Sharing Digital and Physical Contents on and above Distributed Surfaces using Active Infrared Keying

Buntarou Shizuki, Yuichi Murata, and Jiro Tanaka
Department of Computer Science,
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8573, JAPAN
{shizuki,murata,jiro}@iplab.cs.tsukuba.ac.jp

***Abstract*— This paper presents SyncSurface, a system framework for distributed collaboration, which shares digital contents and physical contents on and above distributed surfaces. In such sharing, one important technique is to segment each frame, in a video stream from a camera placed above a surface, into its foreground (e.g., objects and users' bodies) and its background (i.e., displays). This paper describes a simple and low-cost segmentation technique called "active infrared keying" that uses a display that emits infrared light from its surface. We have implemented two systems using the framework to examine the feasibility of the technique. The results of our pilot studies indicate that the technique enables users to communicate with each other using physical miniatures on the surfaces, gestures for pointing out physical objects, and facial expressions.**

***Keywords*—tabletop, large display, whiteboard, CSCW, groupware, gestures, shared spaces, chroma keying, image processing, user interface.**

I. Introduction

During traditional co-located collaboration using surfaces (e.g., tables and whiteboards), physical contents, other than writing (e.g., documents and sketches), can also be shared and play important roles in collaboration. For example, miniatures of furniture are useful for discussing building plan; they are placed and moved by users on room maps on horizontal surfaces. Moreover, physical objects above ("in front of" in the case where vertical surfaces are used) surfaces are also important. For example, users hold miniatures in the air during discussions. Furthermore, users' hands, bodies, and faces are also physical objects above surfaces that play important roles in collaboration. That is, they use their hands and bodies to give gestures (e.g., deictic gestures such as pointing and tapping to point out some physical object on a surface), and their faces to express their emotion.

This also occurs in co-located collaboration using computer displays (e.g., co-writing technical documents and pair-programming). While users share digital contents on one or more computer displays, they also share physical objects including miniatures they hold in their hands, memoranda pasted on large displays, and themselves, providing gestures and expressing themselves.

In contrast, traditional CSCW systems support collaboration by sharing only digital contents such as still images, animations, GUI windows of applications, and cursors between distributed sites. Some CSCW systems and researchers have tried to improve distributed collaboration by sharing physical contents, including users' hands, bodies, and faces, as well as digital contents to provide rich expressiveness in communication. However, there is sill plenty to be explored, especially in the way to implement such sharing.

In this paper, we present an alternative framework, called SyncSurface (Fig. 1), for sharing digital and physical contents on and above distributed surfaces to support distributed collaboration. For such sharing, one important technique is to segment each frame, in a video stream from a camera placed above a surface, into its foreground (e.g., objects and users' bodies) and its background (i.e., displays). To this segmentation, we present a simple and low-cost technique called *active infrared keying*. The technique uses a display that emits infrared light from its surface for robust segmentation. We also describe two systems we implemented using the technique to examine its feasibility.

II. Related Work

Much research has investigated distributed collaboration by sharing physical contents on and above surfaces using cameras, aimed at the surfaces, to capture images of the contents, which are then sent to remote sites. An example

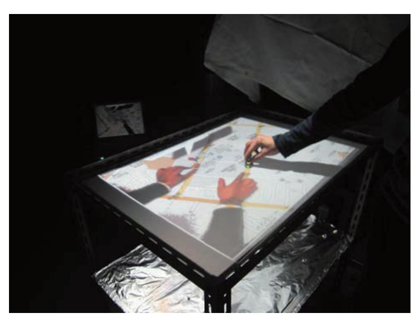

Figure 1. Communication using SyncSurface Table.

of such pioneering research is VideoDraw [1], [2]. It uses video cameras, each of which is aimed at a television set, to send physical contents including marks, which users draw with whiteboard markers directly on the surface, as well as the accompanying hand gestures on and above the surface. VideoWhiteboard [3] places cameras behind screens to capture an image of the marks that users draw on the front surface of the screens using whiteboard markers, and the shadow of the users. Then the system sends these data to the other site and projects them on the screen using a video projector. This enables the users to share physical contents: both real and remote marks, as well as shadows that contain the remote users' gestures and actions. ClearBoard-1 [4] uses a camera mounted above a surface (i.e., a half mirror with polarizing film) to obtain physical contents on and above the surface, including users facing the surface, and drawings made with whiteboard markers. The system sends the images from the camera to a remote surface and projects them onto the surface from its rear after mirror-reversing them. This allows eye contact with the remote user and makes visible each other's gestures, actions, and facial expressions as well as the drawing. Agora [5] and AgoraG [6] are tabletops that use multiple cameras at each site. The systems use cameras mounted above the surfaces to capture physical contents on and above the surfaces, and two cameras to capture users. The former images are sent to a remote site and projected onto the surface with a projector mounted above the surface. The latter images are projected onto vertical screens around the surfaces at the remote site, enabling users to use eye contact, actions, and gestures in their collaboration.

