ICAT 2001
December 5-7, Tokyo, JAPAN

Visualizing Focus of Attention in Mixed Reality Community Space

Yuji Sato, Masayuki Takemura, Itaru Kitahara,
Junichi Hoshino, Yuichi Nakamura, and Yuichi Ohta

University of Tsukuba,
Institute of Engineering Mechanics and Systems
Tsukuba, Ibaraki, 305-8573, JAPAN
TEL:+81-298-53-6556 FAX:+81-298-53-5207
http://www.image.esys.tsukuba.ac.jp

Abstract

Mixed Reality Community Space is a space where users can share a mixed reality environment with each other. In this space, users can share the information of real and virtual worlds together. However the mutual attention among the users are often lost because of the HMDs they using. The attention is important for natural conversation discourse in the Mixed Reality Community Space as well as in an ordinary real space.

In this paper, we propose a new technique to support the attention sharing in the Mixed Reality Community Space by visualizing the attention among the users.

Key words: mixed reality, attention visualization, attention sharing,, eye-contact, HMD

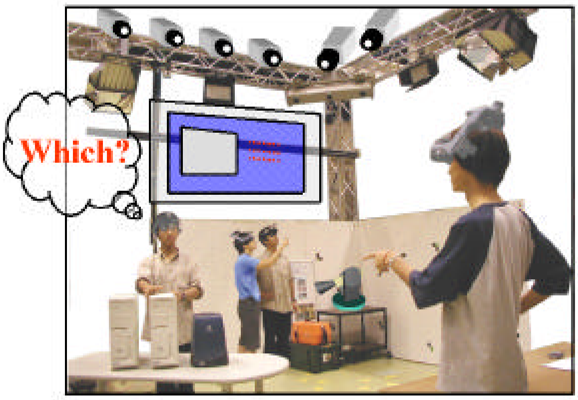

Figure1: Multiple users communicating in MR environment.

1. Introduction

Mixed Reality (MR) is a technology that merges real and virtual worlds and presents the merged world to users. MR technology is useful for many applications because we can experience both of the real and virtual worlds simultaneously by using a head mounted display (HMD) [1][2].

When multiple users are communicating in an MR environment, it is often important to understand which objects are referenced during the conversation. However, when the working space is large, and the referenced objects are far from the users, it is often difficult to recognize them by looking at the pointing gestures made by the users. In addition, the gaze awareness among the users is lost by wearing the HMDs. It plays an important role for recognizing the referenced objects and the user's intension.

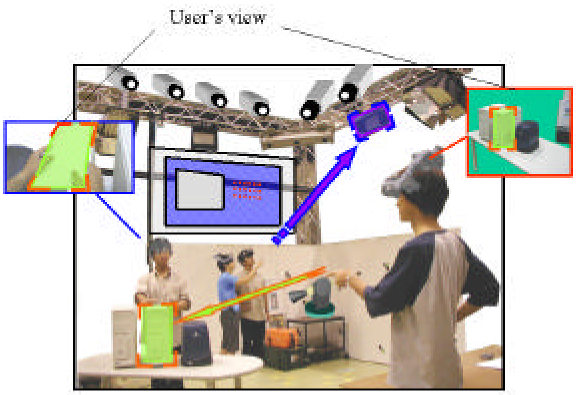

Figure2: Visualizing focus of attention in MR community space.

In this paper, we propose a new MR system, which can support the attention sharing among users by visualizing the referenced objects in the 3D environment. The

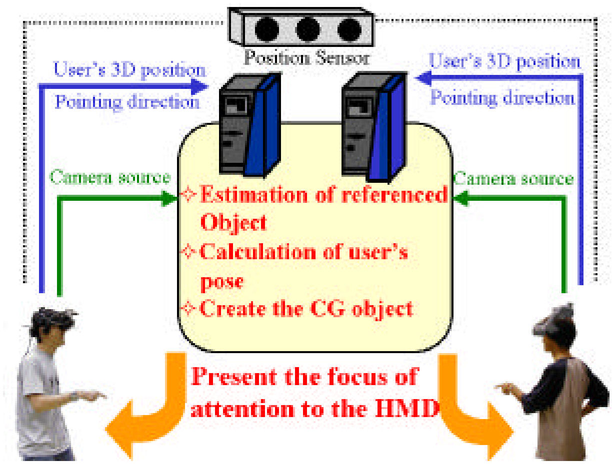

Figure3: Overview of system

Figure5: An example scene.

