as simple as possible to meet the demands of various styles of group activity.

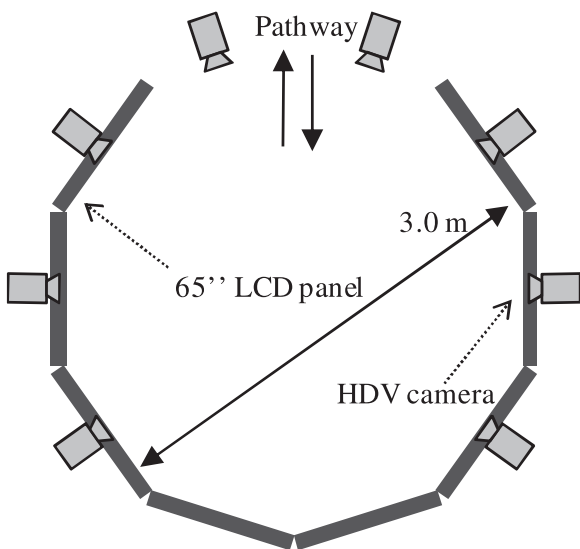
### 4.1 Hardware Design

Considering a scene in our daily life, we assume that the t-Room system supports the group activity of a few to 15 distributed users, who share physical objects and carry out pointing and gestures. To support such a scenario, we may need a 360-degree view angle display with a height comparable to an ordinary classroom or meeting room (2.5–3 m). The room's diameter, considered the diagonal line passing through the space, may span 3–5 m.

Figures 6 and 7 show the hardware configuration of the current t-Room system. A single t-Room consists of eight building modules (called Monoliths) arranged polygonally. With this setup, a t-Room surrounds a user space with LCD displays showing life-sized images. We installed



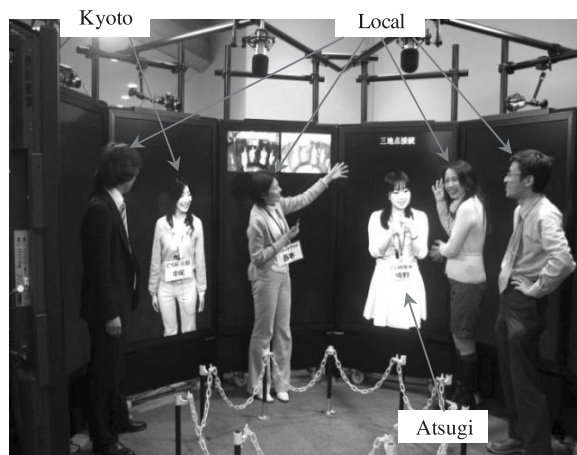
**Fig. 6** A “Monolith” building module: side view (left) and front view (right). A 65-inch display is set upright on a wheeled platform, and a video camera is placed at the top of the display.



**Fig. 7** Top view of t-Room system: eight Monoliths facing inside are arranged decagonally. Since there are no displays at the entrance pathway, there are no cameras on the opposite sides.

three nearly identical t-Rooms in our laboratories located in Atsugi City and Kyoto Prefecture (Atsugi is in the Tokyo area, and Kyoto is approximately 400 km from Tokyo). Currently, commercially available 100-Mbps optical fiber lines connect the Atsugi and two Kyoto t-Rooms.

Figure 8 shows a partial view of a working t-Room system in which four persons are standing and two others are displayed on the screen. The enclosed space is shared with other enclosed spaces by overlaying them. As a result, users can freely come from and go into others’ spaces, since there is no spatial barrier separating users such as the screen employed in a conventional VCS [20], [21]. At “local” (one of the Kyoto t-Rooms), three spaces are overlaid: Atsugi, the other Kyoto t-Room, and local itself. These spaces are similarly overlaid at Atsugi and at the other Kyoto t-Room. At local, the current images of Atsugi and Kyoto are displayed by overlaying them with a transparency ratio of 0.6. Consequently, the overlaid enclosed spaces can provide full



**Fig. 8** Demonstrating a t-Room made from eight Monoliths arranged decagonally, connecting three locations. Four local persons are standing, and two others at Kyoto and Atsugi are displayed on the screen. Five central Monoliths are shown in this photo.

sharedness and minimum exclusiveness.

#### 4.2 Visual Echo Cancellation

In the current system, visual echo cancellation (the function of  $\square$  in Fig. 2) is achieved by placing polarizing film over a video camera’s lens. Since an LCD panel inherently emits polarized light, the polarizing film enables the camera to capture the scene without capturing the image shown on the display [13]. By appropriately using polarizing film, this visual echo cancellation method provides a uniformly black video image of the surface of the LCD panel opposite the camera. An area appearing black gives the impression that there is no opaque object on the display surface, since the camera is directly aimed at the display surface. Accordingly, in the current implementation, the color black can be interpreted as being transparent.

