

Tiled Stereo Display System for Immersive Telemeeting

Ig-Jae Kim*, Sang Chul Ahn, and Hyoung-Gon Kim

Abstract

In this paper, we present an efficient tiled stereo display system for tangible meeting. For tangible meeting, it is important to provide immersive display with high resolution image to cover up the field of view and provide to the local user the same environment as that of remote site. To achieve these, a high resolution image needs to be transmitted for reconstruction of remote world, and it should be displayed using a tiled display. However, it is hard to transmit high resolution image in real time due to the limit of network bandwidth, and so we receive multiple images and reconstruct a remote world with received images in advance. Then, we update only a specific area where remote user exists by receiving low resolution image in real-time. We synthesize the transmitted image to the existing environmental map of remote world and display it as a stereo image. For this, we developed a new system which supports GPU based real time warping and blending, automatic feature extraction using machine vision technique.

Keywords : Tiled display, Telemeeting, Real-time warping, GPU

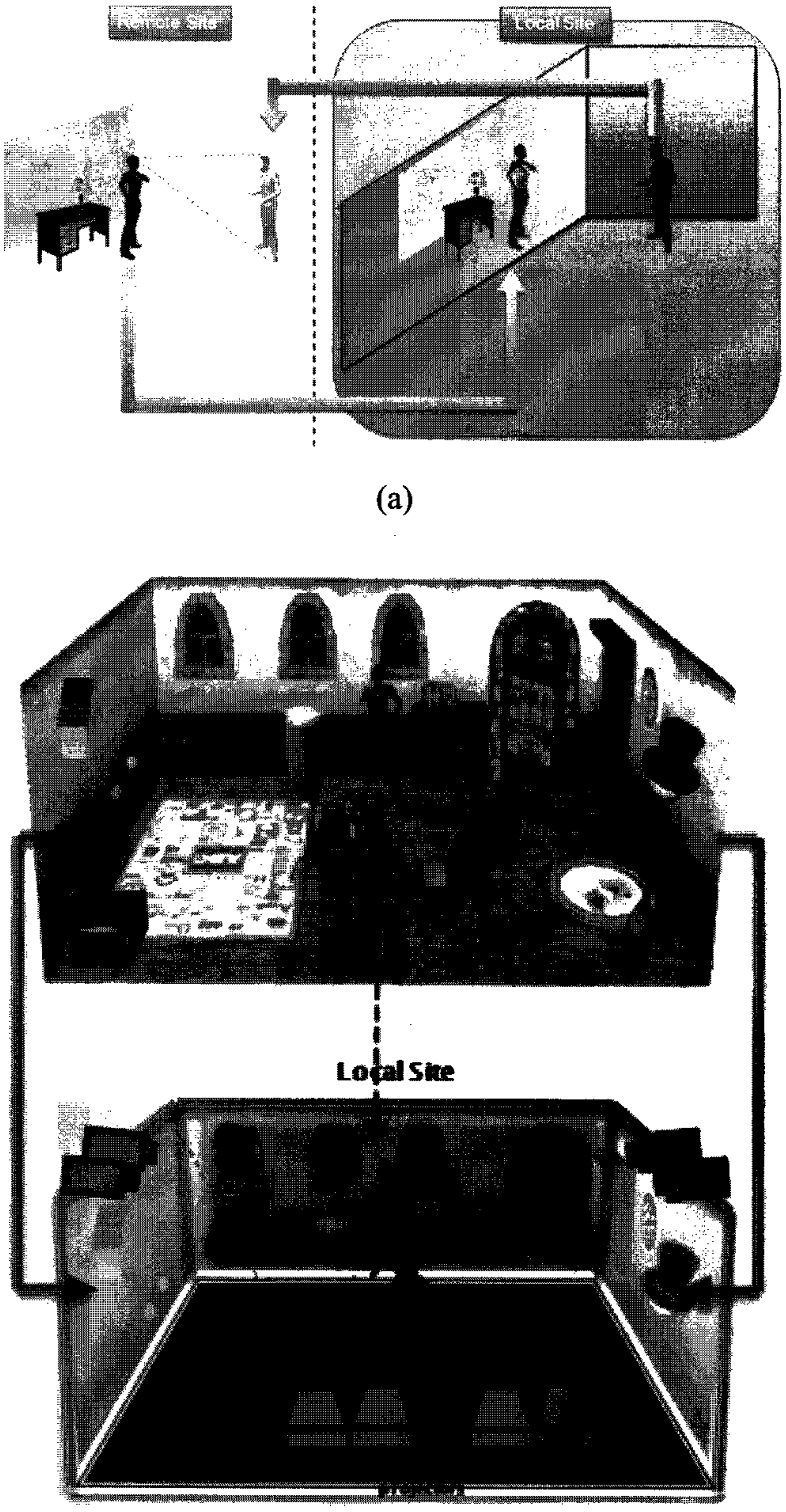
1. Introduction

Immersive and tangible meeting is a new system that enables a user to share a virtual space with remote participants. The main objective is to offer enriched communication modalities that is similar as those used in the face-to-face meetings. This includes gestures, gaze awareness and realistic stereo images. This allows the user to become immersed and have interaction with virtual objects including remote participants. This helps overcome the limitations observed in both the conventional video-based telecommunication and also the VR-based collaborative virtual environment approaches. For immersive and tangible meeting with a remote user, it is important to provide large scale display with high resolution to cover up the field of view. To display high resolution image of remote environment, we must have high resolution image source. However, the transmission of high resolution images needs high network bandwidth. Although we can use gigabit Ethernet owing to the development of network device industry, we still need higher bandwidth network to transfer the high resolution image which can cover the cubical large screen environment, especially