Real-time Interaction in Mixed Reality Space: Entertaining Real and Virtual Worlds

Hideyuki Tamura
Mixed Reality Systems Laboratory Inc.
6-145 Hanasaki-cho
Nishi-ku, Yokohama 220-0022, Japan
tamura@mr-system.co.jp
http://www.mr-system.com/

1. INTRODUCTION

In Japan, we have been participating in the “Key-Technology Research Project on Mixed Reality Systems” (MR Project). The task of this project is to build an innovative information technology and a human interface technology that could be pragmatically utilized in the first decade of the 21st century while going beyond the limitations of traditional virtual reality (VR) technology. At the planning stage, people were already using the term “augmented reality” (AR) in reference to a concept to explain augmentation of the real world with electronic information using the power of a computer. The concept of AR is the antithesis of the closed world of virtual spaces. Some problems were already pointed out such as the emotional impediments created by being absorbed in a virtual world, and the physiological influence of head mounted display (HMD) covering the entire view field of the observer. The AR using a see-through HMD was evaluated as having the potential of solving these problems since an observer could see the surrounding space through the HMD.

On the other hand, there was a trend toward the effective use of cyberspace. Note that the rapid growth of the Internet made the information space much greater than the general population. Cyberspace on the Internet is neither a scientific calculation result nor a fantastic illusion produced by imagination or hallucination. It is a place where people can perform realistic business or enjoy their communication. Popularization of this type of cyberspace gradually requires a virtual world that can be used without awareness of the border between the real and virtual worlds. It is also obvious that as the data transmission bandwidths of the computer network become wider, the quality of visualization of this type of cyberspace improves.

Given this, we have adopted Paul Milgram’s “Mixed Reality” (MR) [1] as a theme of our project. His MR includes “augmented virtuality” (AV), the counterpart of AR, that enhances or augments the virtual environment with raw data from the real world (of course, AR is a subset of MR). He considers that AR and AV are continuous (Fig. 1). By adopting the relatively broader concept of MR, the goal of our project has been set to develop a technology seamlessly merging the real and virtual worlds.

Merging or integration of the real and virtual worlds should not be considered from the point of augmentation, which makes one world primary and the other secondary, but rather should be considered from the point of mixture as in MR technology.

Our Mixed Reality Systems Laboratory Inc. was established to conduct this project in January 1997 using funds provided by the Japanese government and Canon Inc. This national project will be extended to March 2001 with the collaboration of three universities in Japan, the Univ. of Tokyo (Prof. M. Hirose), the Univ. of Tsukuba (Prof. Y. Ohta), and Hokkaido Univ. (Prof. T. Ifukube).

In this paper we introduce the outline of the project and three major prototypes developed in the first half of period. On the emphasis of application to future entertainment industries, this paper also describes the function overview, content design, and system configurations of a newly developed multi-player MR entertainment RV-Border Guards.

2. OUTLINE OF THE MR PROJECT

The research themes of the MR project, which has been founded by the Ministry of International Trade and Industry, are officially classified as shown below.

(1) Technologies for merging real and virtual worlds

- To develop technologies for building a mixed environment model from the geometric and radiometric structures of the real world, using 3D image measurement and computer vision.
- To develop technologies those enable the seamless and real-time registration of physical space and cyberspace.
To totally evaluate a mixed reality system integrated with 3D image display.

(2) 3D image display technologies

- To develop a compact and lightweight head-mounted display, with the aim of achieving a mixed reality system that incorporates state-of-the-art optics design theory.
- To develop a high-luminance and wide-angle 3D display without eyeglasses.
- To establish methods to quantitatively measure, evaluate, and analyze the impact of 3D display on people, as well as to obtain physiological data to prevent and minimize hazardous effects. (Such results will be fed back into the design of displays and other equipment to develop imaging and display equipment that reflects the importance of safety and physical comfort considerations.)

One of the characteristics of this project is that it includes the development of new 3D (stereoscopic) displays, as well as research on the methodology or algorithms aimed at a “seamless MR.” It was thought that the development of a new 3D display was inevitable to reproduce the merged results of the real and virtual worlds as realistically as possible. The goal of this project is not only to write a paper or to obtain a patent, but also to realize prototypes that will work in real time and that will be applicable to the pragmatic or commercial systems of 21st century. Thus it is necessary to develop new 3D image displays conforming to MR technology.

