8

Collaborative MR Workspace with Shared 3D Vision Based on Stereo Video Transmission

Shengjin Wang , Yaolin Tan , Jun Zhou, Tao Wu, and Wei Lin
Tsinghua University
P.R.China

1. Introduction

Mixed reality (MR) research aims to develop technologies that inputting or mixing the real world objects into computer generated three dimensional (3D) virtual space. The mixed reality, including augmented reality (AR) and augmented virtuality (AV), produces an environment in which the real objects are superimposed on user's view of the virtual environment or the virtual objects are superimposed on user's view of the real environment. Mixed reality has received a great deal of attention as a new method for displaying information or increasing the reality of virtual environments. Many research results have been reported and demonstrated [P. Biermann and B. Jung, 2004 - Shintaro Ono et al., 2005]. Recently, some new assessments have been developed that such new model of virtual systems which are different from traditional form is efficient for exciting user's sense [Yi Cai et al., 1997 - Raphael Grasset, 2005]. However, most of such large-scale assessments still remain at the level of viewing computer graphics (CG)-generated virtual objects. It has revealed that such model of virtual system is no longer functional or fascinate based on virtual reality techniques only. The new, innovative, and mixed reality based approaches are required for operation or exhibition in 21st century. That means the new virtual reality model is focusing on cooperation-centred, real object-based operation. The features of such operation are mainly real or mixed, natural interactive, three dimensional with stereo vision, and collaborative.

Mixed reality environments are defined by Milgram as those in which real world and virtual world objects are presented together on a single display [Milgram P. and Kishino, 1994]. The single user based mixed reality interfaces have been developed for computer aided instruction [S. Feiner et al., 1993], manufacture [Cruz-Neira et al., 1992] and medical visualization [Bajura et al., 1992]. Recently, Seon-Min et al. [Seon-Min et al., 2006] presents a method for merging a live video stream of multiple users into a shared virtual space. These applications have shown that mixed reality interfaces can enable users to interact with the real world in ways never before possible.

Furthermore, the combination of mixed reality and network communication is becoming a more interesting research subject. Although mixed reality techniques have proven enough valuable in single user applications, there has been less research on group collaborative applications. We believe that mixed reality is ideal for collaborative interfaces because it addresses two major issues in three dimensional computer supported collaborative work:

Source: Scene Reconstruction, Pose Estimation and Tracking, Book edited by: Rustam Stolkin,
ISBN 978-3-902613-06-6, pp.530, I-Tech, Vienna, Austria, June 2007

seamlessness and enhancing reality. In this paper we mainly describe a framework of our proposed collaborative system based on stereo vision for a shared mixed reality application and report the development experimental results. This research is sponsored by the national project of demonstrative application of China Next Generation Internet (CNGI2006).

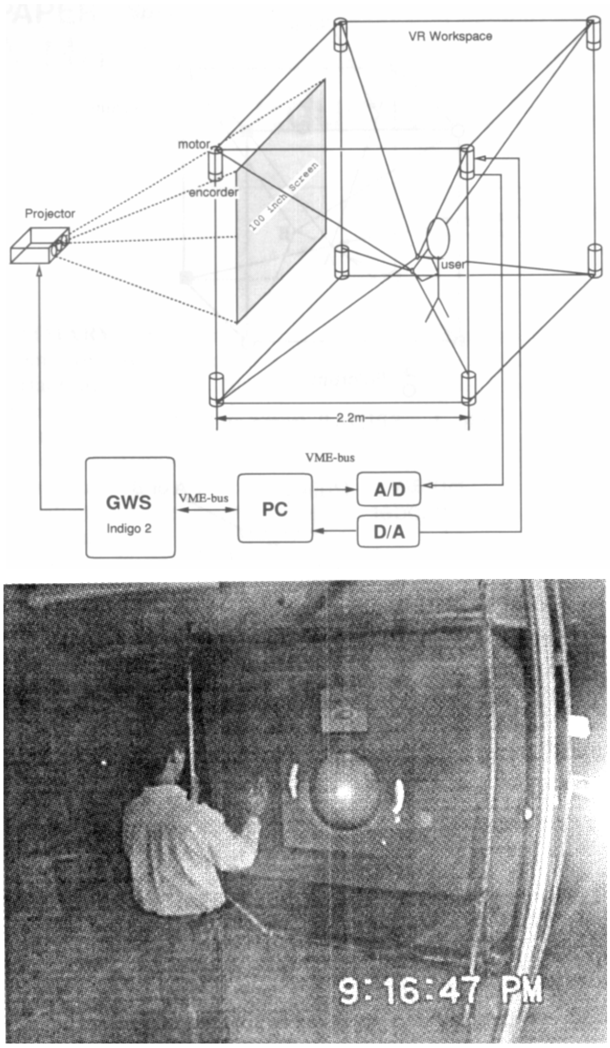


Figure 1. A human-scale direct motion instruction system of virtual reality[Yi Cai et al., 1997]

There are several different approaches for facilitating three dimensional collaborative operations. The most obvious is adding collaborative capability to existing screen-based three dimensional packages. The CAVE Automatic Virtual Environment (CAVE™) -like systems [Cruz-Neira et al., 1992] allow a number of users to view stereoscopic three dimensional images by wearing LCD-shutter glasses. These images are projected on multiple large screen projection walls in the case of the CAVE™. Li-Shu [Li-Shu et al., 1994] developed a workstation based collaborative computer-aided design (CAD) package but users found it difficult to visualize the different viewpoints of the collaborators making communication difficult. Alternative ways include using large parabolic stereo projection screens or holographic optical systems to project a three dimensional virtual image into space. We had developed a human-scale direct motion instruction system for education and training shown in Fig.1 [Yi Cai et al., 1997]. However the system mainly emphasizes the haptic senses by way of virtual objects. Unfortunately in these cases there is still no suitable interface way for multi-users to communicate their view intension each other in three dimensional environment.

Multi-user immersive virtual environments provide an extremely natural medium for three dimensional system. In this case computers provide the same type of collaborative information that people have in face-to-face interactions, such as communication by object manipulation and gesture [Wexelblat, 1993]. Work on the DIVE project [Carlson et al., 1993], GreenSpace [Ishii H. and Miyake, 1991] and other fully immersive multi-participant virtual environments has shown that collaborative work is indeed intuitive in such surroundings. Gesture, voice and graphical information can all be communicated seamlessly between the users.