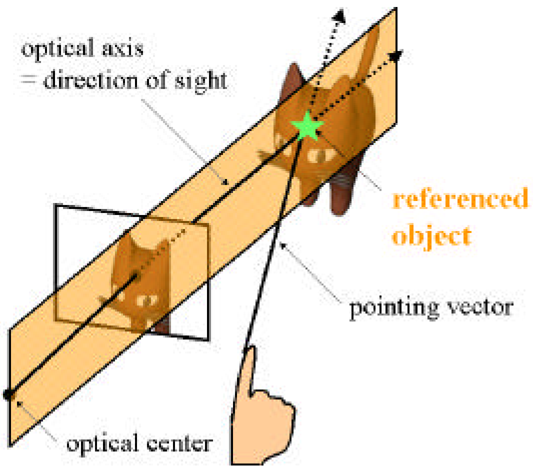

Figure 4: Estimation of referenced object.

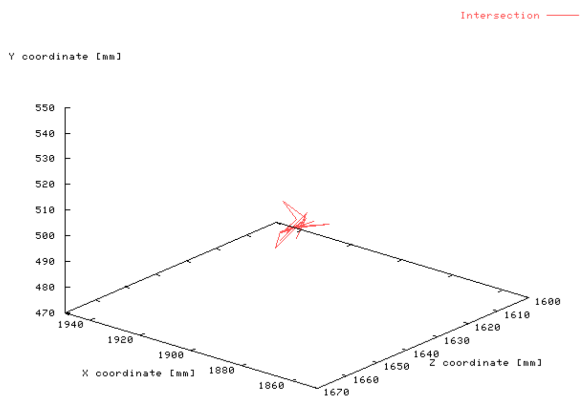

Figure6: Trajectory of the location estimated for the pointing action to the right camera in Figure 5.

system also restores gaze awareness by synthesizing and overlaying the facial image of the user to his head with HMD.

2. Overview of System

Figure 3 shows the overview of our system. We get the 3D position of user and the vector of arm orientation by using a position sensor. In order to estimate the location of a referenced object, we also use the direction of sight of the user. The direction of sight, in this case, means the optical axis of camera attached to the HMD. By calculating the intersection of the direction of sight and the arm orientation, we can get the attention point in the 3D space. Finally, the referenced object is highlighted by overlaying a graphic pointer to the HMD. This function enables users to recognize which object is referenced by the users even in a wide workspace.

3. Estimation of referenced object

In order to estimate the location of object referenced by the users, a virtual plane including the user's direction of sight and the pointing vector determined by the arm orientation are used. Figure 4 shows the relationship between the virtual plane and the pointing vector. The user's direction of sight is determined by the optical axis of camera attached to the HMD.

The pointing vector is determined by a straight line connecting two markers attached to the user's arm. The intersection of the direction of sight and the pointing vector in the 3D environment is estimated as the location of referenced object by the user.

Figure 5 is an example scene and figure 6 represents the trajectory of the estimated location. In this scene, a user pointed out the right camera on the floor.

4. Relationship of coordinate systems

In order to present the estimated location of referenced object to the HMD, we need to acquire the position and orientation of each HMD. The position and orientation are represented by 4x4 matrix.

We define four coordinate systems as follows:.

- World coordinate system W: (Xw, Yw, Zw)

The 3D coordinate system of the MR community space.

- Sensor coordinate system S: (Xs, Ys, Zs)

The 3D coordinate system of the infrared position sensor.