Although sharing a computer's desktop or GUI windows of applications (i.e., a kind of digital contents) among distributed sites using remote desktop software such as VNC, Apple Remote Desktop, and Microsoft Remote Desktop Connection, is the simplest and most common way for distributed collaboration, much research has been trying to share not only digital contents but also additional contents derived from physical contents by using cameras to give users awareness of other users. One such example is LIDS [7]. In LIDS, users can directly write on a horizontal display using a stylus, the position of which is tracked on the surface and is used to draw strokes in digital ink on the surface. At the same time, a camera placed behind the users captures images around the surface, including the users. The system converts the images into shadows (i.e., physical contents) and transmits them to other sites along with digital contents including the strokes. A projector placed behind the surface projects them from its rear. Thus, sharing the shadows allows the users to perform gestures, and provides the users with awareness of other users at remote sites. Distributed Tabletops [8] also shows shadows of users. The system uses cameras placed above the horizontal surfaces to capture the users' hands and other objects on and above the surfaces. A projector above each remote tabletop projects the users' shadows (i.e., physical contents) onto the the surface. VideoArms [9] extracts the users' hands and arms (i.e., physical contents) by detecting skin-colored pixels among images from cameras above the surfaces and sends them to remote surfaces, which are then combined with digital contents and presented to the remote users.

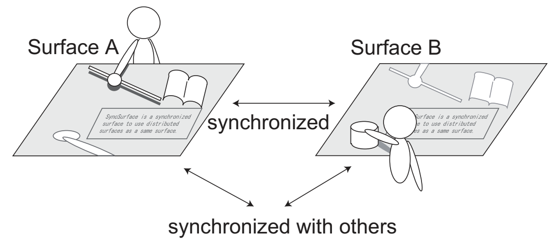

Figure 2. An implementation style of SyncSurface: table style.

Some researchers have explored sharing both digital and physical contents between distributed sites, in which cameras are also used to capture physical contents. DoubleDigitalDesk [10] uses two sets of DigitalDesk, which projects digital contents onto the desk and uses cameras above the desk to capture physical objects including paper documents on the desk, and recognizes the content, enabling the user to interact with them. By placing one set of DigitalDesk on each site, the system makes it possible to share physical contents on and above both desks as well as digital contents. ClearBoard-2 [11], [12] can also share both digital and physical contents on and above surfaces. It uses a camera mounted above a transparent digitizer with a half mirror with polarizing film to obtain physical contents on and above the surface, including users. The system uses the digitizer to capture strokes drawn on the surface. Then it sends the images from the camera and the strokes as well as a computer's desktop to a remote surface and projects them onto the surface from its rear.

Our proposed framework, SyncSurface, enables users to share both digital and physical contents on and above distributed surfaces to support distributed collaboration. In this sense, SyncSurface shares the same goals as DigitalDesk and ClearBoard-2. However, in contrast, our focus is on developing and examining a technique to achieve the sharing of both digital and all physical contents (e.g., books, miniatures, users' bodies and hands) on and above the surfaces in a low-cost and robust manner. To this end, we developed a technique to extract foreground objects from the background. We implemented two systems using this technique, and discuss its feasibility based on the results.