The mixed reality can superimpose real world objects into the virtual world. This allows the creation of mixed reality that combines the advantages of both virtual environment and seamless cooperation. Information overlay may be used by remote collaborators to annotate the user's view, or may enhance face-to-face conversation by producing shared interactive object models. In this way mixed reality techniques can be used to enhance communication regardless of proximity. There are few examples of multi-user mixed reality systems. Amselen [Amselen, 1995] and Rekimoto [Rekimoto and J. Transvision, 1996] have explored the use of tracked hand held LCD displays in a multi-user environment. Klaus et. al. [Klaus et al., 1995] also use video compositing techniques to superimpose virtual image over a real world view. Ogi et al. [Ogi et al., 2003] take the basis of segmentation by stereo cameras to a further step and generate video avatars. S. Wuemlin et al. generate a 3D video of a user in virtual world [S. Wuemlin et al., 2004]. Vorozcovs et al. present an optical tracking approach for a spatially immersive display [Vorozcovs et al., 2005].

Our collaborative operation system presents a framework to use stereo video and 3D CG model with combination form for showing and indicating a real world object in a shared collaborative virtual workspace. These types of MR interfaces allow multiple users in the different location to see a shared stereo vision simultaneously. This approach is most closely related to that of image-based rendering and efficiently to allow users to collaboratively view and discuss the real world arts and products in stereoscopic types.

The purpose of this research is to develop a platform for group user cooperation with stereo video and shared three dimensional vision. This paper presents a novel framework of collaborative MR workspace with shared three dimensional vision based on stereo video transmission. Our approach combines techniques of stereo video capturing, image/CG fusion and combination, and data transmission for shared stereo vision.

The remainder of this paper is organized as follows. Section 2 introduces the distributed virtual environment. Section 3 discusses the framework of the proposed collaborative operation system. Section 4 describes the development of video based stereo vision system.

Collaborative operation in distributed virtual environment is shown in Section 5, and communication of collaborative workspace is drawn in Section 6.

2. Shared Vision of Collaborative Environment

With the development of mixed reality techniques, a group cooperation system for on-line distributed application, instead of an individual user system, becomes more and more eager. That means a world wide workspace environment via IP network needs to be considered. Such applications likes scientific discussion, product design, indoor virtual tourist, and etc.. In virtual reality (VR) system, it is called distributed virtual environment (DVE). Generally, the distributed virtual environments consist of computer graphics, virtual reality, and distributed servers. In such environments, multiple users can do collaborative operations and view shared interactive scenes. Each participant can indicate object and change status in the scene or change viewpoints to walk-through the virtual space. The scenes are transmitted to all of other participants connecting to the distributed system via network and all of other participants can see the object or scenes change in real time. This framework guarantees that all users can see the same scene representation.

For the early stage of research work on distributed virtual environments, the most of researchers focus their efforts on the use of function simulation, of which the almost of the scenes are generated by computer graphics. However, recent years the development of virtual reality and mixed reality requires the researchers to exploit new technologies satisfying many increasing interesting applications, including:

- Scientific discussion. Scientific discussion based on distributed VR/MR system requires a collaborative scientific visualization environments. Such environments can be regarded as physical extensions of the model based discussion space. In such three dimensional virtual spaces, researchers can represent a data-driven model, view an image being captured from a real world object by a desk top camera, or show a dynamic module test to the participants in real time. The researchers can meet in a shared vision of three dimensional scenes and furthermore they can talk and write within such visual space.
- Product design. Recently number of product designs are entrusted to special designing companies. Whether in the stage of CG-based three dimensional model design or in the stage of prototype module fabrication, the product design should be shown to consigners with frequently discussion. For this application we intend to extend the use of DVE to facilitate computer-supported collaborative working (CSCW) space between two separated physical locations.
- Indoor virtual tour. A new application may come to our daily life soon, called indoor virtual tour. Actually the application does not view video at home alone only. It is a group indoor virtual tour which shares a common vision, may be three dimensional vision, among connected physical locations. With the shared vision each viewer can see the same scenes simultaneously and share his knowledge to the touring fancy community. DVE is a logical prototype of indoor virtual tours and can provide a rich framework for advanced touring applications.

- Arts appreciation. Supposing that you and your friends want to appreciate a cultural relic located in British Museum, without going to spot there. A shared vision of collaborative environment may provide such applications. This means the system has the ability to enable participants to navigate through a cultural relic around and to appreciate it based on video sequences captured by a network video camera. Such shared vision environments hold great promise for a broad range of arts appreciation and so far as to e-commerce applications.
- Demonstration lecture of medical operation. Another interesting application of shared vision is demonstration lecture. Typical one is demonstration lecture of medical operation. A video based, not CG based or image base, scene capture system delivers sequences operation scenes to the physical separated demonstration hall. The participants view and learn from demonstration lecture of medical operation and the teacher can interpreter each operation procedure simultaneously.

The technologies concerned with above applications include, but are certainly not limited to, image/video capturing, CG based modelling for the accurate and realistic real-time representation of scenes, fusion of image and CG modelling, natural user interface, group communications via network for updating the shared consistent scenes by transmitting streamed combination data of CG model, image and audio into the shared vision system.

In order to provide the visual realism for such high demands on the underlying distributed systems and enable DVE to accomplish such applications within highly presence feeling for distributed systems, two essential factors play an important roles in their domain which are listed below.

Vision of real scenes. For shared visual spaces of DVE, especially those used for existed objects, they are often required to provide lifelike scenes not only the CG generated model. The desire for visual realism has driven different visual medias, for example, the use of arts appreciation based on a group still images which are captured from objects or natural landscape around and well arranged in visual spaces is an typical instance. Furthermore, the scenes generated from CG based model have already no any fascination for those of real arts appreciators and the pioneered information techniques should take up the mission for exploiting a novel medium in the shared space. To realize such DVE system, several of the media data need to be applied to three dimensional visions, including image, video, audio, and CG models, in any scene updates. Thus, the lifelike scenes of vision can be obtained. Especially, in this occasion, the effects of data fusion on illumination, position, and scaling principally determine the scene presence.

Real time communications. To support a group of users simultaneously with large shared scenes, audio, and operating commands in DVE causes a heavy load of data transmission. Therefore two main problems cannot be neglected, real time response (no delay) and media data synchronism, since either of them may obviously affect the specific of the shared vision systems. To guarantee the update of multiple viewers of a particular scene at the right moment, for example the operation happened or object moved, requires that such updates should be propagated quickly. However, the scene changes from any one viewer side usually cause quantity data updates that are sent to the network for data transmission. Several techniques are used to minimize the transmission amount, including the use of media data combinations, image/video compressions, operation detection, and simplex data transmissions in particular updates to reduce actual information traffic. An efficient approach to reduce the traffic is to partition updates to related objects or users of a scene in general group communications, despite of P2P or multicast types.