- User coordinate system U: (Xu, Yu, Zu)

The 3D coordinate system defined on an HMD.

- Camera coordinate system C: (Xc, Yc, Zc)

The 3D coordinate system of the camera attached to the HMD.

Figure 7 illustrates the relationship among these four coordinate systems. We need the following transformation between the coordinate systems.

1. The transformation between the sensor and world coordinate systems

This transformation is used to represent the positions of infrared markers attached to the user's arm and HMD in the world coordinate system. The infrared sensor is fixed in the world. Then, we can determine this transformation in advance. We denote it as Ms2w.

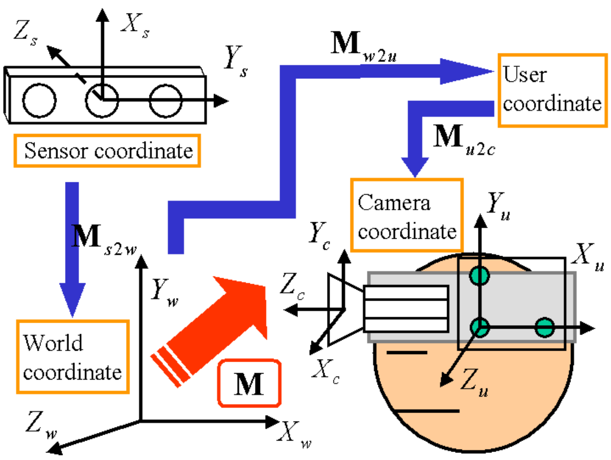

Figure 7: Coordinate systems in MR community space.

2. The transformation between the world and user coordinate systems

This transformation represents the position and orientation of the HMD in the world coordinate system. This can be calculated using three marker positions attached to the HMD[3] as illustrated in figure 8. We denote it as Mw2u.

3. The transformation between the user and camera coordinate systems

This transformation represents the pose of camera attached to the HMD. This can be calculated by multiplying inverse transformation of Mw2u and the initial Mw2c. The initial Mw2c represents the initial position and orientation of HMD. This can be calculated by a camera calibration process. We denote it as Mu2c.

Finally the position and pose of an HMD, M in figure 7, is obtained by multiplying these transformations.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = M_{u2c} M_{w2u} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \qquad (1)$$

The pose of a user changes when he moves. The transformation between the sensor and world coordinate systems is constant because the two coordinate systems are fixed. The transformation between the user and camera coordinate systems is constant because the camera is fixed on the HMD. So the transformation between the world and user coordinate systems changes only by the user motion. We can calculate the pose of

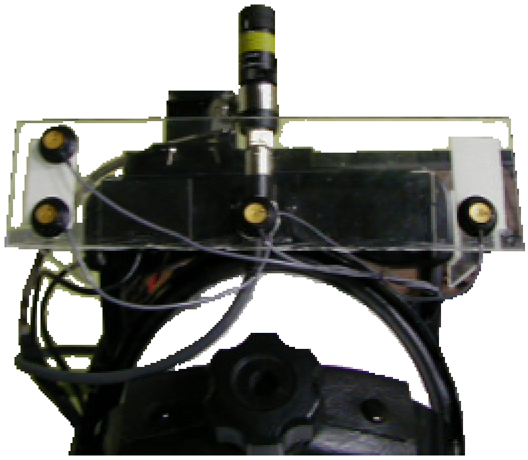

Figure8: Infrared markers on HMD.

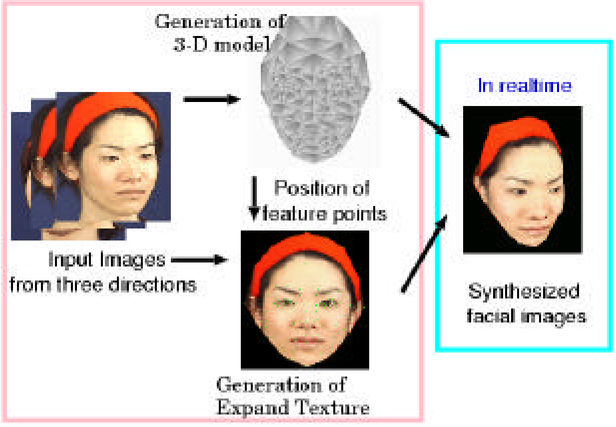

Figure9: Flow of face synthesis.