“Seamless” is the final goal or slogan of our project, although we do not really think that it is possible to perfectly fuse the two worlds. A technology or a system of technologies that can be flexibly applied according to pragmatic precision or cost requirements is preferable. The term “seamless” implies a fine balancing between opposite requirements at various levels, granulation of the predefined classes into subclasses, and continuous stacking of the technologies that conform to each subclass.

3. MAJOR RESULTS AT THE INTERMEDIATE STAGE

We have developed a number of MR systems from AR to AV. All of these systems are designed to work interactively in real time. Three prototype systems below are the major results developed in the first half of our MR project.

3.1 CyberMirage: Mixed Rendering of Model-Based Data and Ray-Based Data

A few years before starting MR project, we had studied AV. In that study we had been seeking a way to handle objects and their backgrounds having complex shapes that could not be drawn using conventional computer graphics techniques in virtual space. We tried to reconstruct a scene that coincides with the viewpoint of observers from various real images without expressing virtual environment with data based on the geometric models.

Our goal was to find a method to reconstruct an image that produced motion parallax when an observer moved around it. The method had to reconstruct the required image from images captured by multiple cameras placed evenly on a line by interpolating images from those cameras. The problem eventually became the simpler problem of finding a straight line from an epipolar plane image (EPI) [2]. This was really a technique of computer vision or image processing. Applying this theory, we developed the Holomedia system [3] that gives an observer stereoscopic images through liquid crystal shutter glasses with a head tracker. No geometric data was used in this method. Now, such approaches are called “image-based rendering (IBR).”

By generalizing the method based on the EPI, we advanced to image-based rendering based on the “Ray Space.” This method, advocated by H. Harashima and others [4], is one that produces a radiometric representation of an object as a bundle of rays which go through a certain point on a screen at a certain time. Fig. 2 illustrates this. The theory has the same basis as the Lumigraph [5] or light field rendering [6]. All these methods perform image-based rendering from a lot of pictures captured from the real world.

The method stated above has realized a procedure to render a photo-realistic scene without describing an explicit geometric shape. Note that just collecting necessary raw images can generate the image seen from a desired viewpoint. Theoretically, it is proved. However, there is an actual problem involving the image acquisition method and the large amount of data.

We then tried to draw an image by merging geometric model-based data and ray-based data. Finally, we were able to complete a system in which an observer could walk through MR (or AV) space that is constructed by complex objects represented by ray-based data placed in circumstances of polygon-represented graphic data. Fig. 3 shows an example of this type of data structure. The system expanded from the VRML viewer is called CyberMirage [7].

Collaborative CyberMirage [8] is an expanded version of the CyberMirage in which multiple remote participants can visit cyberspace on a network and communicate with each other in real time while recognizing other participants.
as avatars. This system was tested by linking multiple points several dozen kilometers apart from each other with lines of 6 Mbps. The research is mainly reviewed from the point of telecommunication system such as how to compress and transmit huge image-based data such as some megabytes per object.

In the MR project, we studied methods to merge the image-based data in any class of implementation from AR to AV. A series of CyberMirage systems target cybershopping at a virtual mall (Fig. 4). We have already achieved a certain success for the photo-reality of a single object not affected by circumstances.

The next problem to be solved is the shading of an object placed under transitional lighting condition. Shading becomes fixed when we reproduce an object from images captured under fixed lighting. For this issue we developed a real-time rendering method that changes the shading of image-based objects and casts appropriate shadows according to the motion of viewpoint or objects and transitions in local lighting [9].

3.2 AR\(^2\) Hockey: A case study of collaborative AR

We have developed the AR AiR Hockey (AR\(^2\) Hockey) system as a case study of a collaborative AR for human communication. In this study, collaborative AR is a method for establishing an environment in which participants get together and collaborate while sharing physical space and cyberspace simultaneously.

Air hockey is a game in which two players hit a puck with mallets on a table, attempting to shoot it into goals. In
our AR² Hockey, a puck is in virtual space. Each player wears an optical see-through HMD and hits a virtual puck on a real table. Fig. 5 (a) shows the scene of playing AR² Hockey and Fig. 5 (b) is an image seen through the HMD when the system is operating.

Fig. 6 (a) shows the typical coordinate systems used in a simple AR. Registration is the process that transforms the viewing matrix $C_C$. In collaborative AR, all the participants share physical space and virtual space. Thus the coordinate systems $C_R$ and $C_V$ exist in the system, and are shared by the participants. At the same time, the coordinate systems $C_C$ and $C_D$ that relate to the viewing transformations exist for each participant. Fig. 6 (b) illustrates this situation. Thus, the registration algorithm is implemented independently for each participant.