3. System Framework

In this section, we will briefly describe our system framework, device setup for stereo video, and shared collaborative MR environment. The main process in our system is also explained here, including stereo video capture, operation detection, image/CG model combination, and media transmission for stereo video streams of group users.

3.1 Stereo Video and Shared Mixed Reality Cooperation

Recent developed VR and MR system do provide an alternative medium that allows people to share the same object in communication space. However, most of these system are within CG model based workspace. When people talk to one another on discussing an object, the object is usually the CG-generated model, shape-like but no sense of reality, and with insipid texture status, let alone the complex CG modeling of the object. Actually, people do likely appreciate real world arts and products, as well as CG models in a shared common three dimensional environments even by applying a commentator instructions. To construct a real scene vision system, the scenes of real world and real world objects need to be obtained by image/video capture devices and inputted into the virtual environment.

As mentioned above, our research is the part of the project of exemplary application of China Next Generation Internet (CNGI). This research project is aiming at a shared collaborative environment based on high performance video transmission for the increasing demands on developed VR and MR application, including man-machine interfaces, IP network communication, real world stereo vision, and collaborative operation.

To achieve the goal of developing a platform for stereo vision of real scenes, it needs to develop a stereo video capture deveice with two cameras. The device can capture image sequences in dual video channels. For doing fusion of images and CG models, it needs to generate an assigned CG model and further more to determine the corresponding position within right and left images; then using compression algrithm to form video streams again for network transmission. Also, each user's operation in the connected group should be detected for scenes updates. Finally, in the physical receiver side of the DVE system, decompression process is applied and a stereo display device is used for stereo vision which is active and real based on video streams. The parts above are the principal descriptions of our DVE concepts, which focuses on stereo video-based collaborative environment.

In our system there are several subjects need to be solved. Firstly, stereo video based vision system needs to be developed, including stereo video cameras (see Fig. 5) and left and right channel image sequences processing, media fusion, data compressing, and stereo video display. Secondly, operation interface with model based vision indicator is required for collaborative workspace. Thirdly, a key subject, media data transmission via network either for IPv4 or IPv6 environment needs to be established. For each part we will explain detail in the following.

3.2 System Structure

Now we briefly describe the structure of the system. Though in this paper we focus on a P2P application platform it is easily to extend it to server-client type. Figure 2 shows the flow diagram of our video based collaborative operation system. The system mainly consists of four functional modules which can be described as side of user A (as teacher), side of user B (as student), dotted line block 1, and dotted line block 2 in Fig. 2. Actually, the fact that we regard as user A as a teacher and user B as a student is just for system flow explaining and they certainly can exchange their status freely in the system. In this P2P collaborative platform user A and user B locate separately in physical location and are connected with IP network. The system may contain more users with user C, user D, and etc. by extension. Here, the side of user A is assigned as a teacher side, in where the real object is captured, an indicator is operated, and the combined image/CG sequences are transmitted to the side of user B. User B is assigned as a student who can see stereo video scenes of the interested object and the generated indicator which tracks the surface of the real object.

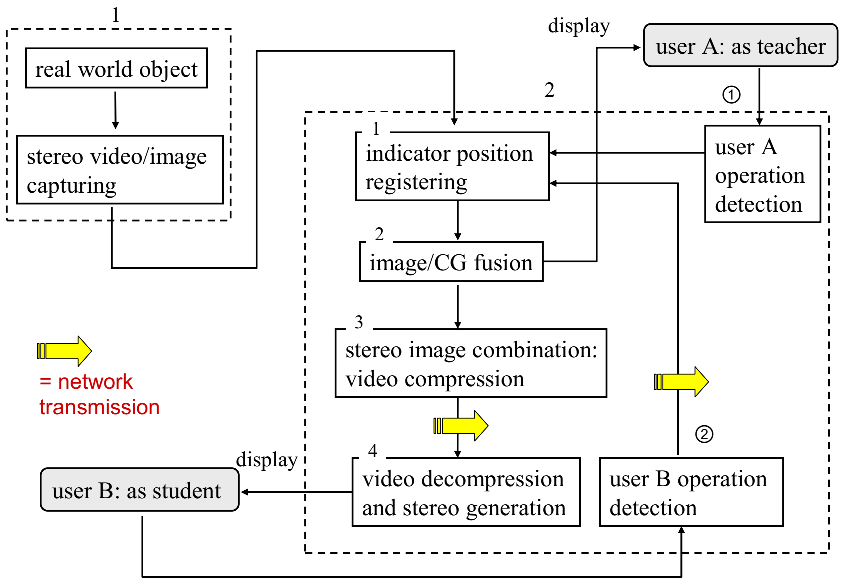


Figure 2. Flow diagram of MR based collaborative operation system

Meanwhile, user B can also operate the indicator which is already displayed within the scenes. The system limits that only one operation is permitted at the same moment, either user A or user B. That means when user A operates the indicator the operation of user B is locked and contrariwise, user A side is the same status when user B do operation since these two possess identity status. This guarantees the identical vision of the system. Here, in side of user B, the system does not transmit video streams to side of user A, in where the side of user B is just a receiver for visual media. In side of user B, the system only transmits audio and operation information to side of user A. This mechanism can reduce the transmission complexity of the collaborative operation system. The procedure of scene updates according to the operation of user B will be illustrated in the following description.

The dotted line frame 1 in Fig. 2 describes the module of stereo video/image capturing. The output of the module block is stereo video sequences of real object or real world scene.

Moreover the capture device consists of two cameras for capturing two channel image sequences which is shown in Fig. 5. Figure 3 is the structures of two channel image capture and process diagram in outgoing system. The central processing module is based on DirectShow® framework where handles capturing image sequences in buffer, three dimensional indicator model generation, indicator position registration, and stereo video stream combination. The corresponding outputs of the central processing module are then delivered to stereo display and network transmission modules.

The dotted line frame 2 in Fig. 2 describes the main function of the video based collaborative environment. In order to reduce the complexity of computation in this paper only indicator model is used and is regarded as a CG model for image/CG model fusion. By that as it may, it is easily to add other CG models if there are demands. The dotted line frame 2 contains the four modules of processing unit, besides user operation detection modules being described below:

The first module (block 1) processes indicator position registration. When two images are read from image buffer captured from two cameras, the right indicator position is determined at once by, for example, user A operation detection which is shown by circle 1 in Fig. 2. The operation detection can be mouse movement distance checking (in this paper) or other natural interface detection by some other special devices. Next, in order to get the left indicator position image registration technique is used to find the suitable matching position in left image.

The second module (block 2) takes image/CG model combination processing. When position of the indicator in left image is found two CG models within right and left two images will be generated, respectively. The process includes generation of right eye indicator CG model and left eye CG model, shading with illumination, and image /CG model fusion of stereo image based on the determined indicator positions. As a result, two image/CG model fusion images are obtained.