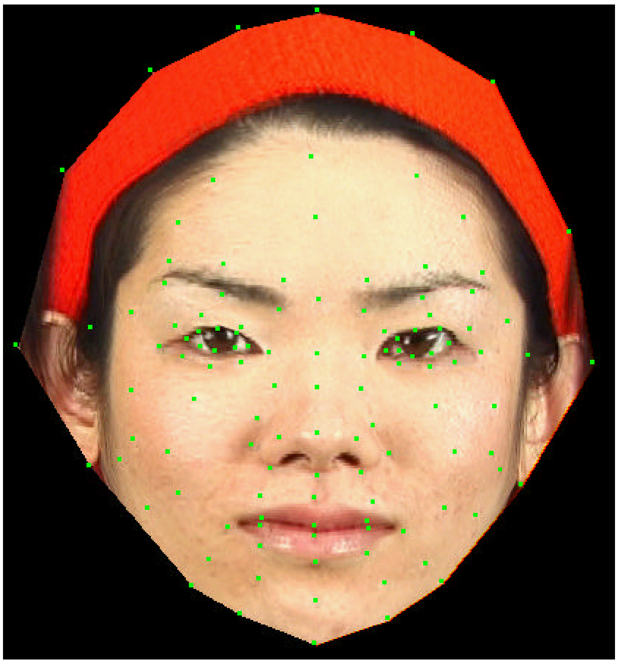

Figure10: Expanded texture.

the camera in real time by measuring the pose of HMD using the infrared sensor.

5. Face synthesis

As shown in Figure 9, we try to synthesize a face by the combination of several face images.

First, we reconstruct a facial 3-D shape with three input facial images. The correspondences of the feature points on the three images are manually given. In this method 3-D facial model, rotation matrix, and components of translation are calculated from the set of 2-D coordinate values.

Next, we synthesize an expanded-texture in order to get facial texture that can cope with arbitrary pose. The

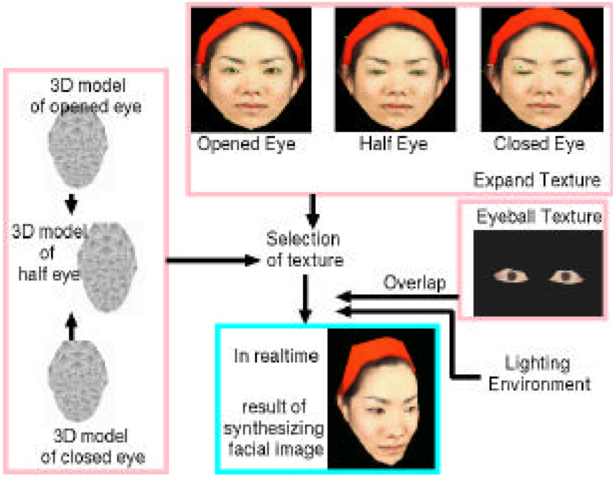

Figure 11: Flow of real time face synthesis with eye expression.

Figure12: Example of synthesized facial images.

Figure13: Synthesized face overlaid to a real head with HMD.

calculation time of texture mapping including blending is the longest part in the real time rendering process of facial image. To reduce the computational time for blending, we generate expanded-texture by blending multiple input images. Keeping the connectedness between the patches defined on the face and representing the expanded-texture with good resolution, the real time face synthesis without texture degradation has been enabled.

Figure 10 shows an example of expanded-texture. The local deformation of the texture on the image plane becomes large when the distance from the central feature points becomes large. However we do not lose textural information because these deformations tend to make the texture patch larger compared to the original one. Geometric deformation of patch shape on the expanded-texture never affects the image quality of the synthesized facial images.