for stereo display. In this paper, we propose a new system for efficient large scale display. Fig. 1 shows the conceptual view of our proposed system. From this figure, you can see that a user in local site can see a remote user and its actual environment in our system. It makes the user feel as if s/he is actually at that remote site. To do this, we first construct a high resolution environmental map[1] for remote site in advance and then display to each wall with 4 projectors in large screen environment. However, we don't need to update the whole remote environmental map because it doesn't change dynamically, except for the area around the user. If we can segment the remote user from a camera image automatically at real-time, we only need to synthesize it into the environmental map.

However, it is not easy to segment an object from natural scene in real-time. Therefore, we only receive a frame, which includes a remote user, captured by remote-camera through the Ethernet. Then, we synthesize the transmitted frame to the high resolution environmental map. To synthesize the transmitted image to the environmental map, we have to extract the feature points and then match it

Manuscript received November 10, 2007; accepted for publication December 25, 2007.
* Member, KIDS
Corresponding Author : Ig-Jae Kim
Imaging Media Research Center, Korea Institute of Science and Technology
39-1, Hawolgok-dong, Seongbuk-gu, Seoul 136-791, Korea.
E-mail : kij@imr.kist.re.kr **Tel** : 02-958-5766 **Fax** : 02-958-5769

[1] Environmental map is a kind of high resolution image, which can be made with several images taken from remote environment using panoramic image technique.



(a)

(b)

Fig. 1. (a) Conceptual view of our proposed system (b) System configuration of local site with stereo projectors

to find the corresponding features. After the matching process is completed, we can warp the transmitted image to align the environmental image seamlessly. Before describing this process in more detail, we will explain how to build a tiled display system for displaying a high resolution image in the following section.

2. Warping process

It is difficult to display a high resolution image with conventional display systems, such as projectors and monitors. Therefore we have to use multiple conventional projectors by tiling several projectors seamlessly in order to display a high resolution image.