The third module (block 3) does stereo image combination and data compression. Since there are two channel images, right and left eye image, two video streams are necessary to be transmitted. This transmission pattern will cause communication complexity and encoding/decoding loads. To simplify system transmission we try to combine two images sequences into one combined image sequence which has the same size in width and double size in height comparing to the original single image. After that, the compression algorithm of mpeg4 is used to execute data compression for the combined image sequences to obtain one video stream. For audio communication mp3 is used for audio compression.

The fourth module (block 4) accomplishes video stream decompression and stereo image generation. By decompressing the video streams received via IP network each combined image, having the same size in width and double size in height of the original one, is obtained sequentially. Followed up with decompression each combined image is then parted into right eye image and left eye image, respectively. Moreover the indicator modelled by CG is simultaneously shown in the vision scenes of user B. Though the description above is based on the operation in side of user A, the procedure is the same when system detects user B operation which is shown by circle 2 in Fig. 2. The difference is only that the sending engine in side of user B will send user B's operation information to the receiving engine in side of user A via network transmission.

The collaborative workspace works via network environment and adapts to IPv4 connections. To establish the system connection between user A and user B the receiving engine in side of user B needs to start up firstly. When receiving engine runs, the side of user B is in waiting call status. When user B's IP address is known by user A, the sending engine in side of user A then calls user B according to the IP address to establish the P2P connection. We stress that though the network connection of the collaborative workspace above adapts IPv4 environment it also works within IPv6 protocols, just needs to change the protocol module and setting interface of sending engine.

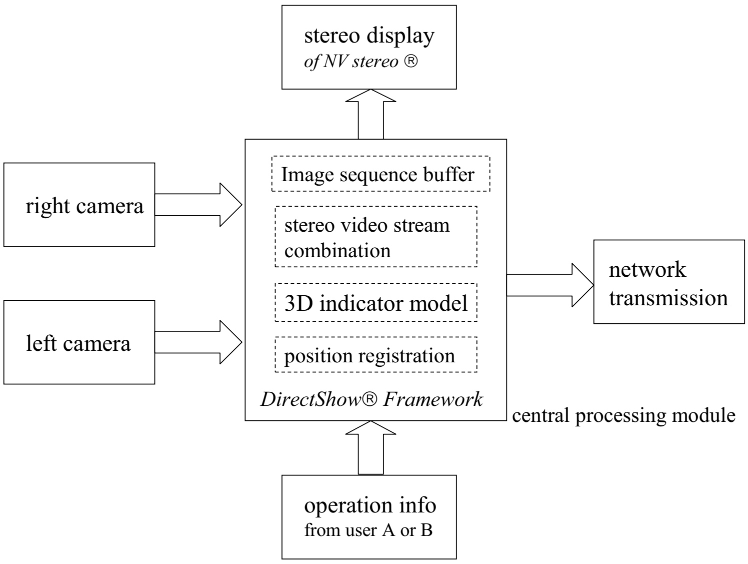


Figure 3. Capture of two channels and process diagram in outgoing system

4. Video Based Stereo Vision

Since the collaborative operation system is a stereo video based MR system we need to develop a video based stereo vision framework, including stereo video capture device, stereo video stream transmission, and stereo video display installation, which enable users view stereo scenes at their separated locations. In this section we mainly describe the component parts of stereo video capture device, stereo video image combination, and stereo video display installation.

4.1 Stereo Video Capture

The sketch of the stereo video capture system is shown in Fig. 4. The stereo video capture device consists of two video cameras, where has 6cm-7cm interval between their optical axises. Figure 5 shows one of our develped stereo camera device. In order to simplify the stereo capture device the usb interface cameras are used in this paper. Each usb camera has high performance in its capture features. It can get 30frame/sec within 640*480 pixels resulution.

The capture process is based on DirectShow® framework shown in Fig. 3. DirectShow® is an important part of DirectX®, which is provided by Microsoft® and is a high performance API to develop graphics, stream media, and audio applications on Windows® platform. The right and left eye images coming from right and left camera of the stereo video device are captured into image buffer sequentially under DirectShow® capture framework. Then the image sequences are delivered to fusion/combination module for the processing of image/CG model fusion, data compression and local side stereo diaplay.

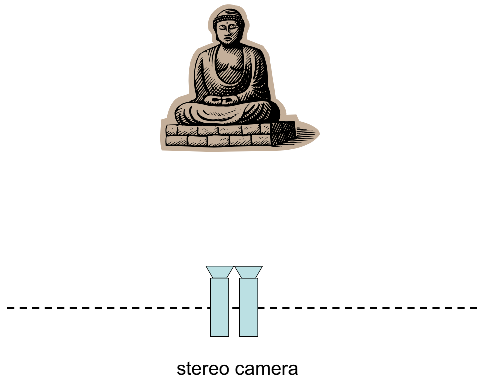


Figure 4. Capture system of video based stereo vision in the collaborative system

Figure 5. The developed stereo camera for capturing image sequences of stereo video

4.2 Display of Stereo Vision

With regard to the stereo display in local side the fused right eye image and left eye image generated respectively by the fusion of image/CG model are directly delivered to 3D stereo display buffer of NV stereo®. Meanwhile, in order to obtain a smooth display for video based stereo vision for far apart side in a DVE system we propose a combination approach to handle the stereo image sequence, named as stereo video stream combination module (see Fig. 2 or Fig. 3). Fig. 6 shows the proposed combination approach, of where image 1 and image 2 is assigned as the fused right eye and left eye image, respectively. To piece the two images, 1 and 2, together in up and down relation a combined image shown in Fig. 6 right side can be obtained. By doing data compression on such image sequence a mpeg4 formatted video streams are generated and then transmitted to the other users via IP network transmission.

In the receiving side of DVE system the video streams received via IP network are decompressed and each combined image, having the same size in width and double size in height of the original one, is obtained sequentially. Then each combined image in the image sequence is then split into the two original image 1 and 2 and afterwards delivered to 3D stereo display buffer of NV stereo®, respectively. Figure 7 shows an example display of stereo vision of the collaborative operation system.