Figure 11 illustrates the flow of face synthesis with eye expression. It works in real time. While the motion of the eyelid is comparatively simple, transition of texture on the eyelid is complex. It is difficult to synthesize the image sequence of moving eyelid by using only the eye-opened images. In order to synthesize realistic eyelid, we use both of the facial images, open eye and closed eye. We can generate a 3D model with an arbitrary state of eyelid by using two 3D models prepared from two sets of facial images with open/close eyes. We also prepare three expanded-textures (open eye, half-open eye, and closed eye). These textures are used to synthesize moving eyelid. An arbitrary view of a face is synthesized by mapping every triangle patch of the 3-D model with the expanded-texture. In addition, expression of eye can be synthesized by synthesizing arbitrary line of sight including moving eyelid. Though the line of sight of human eye comes from a 3-D rotational movement of eyeball, our system represents it by shifting 2-D eyeball texture. Figure 12 shows some examples of the synthesized face images.

In order to overlay the synthesized facial images onto the user's face region in the video image captured by the observer's HMD camera, we need to know the position and orientation of user's head wearing the HMD. This is done by using the infrared markers attached to the HMD. The position and orientation of the head are calculated following the similar process described in the previous

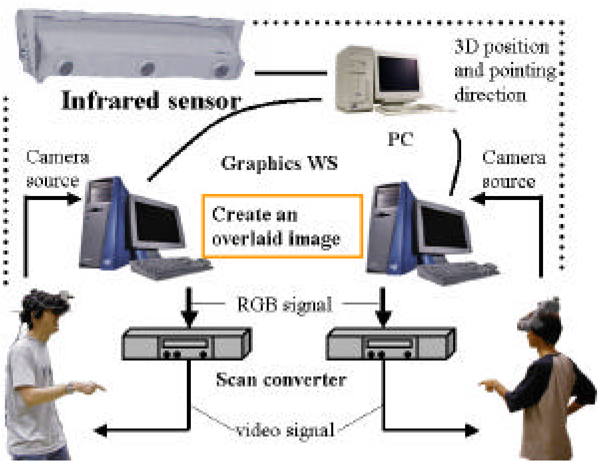

Figure14: Demonstration system configuration


Figure15: Application of MR community space

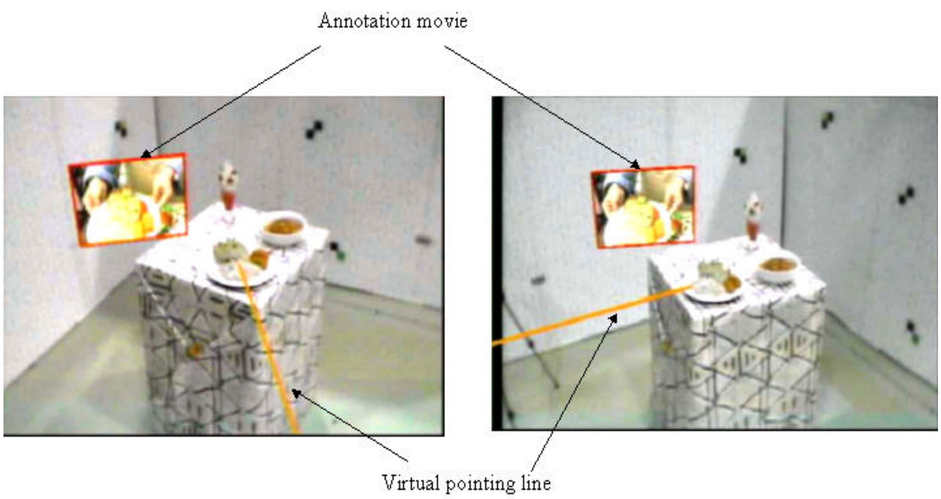

Figure 16: Views presented to user's and partner's HMDs.

section. We can restore the eye contact between the users by overlaying the facial image with eye expression to the region around the real eyes occluded by the HMD [4]. Figure 13 shows a snapshot of the overlay of synthesized face to a real head with HMD.