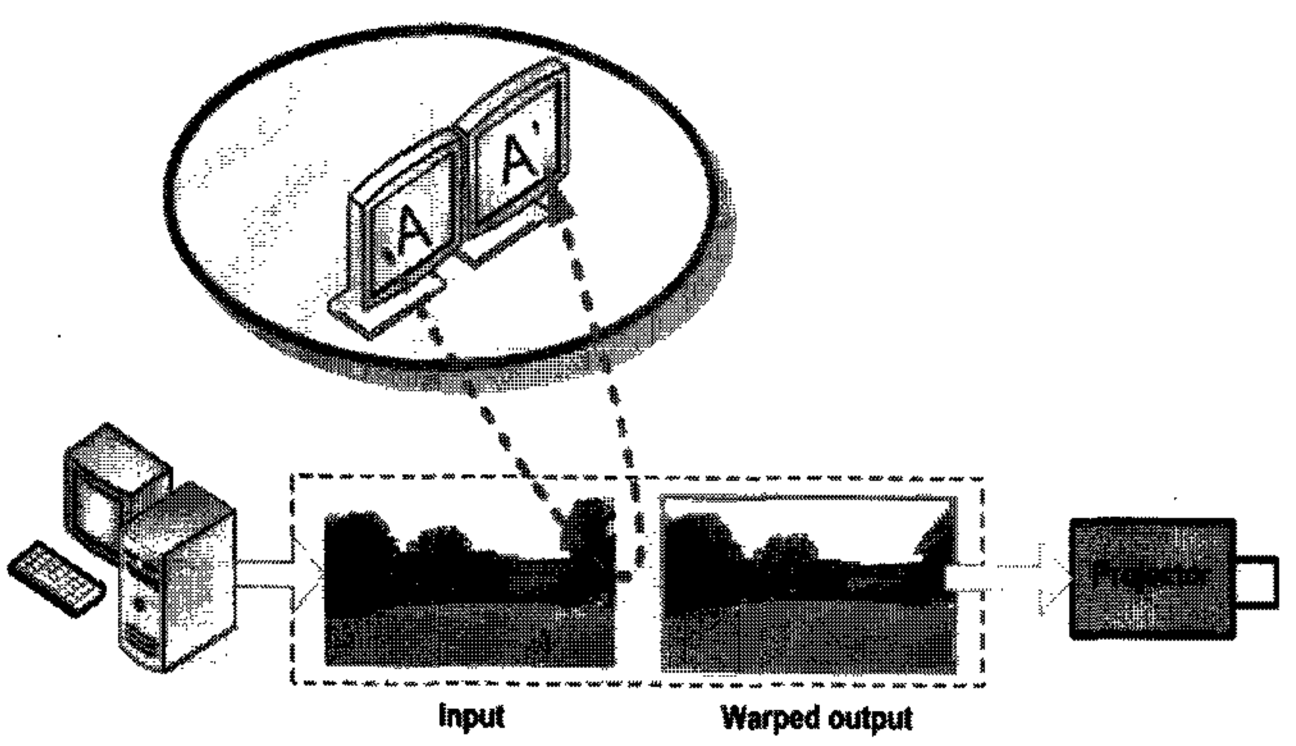
To build a tiled display, we use a real-time warping and blending method using GPU technology and implement as a distributed system to support parallel processing of each projector. Since a real-time image warping method requires large computational complexity, previous approaches have used a special hardware system to support that kind of work. Since then, the performance of GPUs have greatly improved, and so we can utilize the power of GPUs to handle the real-time warping process[1,5].

For this reason, we decided to use DirectX to control GPUs. In warping process, we use dual display functionality of graphic card. We first copy the first display surface ("input" in fig. 2) into the display buffer. In display buffer, we can manipulate the data in real time using GPU. Then, the copied data are warped with warping parameters (hereinafter referred to as "homography") calculated by the alignment process[2]. We display the input image to be warped on the first monitor in dual monitor and capture the displayed image and load it to the texture memory. After mapping it to the geometry model, we can warp the image by transforming the geometry vertices. We finally display the warped image on the second monitor.

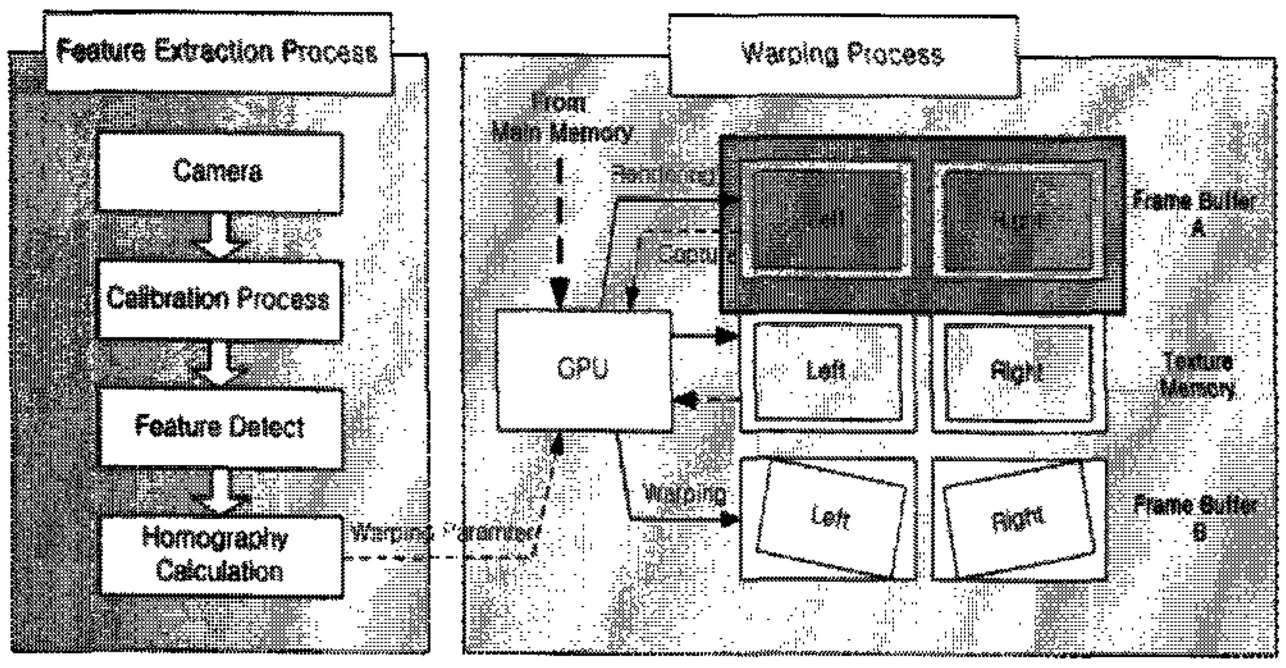
Before taking a warping process, we have to find the parameter for warping. We have developed software for automatic extraction of warping parameter. SOBEL first derivative operators are used to take the derivatives x(D_x) and derivatives y(D_y) of an image, after which a small region of interest is defined to detect corners. A 2x2 matrix of the sums of the derivatives x and y is subsequently created as follows:

$$C = \begin{bmatrix} \sum D_x^2 & \sum D_x D_y \\ \sum D_x D_y & \sum D_y^2 \end{bmatrix} \quad (1)$$

[2] Before warping, we must calculate feature points from each candidate image and find the correspondences from them. After finding the correspondences, we can calculate the homography and warp candidate images with it. We call this process alignment process



(a)



(b)

Fig. 2. Real-time warping process using GPU technology (a) systematic view (b) warping process flow.

The eigenvalues are found by solving $\det(C - \lambda I) = 0$, where is a column vector of the eigenvalues and I is the identity matrix. For the 2x2 matrix of the equation above, the solutions may be written in a closed form: equation (2)

$$\lambda = \frac{\sum D_x^2 + \sum D_y^2 \pm \sqrt{(\sum D_x^2 + \sum D_y^2)^2 - 4(\sum D_x^2 \sum D_y^2 - (\sum D_x D_y)^2)}}{2} \tag{2}$$

From physical meaning of above equation, we know that two small eigenvalues mean a roughly constant intensity profile within a window. A large and small eigenvalues correspond to a unidirectional texture pattern. Two large eigenvalues can represent corners, salt-and-pepper textures, or any other pattern that can be feature points. Therefore, we have to select a point as a corner points in order for it to meet a criterion as follows. If $\lambda_1, \lambda_2 > t$, where t is some threshold, then a corner is found at that location. This can be very useful for object or shape recognition[3]. After finding the adequate corner points, we calculate homography[3] H for warping process[4].

$$H = (h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33}) \tag{3}$$

$$A = \begin{pmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1X_1 & -y_1X_1 & -X_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1Y_1 & -y_1Y_1 & -Y_1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_n & y_n & 1 & 0 & 0 & 0 & -x_nX_n & -y_nX_n & -X_n \\ 0 & 0 & 0 & x_n & y_n & 1 & -x_nY_n & -y_nY_n & -Y_n \end{pmatrix} \tag{4}$$

Given corner points, which are $x_1,\ldots,x_n$ and $X_1,\ldots,X_n$ from left image and right image, respectively, we can make the following A matrix so that $AH = 0$, subject to $|H| = 1$. Finally, we can get the eigenvector (H) of least eigenvalue of $A^T A$.

After warping, regions of the display surface that are illuminated by multiple projectors appear brighter, making the overlap regions very noticeable to the user. Since we can detect the overlapping area from alignment process, which can be done by extracting corresponding corner points from each display surface, we applied multiple texturing with blending mask and transparent overlay technique to make seamless overlapping area[5].

3. Registration process with transmitted image

In registration process, we must combine the environmental map with a transmitted image and update the previous environmental map. As mentioned above, for network efficiency, we only update the active region of remote environment where a remote user occupies, with network cameras. In this case, we have to register the updated region to the static environmental map which can cover the field of view for remote environment and was made in advance[6]. Fig. 3 shows the process of updating active region of environmental map in local site. Fig. 3(a) shows the scene after registration process. Fig. 3(b) shows the reconstructed scene of remote environment in local site. Fig. 3(b) is a static panoramic image of remote environment. As shown in Fig. 3(c), the transmitted image is not synthesized with

[3] Homography is projective transformation that maps points from one plane to another plane. For example, the transformation mapping points in a planar surface in the world to the image plane.