III. SyncSurface

SyncSurface is a system framework for distributed collaboration, which shares digital contents and physical contents on and above distributed surfaces. Fig. 2 illustrates distributed collaboration using SyncSurface. The system displays the same digital contents on all the surfaces located at multiple distributed sites. At the same time, it shares physical contents. Specifically, it captures physical objects on and above Surface A using cameras, then displays them on Surface B and other surfaces (if any). At the same time, the system also captures physical objects on and above Surface B, and displays

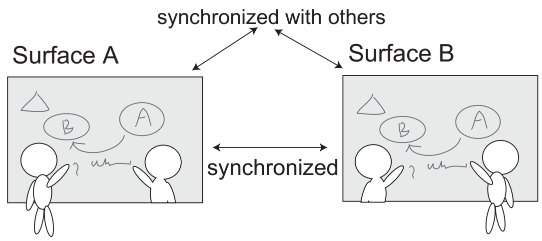

Figure 3. Another implementation style of SyncSurface: board style.

them on Surface A and others. Because all of the physical objects on and above all of the surfaces are displayed on all of the surfaces, distributed users can collaborate using these objects and hand gestures almost in the same manner in which they can engage in co-located collaboration using one physical surface.

There are two possible implementation styles in SyncSurface: table style (Fig. 2) and board style (Fig. 3). Each has its own suitable applications. In table style, the users can place and move physical objects on the surface readily. Thus, this style suits designing, planning, and prototyping. For example, remote users can collaboratively examine the arrangement of the furniture in their new office by displaying a floor map of the office (i.e., digital contents) and then arranging various furniture models (i.e., physical contents) on the floor map. In contrast to table style, the users can use gestures that they use in face-to-face communication in board style, because this style presents the hands, bodies, and faces of the users in front of the surfaces. Thus, this style is suitable for distant presentation and remote lectures.

IV. Active Infrared Keying

To share physical contents on and above distributed surfaces, it is necessary to segment each frame, in a video stream from the camera placed above the surface, into its foreground (e.g., objects and users' hands, bodies, and faces) and its background. To this end, we developed *active infrared keying* (AIR keying). This segmentation technique uses a display emitting infrared light from its surface, achieving robust segmentation with a simple setup.

A. Focus

Techniques to segment regions of foreground objects within images from a camera have been heavily investigated and are widely used.

Many techniques use special backgrounds or utilize visual features of the foreground objects. For example, some techniques use a blue screen as the background, some extract regions where brightness is higher than a threshold, some use skin color-based detection, some extract regions where the shapes are the same as the target objects, some compare them with images captured beforehand, and others use polarizing films.

There are also keying techniques that use sensing technologies in addition to a camera. For example, [13] uses a thermographic camera to detect human regions. This technique is suited to segmentation of human regions while other kinds of objects are recognized as background. A depth-sensing camera can be used to recognize the regions of foreground objects. This method is attractive because it does not impose any visual precondition on the foreground or background objects. However, it is difficult to detect regions of thin objects (e.g., paper) placed on surfaces, because the depths of the regions are almost the same as those of the surfaces.

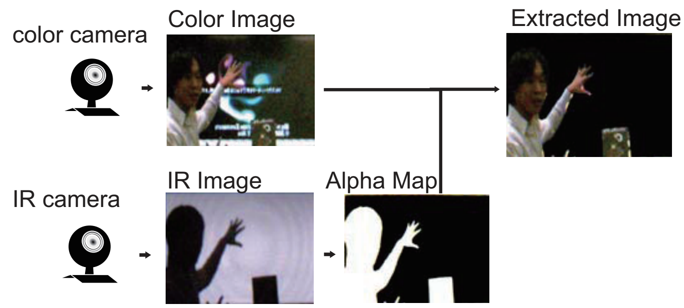

Figure 4. Image processing steps of active infrared keying.

In contrast, our AIR keying takes a more active approach. The technique uses a display emitting infrared light from its surface as the background. Within the images from an infrared camera placed above the surface, it identifies regions where no infrared appears as a foreground region. Thus, it can extract regions of objects that no infrared light penetrates, and does not impose any visual precondition on foreground objects. While this technique does not detect transparent objects that infrared light penetrates, this is not a serious problem because most physical contents used in collaborations, such as paper, stationery, and miniatures, are opaque.

The approach of our technique is similar to that of IR Matte [14], which uses a retroreflector sheet as the background. Infrared light is provided from a ring of bright LEDs attached around the camera lens. In contrast to IR Matte, our contribution is to provide an alternative system, which is simple and low-cost; consequently it is easy to implement.

B. Process of AIR keying

AIR keying uses two cameras: a color camera and a camera with an IR pass filter (IR camera). The two cameras are placed in front of the surface, and a half mirror is situated between them so that the two cameras share the same optical axis (see Fig. 5, left).