6. Experimental results

We have made a demonstration system based on a scenario in which two users communicate about exhibited food samples.

Figure 14 illustrates the system configuration used in this demonstration. Using the infrared sensor attached on the HMD, the user's 3D position can be obtained. The pointing direction is detected from the two markers attached to user's arm. The images to be overlaid on the real scene is generated by a graphic PC assigned to each user.

The scenario is as follows. Two people come to a restaurant, and they begin to talk about the exhibited food samples (Figure 15). When one user points out a food, it is highlighted and at the same time an annotation movie tagged to this food is presented to the partner's HMD (Figure 16). The highlight and the annotation movie are also presented to the user's HMD who made the pointing action. Here the annotation movie is stored as the movie database created in advance[5].

Thus the users can communicate with fully understanding his partner's attention by highlighting the referenced object. Two users can also share the attention by presenting the information about the object even in the situation that the users are isolated temporally and/or spatially.

7. Conclusions and future works

In this paper, we proposed a new MR system that can share user's attention by visualizing referenced objects in 3D environment. We also restore gaze awareness by synthesizing the facial image of the user that is hidden by the HMD.

Future works include the development of stabilization method of the user's position and orientation obtained by the infrared sensor. We are also planning to develop the methods to estimate the location of referenced objects in 3D space more precisely. One possible solution is to use user's line of sight. We think that the use's line of sight has rich information to determine the referenced object. We can get this information by using the HMD, shown in Figure 17, with the capability of detecting the line of sight of the user.

Secondly, we try to use a high-resolution fisheye camera that can observe the whole 3-D space for estimating the user's referenced object. In this view, we can extract the candidate objects that may be pointed as referenced object. A candidate object near the intersection of line of sight and the pointing vector by arm is selected as the referenced one.

REFERENCES

1. M.Billinghurst, S.Baldis, E.Miller, and S. Weghorst. Shared Space: collaborative information spaces. HCI international '97, Sep 1997.
2. Zsolt Szalavari, M.Gervautz, A.Fuhmann, and D.Schmalstieg. Augmented reality enabled collaborative work in studiertube. EURO-VR, 1997.
3. Kanbara Masayuki, Takashi Okuma, Haruo Takemura, and Naokazu Yokoya. A stereoscopic video see-through augmented reality system based on real time vision-based registration. IEEE Virtual Reality 2000, March 2000.
4. Masayuki Takemura, Junichi Hoshino, Itaru Kitahara and Yuichi Ohta. Restoration of eye-contact in mixed community space by human image processing. VRSJ The 6th Annual Conference, 2001 (in Japanese).
5. Motoyuki Ozeki, Yuichi Nakamura, Yuichi Ohta. Sharing Attention in Mixed Community Space – Recording Annotation for Objects by Recognizing Human Behaviors, the 6th VRSJ 2001 (in Japanese).

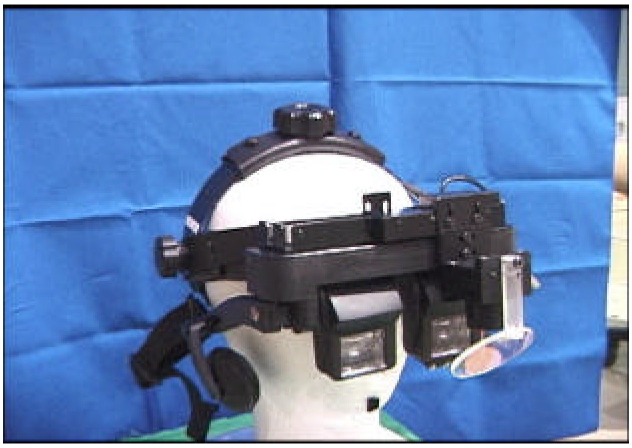

Figure17: HMD with capability to detect the line of sight