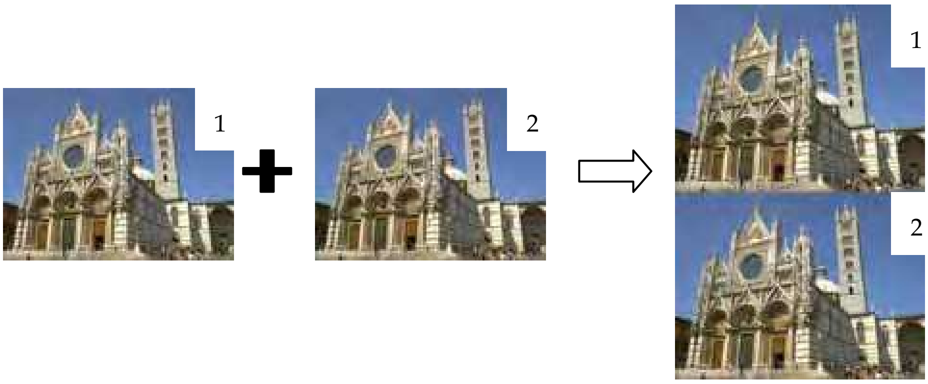


Figure 6. The combination of right and left images and the appearance of combined image

Figure 7. The display of stereo vision with fused (image/CG model) right and left eye images

In our system an active infrared control stereo display device is used for stereo vision in collaborative workspace. The stereo display device consists of an infrared control unit and pairs of liquid crystal eyeglasses, which are shown in Fig. 8. The liquid crystal eyeglasses have an optical shutter, which controls the right and left optical lens opens and closes alternatively in an appropriate frequency. Usually the optical shutter frequency is concerned with field frequency of display device. According to human being's visual response in order to avoid the feeling of scintillation, the frequency of each optical shutter should be more than 30Hz. Since the infrared control unit needs to control two optical shutters open and close for right and left eye, the field frequency of display device should at least be more than 60Hz. In our stereo vision system the field frequency of display device is set to 100Hz. By this means when users wear the liquid crystal eyeglasses and view screen in front of display device the fascinating stereo scenes then really come to their eyes.

Figure 8. An active infrared control stereo display device and a pair of liquid crystal eyeglasses

5. Collaborative Operation of DVE

As mentioned above in this paper, to be an example of fusion application, a 3D indicator is regarded as a 3D CG model for accomplishing the fusion of image and CG model. The other CG models, if it is necessary, can be easily generated and fused in the same way. In this section, we mainly describe the capture of 3D indicator motion, indicator generation and image/CG model fusion, and collaborative operation control.

5.1 Capture of 3D Indicator Motion

The operation interface of our collaborative operation system provides co-located users to interact with shared virtual space while viewing the video based inputted real world objects simultaneously. There are two proposals being presented for collaborative operation in DVE system, complex one with natural interface and smiple one with mouse operation. Though both of the two methods will be mentioned in the following, in order to reduce the sytem complexity and costs the mouse operation method is employed in our collaborative operation sysetm. Therefor, when people construct such collaborative operation system they can select ones between these two interface methods acording to their purpose, space, and research outlay.

A natural interface framework, within camera based detection and tracking of indicator in three dimensional space, is shown in Fig. 9. The configuration of the prototype of natural indicating interface consists of two video based detecting cameras for indicator position detection. The upper camera in Fig. 9 detects the movement of the indicator in a height-based horizontal plane and the front-upper camera detects the shifts information in height for indicator operation, respectively. Using the two detecting cameras, the position of indicator in three dimensional space can be obtained. The indicator being used can be a lightened spot device or a specially colour markered marker and user uses it to indicate the position on the video based stereo object. If there needs to detect user's viewpoint for follow-up scene changing, we prefer to employ an approach of human face detection and tracking rather than body detection and tracking. Both of face detection and body detection have ripe and efficient algorithms developed in motion capture resreach field which can be employed for natural man-machine interface of such collaborative operation systems.

As illustrated above for simplifying the scale of the collaborative operation system a 2D mouse operation style is used in our operation interface. Since the optical axises of the stereo cameras are generally set to pass through the shape center of a real world object, the relations of 3D indicator movement and mouse operation can be assigned as follows for simulation of 3D movement:

- mouse movement in desk plane simulates in a height-based horizontal plane movement of the indicator in three dimensional space.
- The size of indicator scales smaller and larger corresponding to the movement of going in and out in the horizontal plane.
- The movement of mouse contact roller controls up and down movement of the indicator in height in three dimensional space.
- The indicator maintains the size of its shape when user operates mouse contact roller alone.

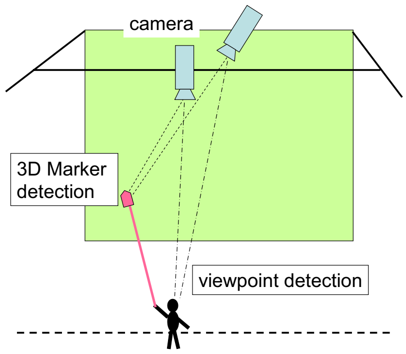


Figure 9. Camera based detection and tracking of indicator in three dimensional space

5.2 Indicator Generation and Image/CG model Fusion

The 3D indicator is employed to indicate a 3D virtual object in three dimensional space. The way of 3D indicator generation is to overlay 2D geometric figures of indicator into right and left eye images, respectively. There is a certain interval between the two geometric figures of indicator, of where the interval is the same as the displacement of 3D object in right and left eye images in screen. More exactly to say the displacement of the 3D object is the distance between the same point on the object surface in right and left eye images. This guarantees an effect that the 3D indicator adjoins closely to the surface of the 3D object. In order to obtain a ideal visual effect the 3D indicator is better to be a solid geometry possessing volume sense and shading effect.

Being with a part of collaborative operation system we adopt the method of addition class and DLL to realize 3D indicator, in which C3DCursor class realizes the modelling and rendering and a Match.DLL we developed realizes the registration (sometimes being regarded as matching) of the corresponding point (actually a block area) in right and left eye images. Using block area to execute the registration rather than a point is for avoiding noise disturbing. Moreover, in order to update the registration algorithm easily a DLL driven type is used in our program. There are two registration algorithm we suggest which are shown in formula (1) and (2), one is based on block histogram and another is based on block vector of grey values.