The image processing steps of AIR keying proceeds as follows (Fig. 4):

1) The IR camera produces an IR Image, with a display emitting infrared light from its surface as the background. In this image, regions corresponding to foreground objects, including persons, are dark.
2) AIR keying uses alpha keying (as in [13]) to make edges of the foreground regions appear natural in the resulting image. To this end, AIR keying constructs an Alpha Map

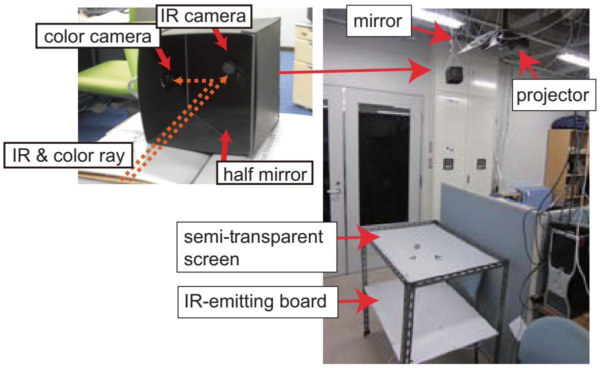

Figure 5. Left: a color camera, a camera with an IR pass filter, and a half mirror that is used so that the two cameras share the same optical axis. Right: our prototype of SyncSurface Table.

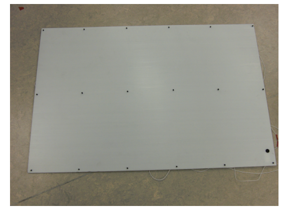

Figure 6. IR-emitting board (before being coated with aluminum foil) that we developed for our prototype of SyncSurface Table.

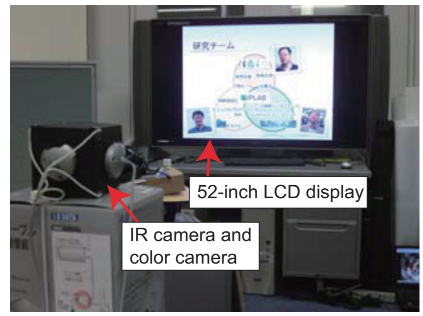

Figure 7. Prototype of SyncSurface Board.

by applying a function with two thresholds to the IR Image.

3) AIR keying uses the IR Map to subtract the background from the Color Image. The result is the foreground image with alpha values (Extracted Image).

The system executes this process at each site, and sends the Extracted Image to other sites. Then, each surface obtains a synthesized image by collecting the Extracted Images from other sites and superimposing them on the digital contents. Finally, the surface displays the synthesized image.

C. Display that emits infrared light from its surface

An important component of AIR keying is a display that emits infrared light from its surface. There are two low-cost approaches to realize this:

- The first approach is to place an infrared projector behind a screen, in addition to a video projector. The video projector projects digital contents onto the screen from its front surface. In contrast, the infrared projector projects infrared light onto the screen from its rear. As a result, it is possible to obtain such a display. The merit of the first approach is that robust segmentation is achieved simply by using a powerful infrared projector.
- The second approach is to use an LCD display with back lighting that emits infrared light. The advantage of this approach is that the setup of the implementation is relatively smaller than the first approach.

V. Study Prototypes

We implemented two prototypes of SyncSurface to examine AIR keying and the communication using SyncSurface: a prototype of SyncSurface Table using the first approach and a prototype of SyncSurface Board using the second approach.

Fig. 5 shows our prototype of SyncSurface Table. The prototype has a video projector (and a mirror to reflect the video from the projector) mounted in the ceiling for projecting digital contents downward onto the semi-transparent screen (80 cm $\times$ 60 cm, roughly 40-inch). Below the screen, the prototype also has a IR-emitting board (same size as the screen) as an infrared projector. The IR-emitter board has an array of 3×6 infrared LEDs on the surface (Fig. 6). We coated the surface of the board with aluminum foil to prevent the board from absorbing the infrared light from the LEDs. The infrared light is captured with an IR camera above the surface, mounted in the ceiling. A color camera is also mounted to capture physical contents on and above the surface. The video projector can be scaled arbitrarily in terms of the size of the surface simply by using a larger board with more LEDs on its surface.