The optical see-through system is used to the AR² Hockey system so that the players (observers) can easily recognize opponents by their eyes. A Polhemus sensor and a CCD camera are mounted on the HMD of the players. The CCD camera is used not to see outside through the captured video image, but rather to register the virtual object based on the captured image.

The paper [10] explains the first version of our AR² Hockey. This system has been modified greatly to be exhibited at the “Enhanced Reality” area of SIGGRAPH 98 [11]. We had to enhance the system throughput in order to operate the system using SGI O2 computers exclusively without any ONYX2 computer. The shape and placement of landmarks were also modified for more accurate registration.

During SIGGRAPH 98, more than 1,000 couples (2,000 players) played the new AR² Hockey. One of the most significant characteristics of the SIGGRAPH 98 version of AR² Hockey is that anyone could play without any difficulty. They did not have to be developers of this system or trained players. It may be said that the new AR² Hockey became the see-through AR system experienced by the largest number of users worldwide.

3.3 MR Living Room: A case study of visual simulation with AR

Using the AR² Hockey system, we studied mainly static and dynamic registrations, that is, positional misalignment and time lag, by taking a game requiring quick motion as the subject of our research. “MR Living Room” is another
experimental AR system for interior simulation. This has been developed using the knowledge obtained from the AR² Hockey project while taking technical problems related to image quality consistency into consideration. This section outlines this project.

The MR Living Room has a 2.8 m x 4.3 m floor made of wooden flooring staff half-equipped with a few pieces of furniture and articles. In this space, two observers with see-through HMDs can experience virtual interior simulation such as selecting and placing furniture. Fig. 7 (a) shows the inside of the experiment space. As shown in Fig. 7 (b), this room is half-equipped with a few pieces of physical furniture and articles. Virtual furniture and articles are merged into this physical space and presented in real time onto the HMDs. The augmented views are shown in Fig. 7 (c) and (d).

As seen Fig. 7 (c) the sufficient image quality that is quite important for this kind of simulation cannot be obtained by the optical see-through mode. The virtual puck of AR² Hockey system was designed to have the
highest brightness and no problem was encountered. On the other hand, the dark virtual objects in MR Living Room such as a tree are almost invisible or look like ghosts under the bright environments. The same issue is sure to arise in a shiny outdoor scene. Thus, we chose the video see-through mode for the purpose of this system.

Geometric registration in this system is attained by fusing sensor-based method and image-based method as in the case of the AR2 Hockey system. In the MR Living Room system, as the physical trackers, an ultrasonic sensor to measure the observers’ position and a gyroscopic sensor to detect the observers’ direction are used. However, no Polhemus sensor is used because the area in which the observer can move around is much greater than in the AR2 Hockey system. Since landmarks (fiducials) placed on the table of the AR2 Hockey system are not as elegant as in the living room, the system uses small devices emitting infrared rays placed on such places as walls or a bookshelf. One of the two CCD cameras mounted on the HMD detects these infrared ray markers and the other is used to obtain a video signal for the video see-through image.

Since the system requires a greater registration area and higher registration accuracy than the AR2 Hockey system, we have developed a new method to detect the position and posture of the observer by using multiple landmarks. In this system, the number of observable landmarks varies depending on the viewing angle of the observer. Therefore, we have to work out an algorithm [12] to adjust misalignment adequately based on the number of observable landmarks. Fig. 8 shows the case when two or three landmarks can be observed.

This algorithm forms a single framework that can treat both the cases that depends only on a physical tracker and that which depends only on landmarks as in the image-based method. It is remarkable that this algorithm can also treat the intermediate state between these two cases (Fig. 9). It is quite important for a pragmatic system that can treat the intermediate state between two extreme methods seamlessly to cope with various occasions in the actual application.

As a guide of this MR application, we have recently embodied an anthropomorphic interface agent who can understand the user’s demand, and move and replace the virtual objects (Fig. 10). Most of other conversational agents so far investigated exist in a rectangular window on a computer monitor, but our MR agent is living in 3D space shared with a user. Such agent’s behavior and the user’s preference are good subjects in HCI research.