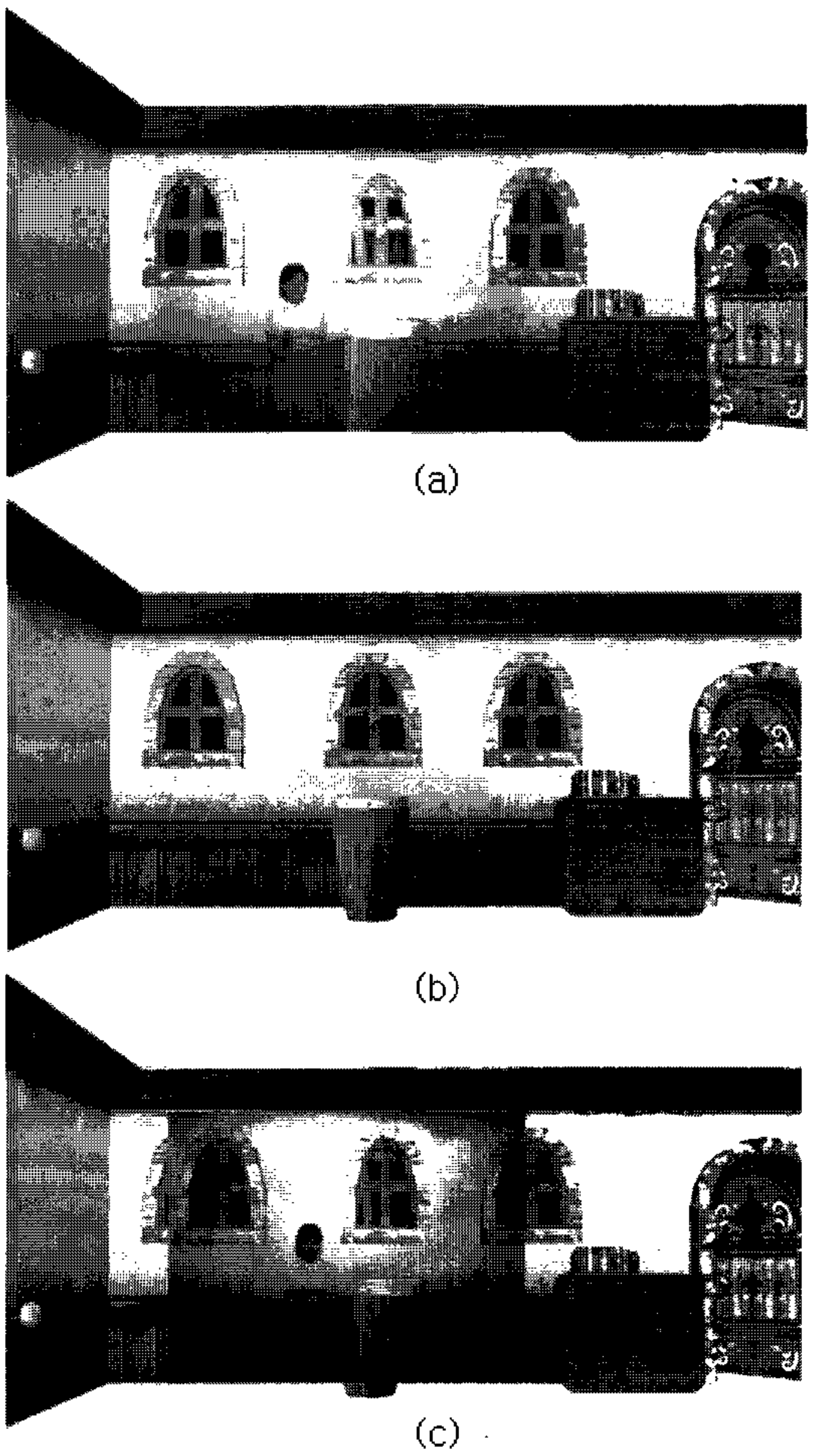


Fig. 3. Registration process with transmitted image (a) after registration process (b) reconstructed environmental map (c) before registration process