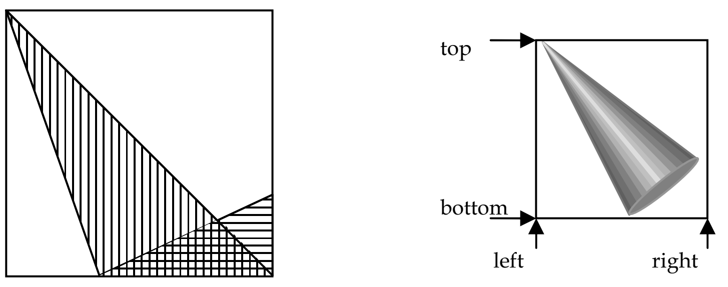
For formula (1) [Swain M and Ballard D, 1991], the colour space of each [0, 255] are quantified into feature colour V=v×v×v. The size of block area is set to level L (L is set to 3) for calculating block histogram , $\boldsymbol{H}$={$H(l, i)$| l=0, …, L-1, i=0, …V-1}. Here v is set to 16. Therefore the similarity value of S varies in [0,1], where 1 means the feature of two blocks is total same and 0 means no similar pixel values in the two registration blocks. $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ indicate the block histogram corresponding to right eye and left eye image, respectively.

$$S = \{\sum_{l=0}^{L-1}\sum_{i=0}^{V-1}\min[\boldsymbol{H}_1(l,i), \boldsymbol{H}_2(l,i)]\}/L \quad (1)$$

The normalized correlation function C shown in formula (2) [David A and Jean Ponce, 2002] is vector of grey value based computation, easier than the algorithm above, where d is a shift distance of a block in left eye image, and w and w′ are the vector formed by scanning grey value of the corresponding blocks sequentially in right eye and left eye image, respectively. Moreover $\overline{w}$ and $\overline{w}'$ are their averages.

$$C(d) = \frac{1}{|\boldsymbol{w}-\overline{\boldsymbol{w}}|}\frac{1}{|\boldsymbol{w}'-\overline{\boldsymbol{w}}'|}[(\boldsymbol{w}-\overline{\boldsymbol{w}})\bullet(\boldsymbol{w}'-\overline{\boldsymbol{w}}')] \quad (2)$$

There are several approaches to accomplishing a 3Dindicator modelling and rendering. Though OpenGL® or Direct3D® can easily be used to construct more complex 3D indicator and more real illumination and rendering can be obtained, the construction of modelling system is rather complicated and moreover, as for our experience OpenGL® and Direct3D® may occasionally emerge conflict with some modules in main program and cause an unstable system. For this reason in our system we adopt a way that drawing geometric figure directly by assigning values to an image matrix to generate a 3D indicator. An example of generated 3D indicator model is shown in Fig. 10. Figure 11 shows a rendering 3D indicator and the fusion of image and the rendering indicator CG model in the collaborative operation system.



(a) (b)

Figure 10. An example of generated 3D indicator model

After four vertexes of the square in Fig. 10(a) are defined the shape of a circular cone is constructed by following loop statement：

$$\text{shape a}=\{(x,y) \mid bottom<y<top,\ left+top/3-y/3<x<right+bottom-y\},$$
$$\text{shape b}=\{(x,y) \mid bottom<y<top,\ 3*(y-bottom)+left+(right-left)/3<x<right\}, \qquad (3)$$

where x and y are coordinates in the square, shape a and shape b are described by vertical line and horizontal line, respectively. Then circular cone can be get from a-b which is shown in Fig. 10(a).

The illumination effect can be simulated by following function of grey values：

$$f(x,y) = 255-K^* \mid A \mid \qquad (4)$$

where

$$A=(left+top/3-y/3+right+bottom-y)/2-x \qquad (5)$$

controls the variance of grey scale of each row pixels, and

$$K=360.0/(right+bottom-y-left-top/3+y/3) \qquad (6)$$

controls the variance of two extremity sizes of circular cone. After assigning the pixel values according to the algorithm above, the circular cone is then possess a stereoscopic visual effect which is shown in Fig. 10(b).

5.3 Collaborative Operation Control on 3D Indicator

For the collaborative operation system the operation control among group users separated in physical location is also a key subject that cannot be ignored. In this paper we adopt two techniques to execute the control of collaborative operation, priority operation and simplex combined image transmission. The priority operation means if one user moves mouse to shift the indicator position the indicator changes in the stereo scenes being caused by operations of other users connected via network are restricted till that user releases the priority by stopping his operation.

Figure 11. The generated indicator and the fusion of image and indicator CG model

The simplex combined image transmission is a mechanism to reduce the system complexity. By employing this mechanism the all mouse move information, despite of user B's or user C's if there is, will be delivered to side of user A. When their sending engines send their operation information to the receiving engine in the side of user A via network transmission the indicator CG model is then generated and set to the corresponding position for the follow-up manipulation of fusion and combination. This mechanism guarantees that an unique CG model generation engine and stereo video streams transmission engine can support system running which greatly cuts down the complexity of the collaborative operation system.

6. Communication of Collaborative Workspace

In order to construct a shared virtual workspace for users separated in real locations the collaborative operation system needs to preserve a real time communications for enjoying in face-to-face meeting, operation, and arts appreciation. All of the connected users can view and indicate the inputted real world objects in such virtual workspace with no bearing time delay. In this collaborative operation system a transmission module based on either IPv4 or IPv6 stereo video has been accomplished for the formed mpeg4 video and mp3 audio streams transmission.

6.1 Data Transmission via IP Network

To interact with shared virtual workspace the system needs to preserve real time communications to enable users separated in real world enjoying in face-to-face liked meetings, leanings, and arts appreciations with inputting real world object into MR based DVE systems. The media data for communication includes human voice, video based stereo sequences, and indicator position. Figure 12 shows the configuration of the collaborative environment with a shared MR workspace via IP network. Based on the applications of the system, the network for communication in this prototype prefers an Internet based distributed network rather than a special-use distributed network. For this purposes all protocols for command and stream media transmission are Internet Protocols with IPv4 and IPv6. We have tested successfully the data transmission in both IPv4 LAN and over three layers of node point with IPv6 environment. The program below shows part of socket communication sour code. The interface of IPv4/v6 network communication in sending side (user A) is shown in Fig. 13 and 14, respectively. Figure 15 is the interface of IPv4/v6 network communication in receiving side (user B). The right figure shows the waiting call status. Noticing that the receiving engine should be starts before sending engine.

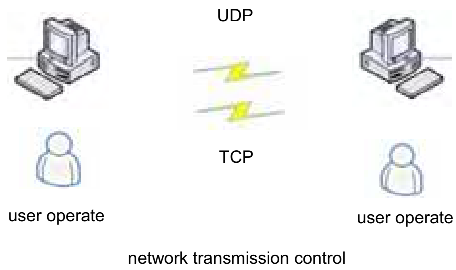


Figure 12. configuration of collaborative environment via IP network

```
mSocket = socket(AF_INET, SOCK_STREAM, IPPROTO_TCP);
if (mSocket != INVALID_SOCKET){
        BOOL sopt = TRUE;
        setsockopt(mSocket, IPPROTO_TCP, TCP_NODELAY,
                (char *)&sopt, sizeof(BOOL));
        setsockopt(mSocket, SOL_SOCKET, SO_DONTLINGER,
                (char *)&sopt, sizeof(BOOL));
        SOCKADDR_IN   saddr;
        memset(&saddr, 0, sizeof(SOCKADDR_IN));
        saddr.sin_addr.S_un.S_addr = htonl(inTargetIP);
        saddr.sin_family        = AF_INET;
        saddr.sin_port          = htons((WORD)inPort);
        if (connect(mSocket, (SOCKADDR *)&saddr, sizeof(SOCKADDR_IN)) != 0) {
                Detach();
                return FALSE;
        }
        mIsConnected = TRUE;
        return TRUE;
}
return FALSE;
```

6.2 Collaborative Workspace Based on Shared Stereo Video

The proposed stereo video based collaborative operation MR system shown in Fig. 16 can be widely used in many applications. Figure 17 and 18 shows the experiment scenes of the system. The video of object scenes are captured into centre computer in side of user A which serves as a server to implement video capturing, video delivering, and operating command exchanging. The computer in both side of user A and B serves the stereo video sequences displaying.