4. ENTERTAINING THE MIXTURE OF REAL AND VIRTUAL WORLDS

4.1 Application to Future Entertainment Industries

Most of the people may think the mixture of real and virtual worlds as the composition of real world images and computer generated images (CGI) popular in feature films. Our MR technology has some similarities with these visual effects (VFX) technologies but is not equivalent to them. Generally VFX used in modern motion pictures or TV commercial films is a result of postproduction. This means that we can composite real and virtual images manually frame by frame. Also, the audience has no freedom to change their viewpoint and the sequence of images since they are determined by the director’s intention.

On the other hand, users can see both real and virtual worlds automatically merged in real time and even they can interact with the resultant mixed world. Many existing VFX techniques can be applied to MR if we can overcome the strict limitation of the real-time composition. We have to degrade the quality of CGIs because of this limitation. However, this will be a little problem since we can expect fast growth of computing power still in the next decade. We think it is possible to apply VFXs used in the current feature films to MR systems in few years.

MR technology has a wide variety of applications as in education, architecture, urban planning, manufacturing, medicine and welfare. Since the beginning of MR Project more interest has been directed to this field and many newcomers have appeared [13][14]. In all of them, the entertainment industries are considered to be the biggest fields of application.

Since AV is a VR technology of faithful delineation, it may be applied to the most of domains in which computer graphics are currently used. AV may be useful to improve photo-reality in movies and video games, because AV does not require time-consuming geometric modeling of complex objects.
Interactive real-time AR system may also be applied to the production for movies or TV programs. It composites real world image and CGI from arbitrary viewpoint and the results of composition can be confirmed in the intermediate stage of production. Therefore the stuff can adjust CGIs and rehearsal may be much simplified. It will certainly improve the productivity of film-making.

Games are most direct fields of application. In fact, many game industries paid great attention to the AR^2 Hockey, developed as a research example of collaborative AR. Most of video games so far, for arcade game machines or consumer game machines, force players to watch TV-type monitors while playing the games. Essentially players can use MR space and direct games with their physical action of the whole body. A multiplayer AR/MR game has an advantage to the full-virtual games using HMDs, since it can utilize the reality of the real world as the playground and players can see their partners and/or opponents in the real world. This increases the freedom in planning and designing of games and improves the quality as an entertainment.

Unfortunately, the AR^2 Hockey did not utilize this characteristic of MR entertainment. Note that there was only one virtual object, the puck, and its action was limited to the 2D plane. RV-Border Guards shown below utilizes the advantage of MR space far more than the AR^2 Hockey, and is maturity as an entertainment is much greater.

4.2 RV-Border Guards: A Multi-player MR Entertainment

(1) Function Overview
RV-Border Guards is a game to utilize 3D-MR space and has the following functions in addition to those of AR^2 Hockey:
(a) to have more than three players,
(b) to render multiple virtual objects lit in the adequate way,
(c) to make virtual objects move and transform in 3D-MR space,
(d) to achieve the occlusion between real and virtual objects,
(e) to achieve the reflection of real world scene onto virtual objects,
(f) to produce spatial sound effects,
(g) to accept action commands by means of gesture, and
(h) to display objective (composite) view besides the viewpoints of players.

Fig. 11 shows a scene of the game played in this system. Three players wearing HMDs place themselves around a game field (Fig. 11 (a)). This system uses the video see-through method as in the MR Living Room since it is easier to adjust visual consistency between the real and virtual worlds such as contrast, color tone, resolution, and latency. Since the composite image by the video see-through method can easily be taken out as video signals, the audience can also see the mixed world that is seen by a player. The image shown in Fig. 11 (b) is taken by placing
a video camera in a location different from players and merged with virtual images in the same way. This image is also helpful for the audience to know how the game is going on.

(2) Content Design

“RV-Border Guards” means guards at the border between the real and virtual worlds (RV-Border). In this game, players compete in earning points by shooting invaders from the virtual world. The reasons why we have adopted a shooting game are because it is possible to achieve 3D positioning of players and virtual objects in MR space and to define several simple rules. The 3D space around the players is spatially used in this game by making invaders (virtual objects) fly around the space.

People play this game by wearing the devices shown in Fig. 12. However, players in the MR space find themselves wearing virtual helmets and virtual guns (Fig. 13). The virtual helmet is designed to cover the projected parts such as a CCD camera and a magnetic sensor. The virtual gun is designed to cover the player’s hand and arm.