The advantages of the current implementation include its ease of implementation and robustness. On the other hand, it has a side effect on the overlayer ( $\square$  in Fig. 5); Since black physical objects (e.g., black hair and black clothes) are interpreted as transparent, when the overlayer superimposes a front layer onto back ones, the images on the back layers can be seen through the black areas on the front layer. Occasionally, this effect renders unnatural images.

#### 4.3 Standing Position

To achieve life-size projection, users have to stand as close to the LCD surface as possible as stated in Sect. 3.3. However, it is generally not easy to make users do this, since they have an unconscious tendency to move toward something that interests them. In fact, when a user walks away from the LCD surface and moves toward the center of the t-Room, the displayed image of the user is magnified, and that person’s life-sized appearance is lost. The image displayed

in the other t-Room depends strongly on the user's standing position. Furthermore, a forward-moving user is then captured not only by the front camera but also by the others (possibly at both sides of the front camera), which leads to the display of plural images at different angles. The yellow chain in Fig. 8 is positioned to restrict such movement by the users and thus prevent this problem.

## 5. Discussion

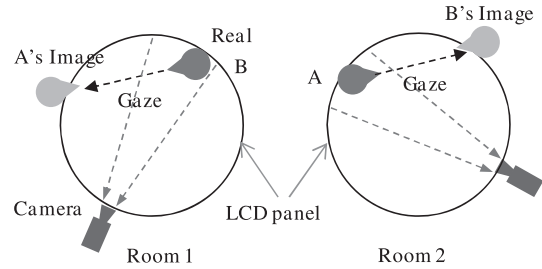
### 5.1 Gaze

The geometrical relationships known schematically in Fig. 9 illustrate people's positions in terms of gaze direction, face image orientation, and camera angle in the t-Room. In Room 1 of Fig. 9, person B looks forward and to his/her right, and the camera captures an image of B's face from the front. However, in accordance with the Mona Lisa effect [22], [23]<sup>†</sup>, the image of B's profile displayed in Room 2 appears too far to the side for person A. The same phenomenon occurs in terms of A's image as seen by B. Therefore, projecting an image at a photorealistic rotation angle may not always be appropriate for exchanging spatial information.

On the other hand, the camera is fixed at the top of the display in the t-Room. Therefore, the person's face in the image appears to be looking slightly downward. It was reported that the detecting limit for eye contact in F2F communication varies depending on the horizontal, upward, and downward directions of a camera angle; in the horizontal direction, it is 4.5 degrees, the upward direction 12 degrees, and the downward direction 8 degrees [24]. With the t-Room, the camera is fixed at a position 205 cm from the floor. For example, when a person with a height of 170 cm stands near the front of the display, the camera angle will be around 10 degrees. This implies that a person in front of the display can observe his/her counterpart's image within nearly acceptable error range.

Since the eight displays are positioned in a decagonal configuration in the t-Room, the angle of the each adjacent display is 144 degrees. Thus, the viewing angle measured from the surface normal of the nearest display is around 54 degrees. Such a very shallow angle might cause the observer to perceive the displayed image incorrectly [25]. This perceptual failure might also affect the gaze error. This remains an interesting issue for future consideration in refining layouts based on surrounding displays such as the t-Room.

When using the t-Room, the awareness of users' positions and body/face orientations are dynamically organized and shared among users in different t-Rooms through the processes of interaction and collaboration. As a result, we sometimes observed that this awareness partly compensates for the gaze error induced by the factors mentioned above. Therefore, a t-Room user is considered capable of imagining who sees whom.



**Fig. 9** The phenomena of gaze are influenced by the face image orientation and the position at which a user looks at the face image. The image of B's side-view facial profile is displayed in Room 2, and A's profile is displayed in Room 1.