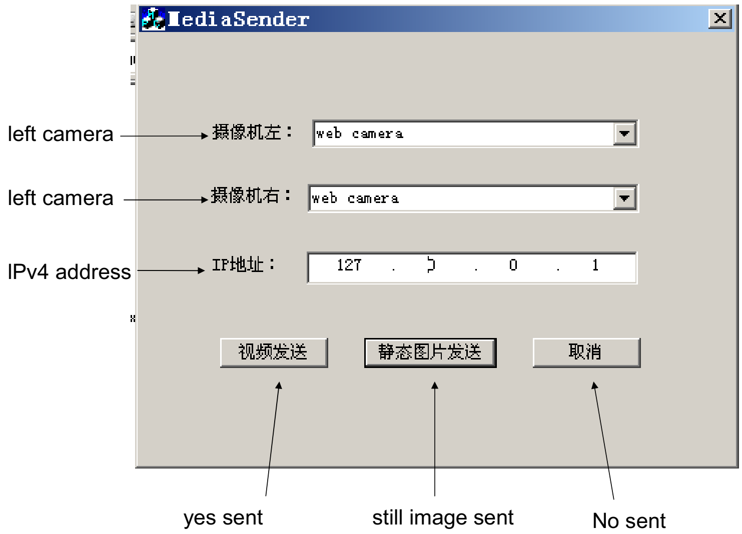


Figure 13. The interface of IPv4 network communication in sending side (user A)

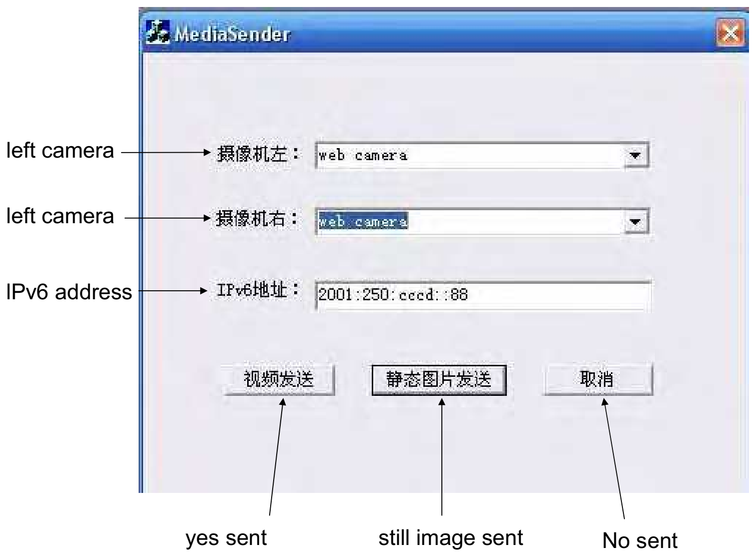


Figure 14. The interface of IPv6 network communication in sending side (user A)

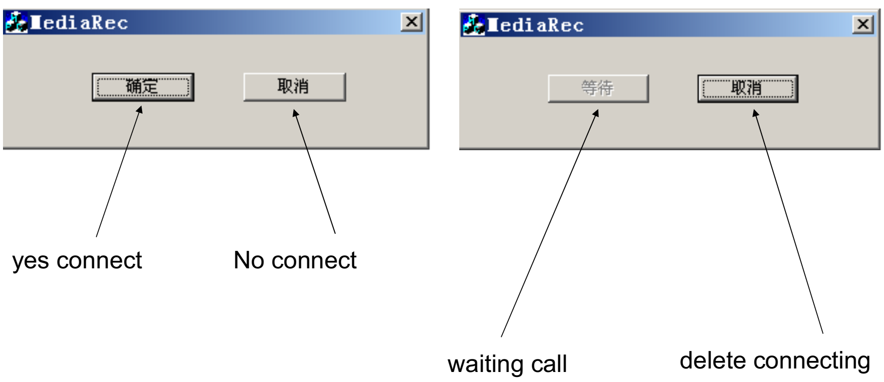


Figure 15. The interface of IPv4/v6 network communication in receiving side (user B). The right figure shows the waiting call status

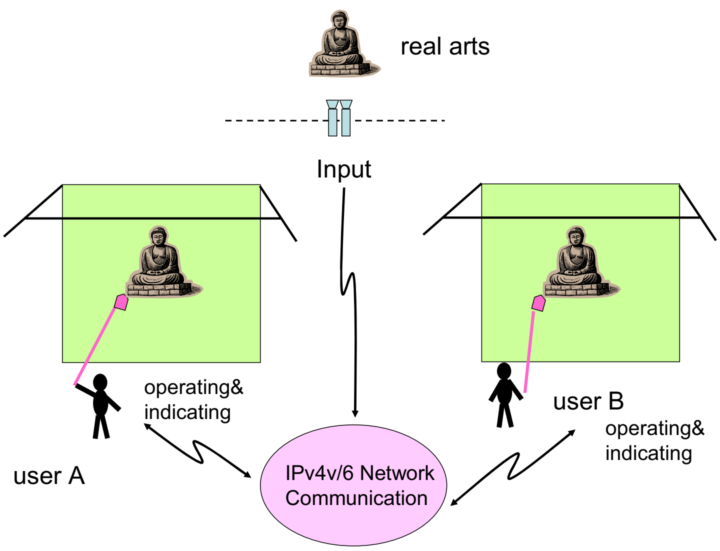


Figure 16. Collaborative operation framework of shared mixed reality via IP network

Figure 17. The experiment scene of collaborative operation MR system

(a) (b)

Figure 18. The experiment scene of collaborative operation MR system. (a) is in side of user A and (b) is in side of user B

7. Conclusion and Future work

We have presented a framework of collaborative MR workspace with shared three dimensional vision based on stereo video transmission, established and experimented for collaborative operation. The system provide a 3D-like operating interface by a CG indicator. Though the system is tested successfully for the data transmission in both IPv4 LAN and over three layers of node point with IPv6 environment there still several tasks need to be accomplished in future work. The one is although we regard a CG indicator as 3D CG models now, for a widely use more 3D CG may need to be added into virtual space and operated in shift, rotate, and spin, as an extended function in future. The second is that though the size of video frame used in now system is 640*480 pixels we plan to test more larger image, such as high resolution image, in the collaborative operation system both for testing its performance and for more wide use. Moreover in future work the performance evaluation for both operation feeling and time response is also an important task which is should be done. Moreover our future work will try to focus on realizing more desired dreams, including multi-stereo vision construction, natural indicating operation, and viewpoint detection, in addition to the ones mentioned above.