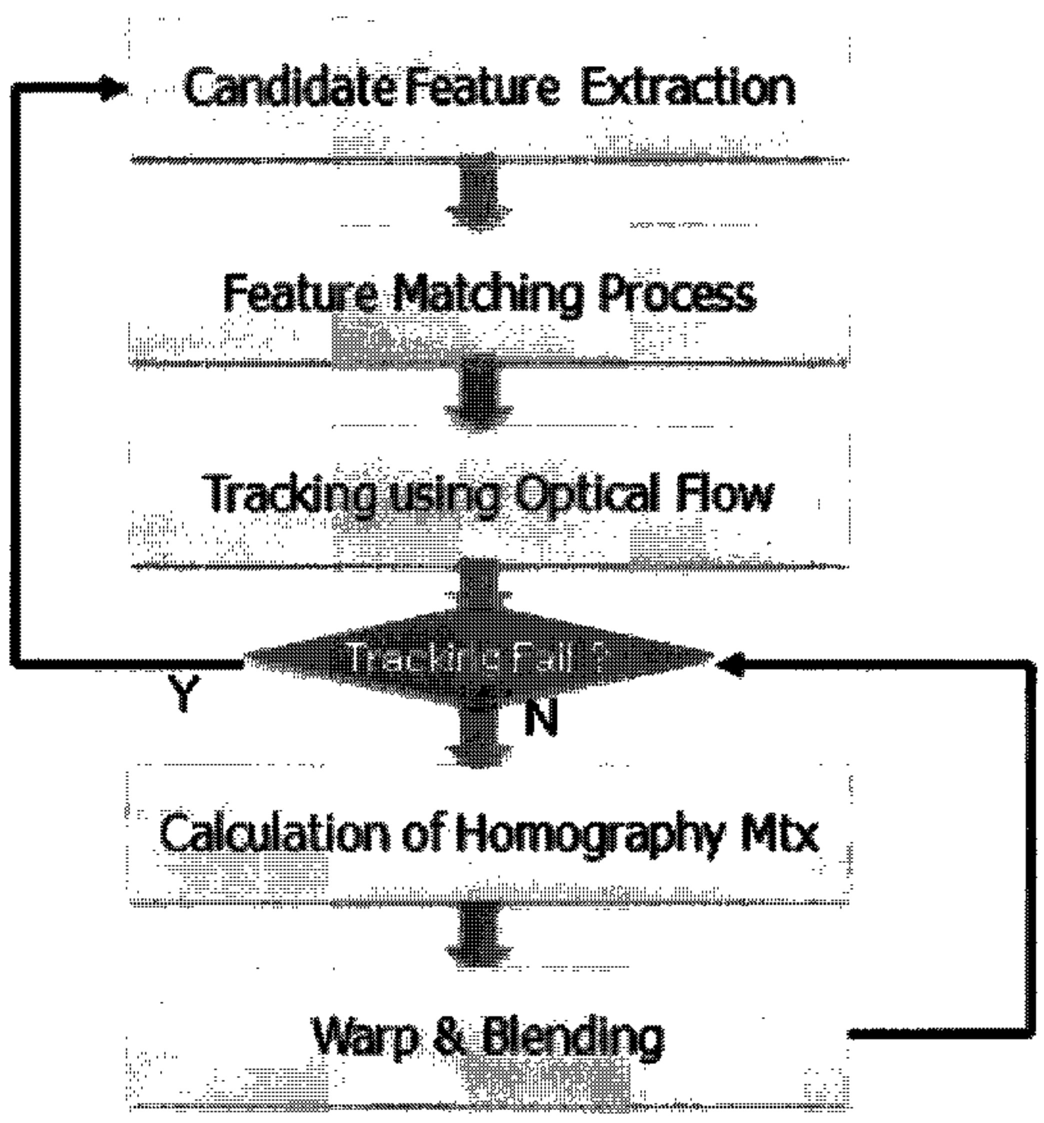


Fig. 4. Flowchart of registration process between transmitted image and reconstructed environmental map.

the background seamlessly. To make a seamless composition, we must extract feature points from both static panoramic image and transmitted image. We apply the Lukas-Kanade feature tracking method to extract candidate feature points from images[2,3]. After extracting feature points, we find matching points and calculate homographies[4]. It is redundant to apply feature extraction and matching process for every frame. After those processes, we apply optical flow method to track the extracted feature points. This process can reduce computation complexity. Finally we can apply warping and blending process with calculated homographies[4]. Fig. 4 shows the flowchart of these steps.