Fig. 7 shows our prototype of the SyncSurface Board. We used a 52-inch LCD display in this prototype. The same cameras with a half mirror described in Fig. 5 were used to build this prototype.

In both implementations, before executing the process described in Section IV-B, a homography is applied to each frame in the video streams from the color camera and the IR camera. To this end, a calibration is necessary for each camera when the surface is set up.

VI. Pilot Studies

We conducted two pilot studies to examine AIR keying and the communication using SyncSurface.

A. Examining SyncSurface Table in remote communication

The goal of this pilot study was to examine whether users can communicate naturally with each other using SyncSurface Table. More specifically, we were interested in examining whether users can see both digital and physical contents on and above the remote surface naturally, use gestures (especially deictic gestures such as pointing and tapping) to physical contents on the remote surface, and use gestures to digital

Figure 8. Digital contents and physical contents are projected onto the users' hands (center) and onto the notebook (upper-left) that a user holds in his hand.

contents (i.e., observe the intention of gestures by remote users).

Setup

We built two separate sets of SyncSurface Table. We displayed a map of our university as the digital content and placed a miniature car on each SyncSurface Table. Two participants were involved. We asked one participant (host) to explain a route to our university to another participant (guest) (see Figs. 1 and 8). After the task, we interviewed both participants.

Results

Both participants said that they could see the place, color, and shape of the car of its counterpart. The host said that he could readily use the car to explain the route by moving it along the route on the map and to suggest a parking lot by placing the car on the map. At the same time, the guest said that he could understand the route and the parking lot.

Moreover, we observed that both participants frequently use deictic pointing gestures to indicate places on the map. The host used them for explanation. The guest used them for confirmation. This result suggests that users could effectively and naturally use gestures to digital contents on the SyncSurface.

However, we identified three problems in this study. First, when a participant picked up the car with his hand, the other participant could not see the car because the hand largely covered it. Second, as shown in Fig. 8, the system projected the digital contents and the physical contents of the remote site onto the participant ' s hands and onto the notebook that one participant held in his hand, because we used a front projection technique in this study. Third, the participants said that they sometimes felt frustration during communication, because one could not touch and move the physical contents of the remote site directly while he could see them.

B. Examining SyncSurface Board in remote presentation

There were two goals in this pilot study: to assess the second approach of AIR keying (i.e., using an LCD display as the display that emits infrared light from its surface) and to evaluate the efficacy of communication with physical objects visible only in front of a board.

Figure 9. Pilot study of SyncSurface Board.

Figure 10. Left: a color image whose background is subtracted using AIR keying (Extracted Image). Right: the result shown on SyncSurface Board, which is the color image synthesized with the digital content.

Setup

We built one SyncSurface Board (due to a lack of our hardware resources) and placed it in a remote room (Fig. 7). Then one author of this paper (presenter), standing before a prototype of SyncSurface Board, conducted a remote presentation, using slides as well as body gestures. We asked a participant to observe the remote presentation in another room, where we showed the synthesized images on a 50-inch plasma display (Fig. 9), almost the same size as the prototype of SyncSurface Board. After the presentation, we interviewed the participant.

Results

The participant said that he could see the gestures of the presenter during the presentation. Moreover, as shown in Fig. 10, the display clearly showed the actions and facial expressions of the presenter.

One problem revealed by this pilot study was the presence of noise in the images of physical contents.

VII. Discussion

In the first pilot study described in Section VI-A, SyncSurface Table successfully provided the users with shared digital and physical contents for communication. However, because the two sets of SyncSurface Table were connected to the same local network in this pilot study, it is necessary to further examine its effectiveness in various cases, especially in a case where there is severe network latency.

The first and third problems observed in the first pilot study (described above) are inherent to distributed communication and are still open for discussion. However, these problems are not serious, because users would quickly get used to avoiding such problems by communicating with each other.

We plan to address the second problem simply by using an Alpha Map (Fig. 4) as a mask to make the regions in the digital contents where physical contents exist black, before synthesizing the digital and physical contents. Such a system would still allow users to see and to indicate both local and remote physical content using pointing gestures.

The noise observed in the second pilot study was due to the half mirror we used. The half mirror was an acrylic panel, the reflection rate of which was low to capture the physical content using a color camera (in contrast, the infrared light penetrating the half mirror was powerful enough for keying). Thus, we will solve this problem by replacing the acrylic half mirror with one made of material that reflects infrared light well but allows visible light to penetrate.