### 5.2 Sharedness and Exclusiveness under Practical Situations

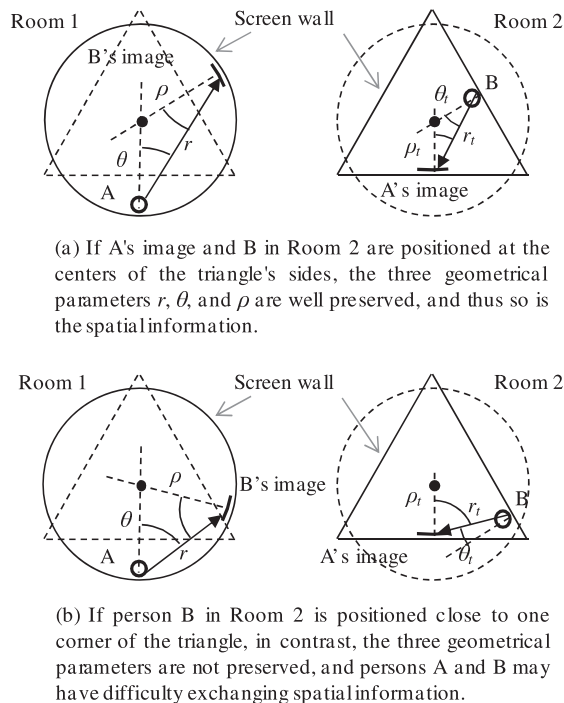
Here, we examine the possibility of distinguishing different display layouts from each other based on the two properties (sharedness and exclusiveness) and three parameters ( $r$ , the distance between A and B,  $\theta$ , the direction of local person to the normal of the back screen, and  $\rho$ , rotation angle of opposite person's image to the normal of back screen) introduced in Sect. 3.1. Even if full sharedness and minimum exclusiveness are achieved, when two t-Rooms with different layouts are connected, we will still find areas where mobile users cannot exchange correct spatial information. Now, assume that one t-Room (Room 1) occupied by person A always has a circular screen layout, while the other t-Room (Room 2) occupied by person B has a triangular one (Fig. 10). The figure depicts the top view, so the height dimension is disregarded. We assume that the two connected t-Rooms have the same overall circumference for the purpose of projecting life-sized images. A camera captures a user's image from a direction normal to his/her back screen, namely perpendicular to the back screen surface<sup>††</sup>.

Figure 10 shows the values of  $r$ ,  $\theta$ , and  $\rho$  for Room 1 (circle) and those of  $r_t$ ,  $\theta_t$ , and  $\rho_t$  for Room 2 (triangle), and it depicts two patterns of user position: (a) A's image and B in Room 2 are positioned at the centers of the triangle's sides, and (b) person B in Room 2 is positioned close to one corner of the triangle. For (a) in the figure, the values of  $r$ ,  $\theta$ , and  $\rho$  are well preserved, and only  $r$  is shortened to  $r_t$  in Room 2. In contrast, for (b), we find that  $\rho$  is positive, while  $\theta_t$  is negative<sup>†††</sup>, which may cause a misleading spatial information, although full sharedness holds. This suggests that users should not move close to acutely angled corners.

<sup>†</sup>A portrait that appears to gaze at an observer will "follow" the observer as he/she moves around and views the image from different angles, despite the motion of the observer inducing relative rotation of the portrait.

<sup>††</sup>In fact, a camera cannot always capture a user from the direction normal to his/her back screen when he/she appears in the area away from the center of the view angle. However, in this discussion, this particular phenomenon is ignored.

<sup>†††</sup>Regarding the rotation angle, the clockwise direction is assumed to be negative, and the counter clockwise direction positive.



**Fig. 10** The t-Room with a circular screen layout is connected to that with a triangular screen layout.

## 6. Conclusions

The contribution of this paper is the surrounding back screen method for designing a shared space in a VCS; the method enables users to accurately exchange spatial information, supports user mobility and provides every user with the direct pointing capability, unlike the case of conventional video communication systems. To investigate the methods to characterize a shared space and distinguish different layouts of displays, cameras, and users that construct a shared space from each other, we introduce two properties and three parameters for classifying the shared space: sharedness and exclusiveness, and  $r$ ,  $\theta$ , and  $\rho$ . These properties and parameters enable us to compare several 2D display and camera layouts and examine their strong and weak points in a more quantitative manner. Using the properties and parameters, we have appropriately examined how well users in two t-Rooms can exchange spatial information.

As t-Rooms become more widely deployed more layout patterns of the 2D displays and cameras will emerge. In a given layout, in spite of achieving full sharedness and minimum exclusiveness, the t-Room system may not accurately transfer the spatial information of mobile users due to its particular shape. Consequently, we believe it is important to deepen our theoretical understanding of such layout patterns and to investigate the effects of layouts on collaborative work in practical situations.