4. System configuration

In Fig. 5, we show the flowchart of our system explained above to display active region stereoscopically. The

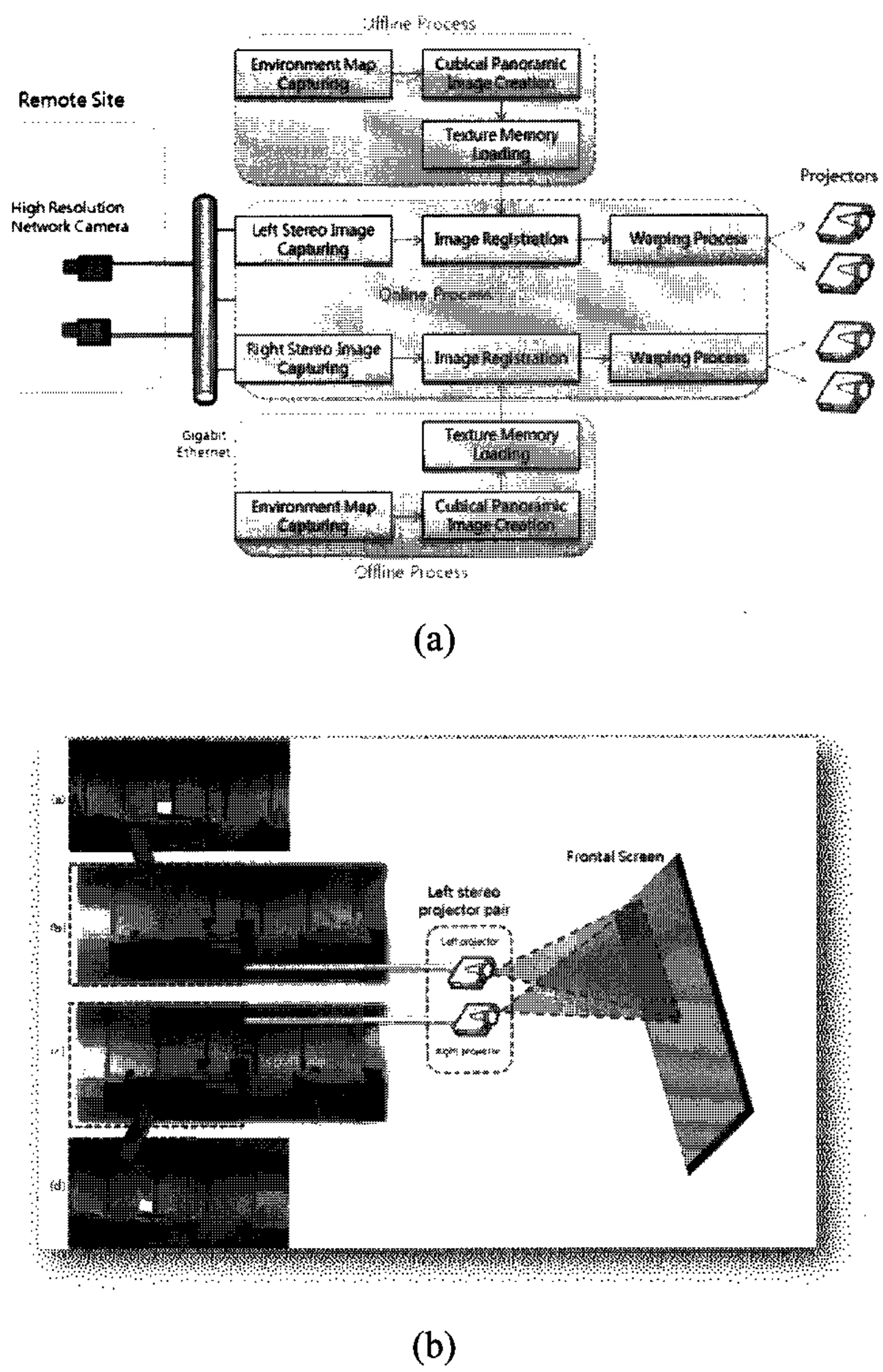
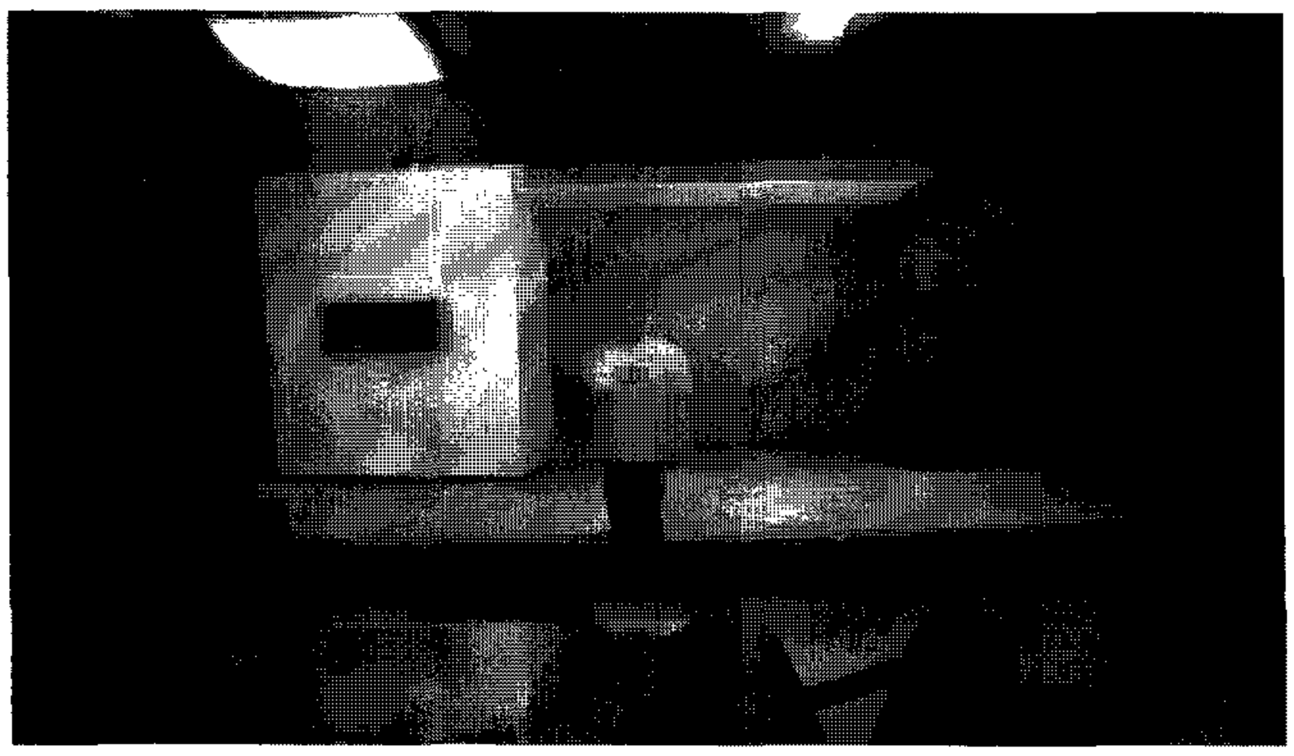