Fig. 14 is an image seen from a player. In order to present a stereoscopic view as clearly as possible to the player in the limited view field, only minimum amount of text such as score and remaining time is superimposed to help playing.

Players interact with the mixed world by means of arm action. A magnetic sensor built into the equipment put on the hand recognize the movement of arm and generate three kinds of commands (ready, fire, and defense) as shown in Fig. 15.

There are also three kinds of targets (invaders) - jellyfish-type, shark-type, and rocket-type (Fig. 16) - are designed to have their own surface property and action pattern. Any of the invaders appear from the MR space above the table, then moves around for several seconds, rushes at the player’s head, and finally crushes against the player. The motion of the target rushing to the player emphasizes the effect of stereoscopic view.

Invaders cast their shadow onto the real objects such as floor, table or objects on the table; they can also hide themselves behind or beneath the objects. It means that the system has the visual consistency between the real and virtual worlds. In order to achieve this, 3D geometric models of a necessary part of the real space must be obtained in advance. The environmental mapping technique is used to render each invader so that each surface reflects the image of the real environments. Actually a real scene reflected by virtual objects depends to each player’s location, and so different environmental texture images are prepared for each player.
Fig. 17 and Fig. 18 show two sequences of images seen by a player through the HMD while playing this game.

(3) System Configurations
This system uses a client/server method based on the de-coupled simulation model [15] in which many processes such as conversation input, positional sensing, audio/visual sensory presentation, and database management are handled separately. Client modules which share the same game servers are placed around the server (Fig. 19). The game server maintains the database of information such as players’ status, interactive command entry and simulated virtual environment. These clients can be classified into the following three categories.

(a) Master module controls the overall system and

- #1: Found. A jellyfish is in the scope. (A player’s gun is visible.)
- #2: Ready. A laser beam is targeted at the invader.
- #3: Fire. A fire bullet is discharged.
- #4: Hit. The invader is vanishing with a flush of light.

**Fig. 17 Sequence 1: A target is destroyed.**

- #1: A shark turns into the offense mode and is rushing to the player.
- #2: Just before the crush. It is too late to defense.
- #3: The player is damaged and his view is out of order.
- #3’: Another player’s view at the crush above. The shark explodes.

**Fig. 18 Sequence 2: A player is damaged.**

Fig. 19 Block-diagram of RV-Border Guards
manages the progression of the game. It also simulates the behaviors of virtual objects and updates them.

(b) Player module enables a player to participate into the MR space. Each player has his/her own player module. This module detects interaction and tracks head movement, and then generates the composite MR view based on the information from the game server.

(c) Observer module provides a composite MR view from an observer’s camera position to people other than players. Highly modular design of this system provides great flexibility and scalability. This means that we can easily increase the number of players and observer’s cameras provided that there is enough computing power. It should be noted that the same framework could be used to various games other than shooting.

5. CONCLUDING REMARKS

This paper describes the concept of the MR technology in our project and provides the major results to date. Now, research in this project is concentrated on visually merging the real and virtual worlds. However, this does not mean that auditory or haptic stimuli are not applicable to our MR system. Though the original and innovative research in our project is concentrated on visual information, we are going to look into ways of incorporating other sensory data into our total MR system. It should be noted that the enhanced version of AR2 Hockey has vibrators in the mallets, allowing players to feel vibrations when they hit a puck. In addition, we are going to implement a 3D sound system into the MR Living Room that generates sound from virtual equipment such as a TV set or audio set. Of course, planners or producers of such games as the RV-Border Guards may expect more diverse and tactile sensing devices.

There is a growing movement to enhance contents of the MR entertainment. PC-game industries and multimedia software industries have organized the Mixed Reality Entertainment Conference (MREC) in Japan. MREC now holds a contest to collect ideas of MR entertainment and commends excellent ideas. Their aim is to find out some mine of new entertainment other than conventional video games. We are expecting contents of new concept quite different from those of engineers participating in the MR project.

We have to solve several problems to make the MR entertainment so popular as the existing video games or movies. The biggest problem is its cost. Although the graphic computing power of the consumer game machine is drastically improving, MR entertainment, especially AR type entertainment, requires devices such as HMDs, video cameras, and position sensors. It also requires a broader playing field for fully utilizing its advantage. Considering these factors, we expect that the MR entertainment will be spread over theme parks and other high-class LBE first and then downsized towards consumer game machines in the future.

REFERENCES