## References

- [1] R.S. Fish, R.E. Kraut, and L. Chalfonte, "The videowindow system in informal communications," Proc. CSCW'90, pp.1–11, 1990.
- [2] E.J. Addeo and H. Tome, "Invited address: The videowindow research prototype: An experiment with a new form of audio/video teleconferencing," Proc. SID'92 DIGEST, pp.927–930, 1992.
- [3] I. Namiki, Y. Meguro, S. Aoki, K. Irie, T. Nomura, and K. Saitoh, "Development of multimedia videoconference system," IEICE Trans. Commun. (Japanese Edition), vol.J80-B-I, no.6, pp.338–347, June 1997.
- [4] D. Nguyen and J. Canny, "Multiview: Spatially faithful group video conferencing," Proc. CHI2005, pp.799–808, 2005.
- [5] A. Kendon, *Conducting Interaction — Patterns of Behavior in Focused Encounters*, Cambridge University Press, 1990.
- [6] R.E. Kraut, S.R. Fussell, S.E. Brennan, and J. Siegel, "Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work," in *Distributed Work*, ed. P.J. Hinds and S. Kiesler, pp.137–162, The MIT Press, 2002.
- [7] K. Hirata, Y. Harada, T. Takada, N. Yamashita, S. Aoyagi, Y. Shirai, K. Kaji, J. Yamato, and K. Nakazawa, "Video communication system supporting spatial cues of mobile users," Proc. CollabTech 2008, IPSJ, pp.122–127, 2008.
- [8] A.J. Sellen, "Speech patterns in video-mediated conversations," Proc. CHI'92, pp.49–59, 1992.
- [9] K. Okada, F. Maeda, Y. Ichikawa, and Y. Matsushita, "Multiparty videoconferencing at virtual social distance: MAJIC design," Proc. CSCW'94, pp.385–393, 1994.
- [10] R. Vertegaal, I. Weevers, C. Sohn, and C. Cheung, "GAZE-2: Conveying eye contact in group video conferencing using eye-controlled camera direction," Proc. CHI2003, 2003.
- [11] K. Nakazawa, S. Shiwa, T. Komatsu, and S. Ichinose, "Proposal of a new eye contact method for teleconferences," IEICE Trans. Commun., vol.E76-B, no.6, pp.618–625, June 1993.
- [12] D. Nguyen and J. Canny, "Multiview: Improving trust in group video conferencing through spatial faithfulness," Proc. CHI2007, pp.1465–1474, 2007.
- [13] J.C. Tang and S.L. Minneman, "VideoDraw: A video interface for collaborative drawing," Proc. CHI'90, pp.313–320, 1990.
- [14] O. Morikawa and T. Maesako, "HyperMirror: Toward pleasant-to-use video mediated communication system," Proc. CSCW'98, pp.149–158, 1998.
- [15] H. Kuzuoka, J. Yamashita, K. Yamazaki, and A. Yamazaki, "Agora: A remote collaboration system that enables mutual monitoring," Proc. CHI'99 Extended Abstracts, pp.190–191, 1999.
- [16] H. Ishii, M. Kobayashi, and K. Arita, "Iterative design of seamless collaboration media," Commun. ACM, vol.37, no.8, pp.84–97, 1994.
- [17] A. Tang and S. Greenberg, "Supporting awareness in mixed presence groupware," ACM CHI Workshop on Awareness systems: Known Results, Theory, Concepts and Future Challenges, 2005.
- [18] K. Hirata, Y. Harada, T. Takada, S. Aoyagi, Y. Shirai, N. Yamashita, K. Kaji, J. Yamato, and K. Nakazawa, "t-Room: Next generation video communication system," Proc. World Telecommunications Congress 2008, IEEE, 2008.
- [19] W. Buxton, "Telepresence: Integrating shared task and person spaces," Proc. Graphics Interface'92, pp.123–129, 1992.
- [20] N. Yamashita, K. Hirata, S. Aoyagi, H. Kuzuoka, and Y. Harada, "Impact of seating positions on group video communication," Proc. CSCW2008, pp.177–186, ACM, 2008.
- [21] N. Yamashita, K. Hirata, T. Takada, Y. Harada, Y. Shirai, and S. Aoyagi, "Effects of room-sized sharing on remote collaboration on physical tasks," IPSJ Digital Courier, vol.3, pp.788–799, 2007.
- [22] V. Bruce and A. Young, *In the eye of the beholder, The science of face perception*, Oxford University Press, 1998.
- [23] D. Todorović, "Geometrical basis of perception of gaze direction,"



Vision Research, vol.46, pp.3549–3562, 2006.

- [24] K. Nagata and A. Miura, "A study of picture quality and video telephone," 1970 IEEE Int. Conv. Dig., pp.250–251, March 1970.
- [25] D. Vishwanath, A.R. Girshick, and M.S. Banks, "Why pictures look right when viewed from the wrong place," Nature Neuroscience, vol.8, no.10, pp.1401–1410, 2005.
- [26] B. Reeves and C. Nass, The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places, Cambridge University Press, 1998.