Fig. 5. Whole system configuration for stereo display (a) Block diagram (b) Stereo display example using left stereo project pair

overall system architecture is similar to the system with single remote camera, however, we need two high resolution network cameras to capture the stereo images from remote site and two more projectors and processing computers. In offline process, we build each environmental map for left and right view from remote stereo cameras. After building environmental maps, we register to each map with transmitted frame for each view in real-time. Fig. 6(a) shows the display example of rendering the reconstructed remote world with user. Fig. 6(b) shows the picture taken from the real demonstration of telemeeting between remote user and local user.



(a)

(b)

Fig. 6. Experimental results (a) display example of rendering the reconstructed remote world with user. (b) real demonstration of telemeeting between remote user and local user

5. Conclusions

In this paper, we propose an efficient tiled display system for tangible telemeeting. For tangible telemeeting, we must support immersive environment such as large screen, haptic devices and surround sound system. Among them, we focused on the immersive large display system in this paper. To display high resolution image of remote environment, we must have high resolution image source. Even though we can have high resolution images of remote site with several cameras, the transmission of high resolution images needs high network bandwidth. This can be said to be a critical condition for building a real-time system. For network efficiency and reducing computation complexity, our proposed system only updates the region changed actively by user. For seamless synthesis, GPU based warping and blending technique is applied in our system. More robust feature extraction and matching algorithm is essential to improve the synthesized image quality. For now, we have focused on developing robust and fast feature matching and tracking algorithm, especially under the occlusion state.

References

[1] S. C. Ahn, I. J. Kim, H. G. Kim and S. Ha, Interactive Immersive Display, in Proc. *International Conference on Artificial Reality and Telexistence, TSI workshop,* (2004), p. 63.

[2] C. Tomasi and T. Kanade, Detection and Tracking of Point Features, in *Carnegie Mellon University Technical Report,* (1991).

[3] J. Shi and C. Tomasi, Good Feature to Track, in *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* (1994), p. 593.

[4] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision,* (Cambridge University Press, Cambridge, 2000), ch. 12.

[5] I. J. Kim, S. Y Lee, S. C. Ahn and H. G. Kim, Augmented Telexistence in Smart Space, in *Proc. International Conference on Artificial Reality and Telexistence,* (2005).

[6] H. Y. Shum and R. Szeliski, *International Journal of Computer Vision,* **36**, 101, (2002).