Thanks again for sponsoring this research by the national project of demonstrative application of China Next Generation Internet (CNGI).

8. References

Shintaro Ono, Koichi Ogawara, and et.al. A Photo-Realistic Driving Simulation System for Mixed-Reality Traffic Experiment Space. IEEE Intelligent Vehicles Symposium (IV2005), June 2005

S. Feiner, B. Maclntyre, and D. Seligmann. Knowledge-based Augmented Reality. In Commun. of the ACM, volume 36(7), pages 53–62, 1993

T. Ohshima, K. Satoh, H. Yamamoto, and H. Tamura. AR Hockey: A Case Study of Collaborative Augmented Reality. In Proc. VRAIS'98, pages 14–18, 1998

Andreas Pomi and Philipp Slusallek. Interactive Mixed Reality Rendering in a Distributed Ray Tracing Framework. IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR) USA, November, 2004

P. Biermann and B. Jung. Variant Design in Immersive Virtual Reality: A Markup Language for Scalable CSG Parts. Proceedings AMDO-2004, Palma de Mallorca, Spain, Springer (LNCS 3179), pp.123-133, 2004

Raphael Grasset, Julian Looser, Mark Billinghurst. A Step Towards a Multimodal AR Interface: A New Handheld Device for three dimensional Interaction, ISMAR2005, pp.206-207, 2005

A. Orimoto, T. Matsuyama, R. Shizaki, R.Masuda, and M. Masuda. Development of education system for handicapped children. 11th Symposium on Human Interface, pp.209-214, 1995

Y. Hirata and M. Sato. 3-D interface device for virtual work space. Proc. 1992 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.889-896, July 1992

Yi Cai, Shengjin Wang, and Makoto Sato. A Human-Scale Direct Motion Instruction System Device for Education Systems. IEICE Transactions on Information and Systems, vol. E80-D, no.2, pp.212-217, Feb. 1997

Milgram, P., and Kishino, F. A taxonomy of mixed reality visual displays, IEICE Transactions on Information and Systems, Special issue on Networked Reality, Dec. 1994

Cruz-Neira, C., Sandin, D. J., Defanti, T. A., Kentyon, R. V., and Hart, J. C. The CAVE: Audio Visual Experience Automatic Virtual Environment. Communications of the ACM, vol. 35 (6), pp. 65, 1992

Bajura, M., Fuchs, H., Ohbuchi, R. Merging Virtual Objects with the Real World: Seeing Ultrasound Imagery Within the Patient. In Proceedings of SIGGRAPH '92, New York: ACM Press, pp. 203-210, 1992

Li-Shu, Flowers, W. Teledesign: Groupware User Experiments in Three-Dimensional Computer Aided Design. Collaborative Computing, vol. 1(1), pp. 1-14, 1994

Wexelblat, A. The Reality of Cooperation: Virtual Reality and CSCW, in Virtual Reality: Applications and Explorations. Edited by A. Wexelblat. Boston, Academic Publishers, 1993

Carlson, C., and Hagsand, O. (1993) DIVE - A Platform for Multi-User Virtual Environments. Computers and Graphics. Nov/Dec, vol. 17(6), pp. 663-669, 1993

Ishii, H., Miyake, N., Toward an Open WorkSpace: Computer and Video Fusion Approach of TeamWorksation. Communications of the ACM, vol 34, No. 12, pp. 37-50, Dec 1991

Amselen, D. A Window on Shared Virtual Environments. Presence, vol. 4(2), pp. 130-145, 1995

Rekimoto, J. Transvision: A Hand-held Augmented Reality System for Collaborative Design. In Proceeding of Virtual Systems and Multimedia '96 (VSMM '96), Gifu, Japan, 18-20 Sept., 1996

Klaus, A., Kramer, A., Breen, D., Chevalier, P., Crampton, C., Rose, E., Tuceryan, M., Whitaker, R., Greer, D. Distributed Augmented Reality for Collaborative Design Applications. In Proceedings of Eurographics '95. pp. C-03-C-14, September 1995

Seon-Min Rhee, Jiyoung Park, and Myoung-Hee Kim (2006). Multi-user Live Video in a Shared Virtual World for Enhanced Group Cooperation, *Proceedings of International conference Edutainment 2006*, pp.1131-1140, LNCS 3942, Hangzhou China, April 2006, Springer, New York

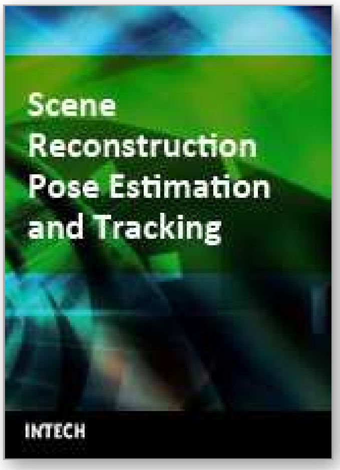
T. Ogi, T. Yamada, Y. Hattori, Y. Kurita, and M. Hirose, Usage of video avatar technology for immersive communication. In *Proc. First Intl Workshop on Language Understanding and Agents for Real World Interaction 2003*, 2003.

S. Wuemlin, E. Lamboray, and M. Gross. 3D video fragments: Dynamic point samples for real time free viewpoint video. *Computers and Graphics*, 28(1):3-14, 2004.

A. Vorozcovs, A. Hogue, and W. Stuerzlinger. The hedgehog: a novel optical tracking method for spatially immersive displays. In *Proc. IEEE VR 2005*. IEEE Computer Society press, 2005.

Swain M J, and Ballard D H. Color indexing. *International Journal of Computer Vision*, 1991, 7: 11-32.

David A Forsyth and Jean Ponce, Computer Vision: A Modern Approach , *Prentice Hall; US Ed edition (August 14, 2002)*, English ISBN-10: 0130851981 ISBN-13: 978-0130851987



ISBN 978-3-902613-06-6
Hard cover, 530 pages
Publisher I-Tech Education and Publishing
Published online 01, June, 2007
Published in print edition June, 2007