**Keiji Hirata** is Computer Scientist, NTT Communication Science Laboratories. He received the D.Eng. degree in information engineering from the University of Tokyo, in 1987. He joined NTT Basic Research Laboratories in the same year. He spent 1990 to 1993 at the Institute for New Generation Computer Technology (ICOT), where he was engaged in R&D of parallel inference machines. In 1999, he moved to NTT Communication Science Laboratories.

Currently, he is heading the project team of the video communication system t-Room. His research interests include computer-supported cooperative work and musical knowledge processing. He received the IPSJ Best Paper Award from the Information Processing Society of Japan (IPSJ) in 2001. He co-translated the book "Computer Music Tutorial" by Curtis Roads in 2001. He is a member of IPSJ, the Japanese Society for Artificial Intelligence, and the Japan Society for Software Science and Technology.



**Yasunori Harada** is Chief Technology Officer of Media Interaction Principle Open Laboratory, NTT Communication Science Laboratories. He received the Dr.Eng. degree in information engineering from Hokkaido University, Hokkaido, in 1992. He joined NTT Basic Research Laboratories in the same year. He spent 1998 to 2001 at PRESTO-JST (Precursory Research for Embryonic Science and Technology, Japan Science and Technology Agency). His research interests include visual language and

object-oriented programming.



**Toshihiro Takada** is Senior Researcher, Media Interaction Principle Open Laboratory, NTT Communication Science Laboratories. He received the B.E. and M.E. degrees in information science from Tokyo Institute of Technology, Tokyo, in 1986 and 1988, respectively. He joined NTT in 1988. His research interests are in networked information systems, networked computation, real-space computing, and human-computer interaction. He is a member of the Association for Computing Machinery (ACM),

JPSJ, the Japan Society for Software Science and Technology (JSSST), and the Human Interface Society of Japan.



**Naomi Yamashita** is Junior Research Scientist, Ambient Semantics Research Group, NTT Communication Science Laboratories. She received the B.E. and M.E. degrees in applied mathematics and physics and the Ph.D. degree in informatics from Kyoto University, Kyoto, in 1999, 2001, and 2006, respectively. She joined NTT in 2001. She is interested in computer supported collaborative work and interaction analysis.



**Shigemi Aoyagi** is Research Scientist, Media Interaction Principle Open Laboratory, NTT Communication Science Laboratories. He received the B.E. and M.E. degrees in information science from Tokyo Institute of Technology, Tokyo, in 1988 and 1990, respectively. He joined NTT in 1990. His current research interests include parallel processing, distributed algorithms, image understanding, object recognition, and distributed systems for multimedia content. He is a member of IPSJ and JSSST.



**Yoshinari Shirai** is a researcher at the NTT Communication Science Laboratories at NTT Corporation. He received a B.A. in environmental information and an M.M.G. (Master of Media and Governance) from Keio University in 1998 and 2000, respectively. His research interests include ubiquitous computing and interaction design. He is a member of the Information Processing Society of Japan (IPSJ) and the Human Interface Society.



**Katsuhiko Kaji** is Research associate, Ambient Semantics Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories. He received the Ph.D. in information science from Nagoya University in 2007. He joined NTT Communication Science Laboratories the same year. His research interests include web annotation and computer human interaction. He is a member of IPSJ and JSAT.



**Junji Yamato** is Senior Research Scientist, Supervisor, Group Leader, Media Recognition Research Group, NTT Communication Science Laboratories. He received the B.E., M.E., and Ph.D. degrees in precision machinery engineering from the University of Tokyo, Tokyo, in 1988, 1990, and 2001, respectively, and the S.M. degree in electrical engineering and computer science from the Massachusetts Institute of Technology in 1998. He joined NTT Human Interface Laboratories in 1990. His research interests are in computer vision, human-computer interaction, and robotics.

He is a member of IEEE, ACM, and the Robotics Society of Japan.



**Kenji Nakazawa** is Executive Manager, Media Information Laboratory, NTT Communication Science Laboratories. He received the B.E. and M.E. degrees in electronics, and the Ph.D. degree in materials science and engineering from Kanazawa University, Ishikawa in 1981, 1983, and 1991, respectively. He joined the Musashino Electrical Communications Laboratories of Nippon Telegraph Telephone Public Corporation (now NTT) in 1983 and was engaged in research on amorphous polycrystalline

Si thin-film materials and on how to use these films. He was also engaged in R&D of large-screen-size super-high-resolution display systems and 3D display systems. He is interested in highly realistic communication. He is a member of the Institute of Image Information and Television Engineers (ITE), and the Society for Information Display (SID).