Our pilot studies were conducted under controlled illumination, avoiding effects from external infrared light (e.g., a fluorescent light and ambient light, which also emit infrared light). In actual usage, however, such effects would need to be addressed. While the second approach requires an environment with little infrared light, the first approach can eliminate the effect, synchronizing the infrared projector with the IR camera. That is, the system turns the projector on and off repeatedly. Bright areas in IR images captured when the projector is turned off are infrared light from the environment. Thus, using bright areas found only in images captured when the projector is turned on produces the Alpha Map, where the effect of infrared light from the environment is eliminated.

VIII. Conclusions

This paper presents SyncSurface, a system framework for sharing digital and physical contents on and above distributed surfaces to support distributed collaboration. The framework utilizes *active infrared keying*, a simple and low-cost technique for extracting physical objects from camera images, by using a display that emits infrared light from its surface. The system proved useful in two different pilot studies with two different implementations (SyncSurface Table and SyncSurface Board). In the future, we will expand our system in a way that addresses these problems found in two pilot studies and use the refined system in further user studies.

References

[1] J. C. Tang and S. L. Minneman, "VideoDraw: a video interface for collaborative drawing," in *Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people*, pp. 313–320, ACM, Apr. 1990.

[2] J. C. Tang and S. L. Minneman, "VideoDraw: a video interface for collaborative drawing," *ACM Transactions on Information Systems*, vol. 9, pp. 170–184, Apr. 1991.

[3] J. C. Tang and S. L. Minneman, "VideoWhiteboard: video shadows to support remote collaboration," in *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology*, pp. 315–322, ACM, April/May 1991.

[4] H. Ishii and M. Kobayashi, "ClearBoard: a seamless medium for shared drawing and conversation with eye contact," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '92, pp. 525–532, ACM, May 1992.

[5] H. Kuzuoka, J. Yamashita, K. Yamazaki, and A. Yamazaki, "Agora: a remote collaboration system that enables mutual monitoring," in *CHI '99 extended abstracts on Human factors in computing systems*, pp. 190–191, ACM, May 1999.

[6] J. Yamashita, H. Kuzuoka, P. Luff, C. Heath, and K. Yamazaki, "Supporting paper mediated video conference with Agora system," in *Proceedings of the 2002 ACM on Computer supported cooperative work video program*, p. 2, ACM, Nov. 2002.

[7] M. Apperley, L. McLeod, M. Masoodian, L. Paine, M. Phillips, B. Rogers, and K. Thomson, "Use of video shadow for small group interaction awareness on a large interactive display surface," in *Proceedings of the Fourth Australasian User Interface Conference (AUIC2003)*, pp. 81–90, Australian Computer Society, Inc., Feb. 2003.

[8] P. Tuddenham and P. Robinson, "Distributed Tabletops: Supporting remote and mixed-presence tabletop collaboration," in *Proceedings of the Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP 2007)*, pp. 19–26, Oct. 2007.

[9] A. Tang, C. Neustaedter, and S. Greenberg, "VideoArms: Embodiments for mixed presence groupware," in *Proceedings of the 20th BCS HCI Group Annual Conference (HCI 2006)*, pp. 85–102, Springer, Sept. 2006.

[10] P. Wellner, "Interacting with paper on the DigitalDesk," *Communications of the ACM*, vol. 36, pp. 87–96, July 1993.

[11] H. Ishii, M. Kobayashi, and J. Grudin, "Integration of interpersonal space and shared workspace: ClearBoard design and experiments," *ACM Transactions on Information Systems*, vol. 11, pp. 349–375, Oct. 1993.

[12] H. Ishii, M. Kobayashi, and K. Arita, "Iterative design of seamless collaboration media," *Communications of the ACM*, vol. 37, pp. 83–97, Aug. 1994.

[13] K. Yasuda, T. Naemura, and H. Harashima, "Thermo-Key: Human region segmentation from video," *IEEE Computer Graphics and Applications*, vol. 24, pp. 26–30, January/February 2004.

[14] H. Mitsumine, Y. Yamanouchi, T. Fukaya, H. Okubo, and S. Inoue, "A novel image compositing method considering studio lighting – IR Matte –," *Proceedings of the IEICE General Conference*, vol. 2008, p. 103, Mar. 2008. (in Japanese).