Interactive Immersive Display

Sang Chul Ahn, Ig-Jae Kim, Hyoung-Gon Kim, and Sungdo Ha
Imaging Media Research Center, KIST
39-1 Hawolgok-dong, Sungbuk-gu, Seoul 136-791, Korea
{asc, kij, hgk}@imrc.kist.re.kr

Abstract

'Tangible Space' is the concept of connecting the cyber space and real world seamlessly so that people can extend their capability over the limitations of each space. The seamless integration of cyber space and real world requires natural and efficient interfaces to the cyber space. We define 'Tangible Interface' as a surface where people meet the cyber space. In this paper, we present technology for building interactive immersive display as a Tangible Interface. In our vision, as computers are disappearing from our sight in the near future, interactive display will be a main interface with cyber space. This paper presents a method to build a tiled display as an immersive large scale display using off-the-shelf graphics boards, and live avatar based interaction with a large scale display. We expect them to be combined into the interactive immersive display as a natural and efficient interface.

Key words: Tangible Interface, Interaction, Immersion, Tiled Display, Live Avatar

1. Introduction

The innovative development of information and digital technology makes it possible to open a new paradigm of usual lifestyle, which was even unpredictable to us before. Nowadays, as computers are getting smaller and smaller, and are embedded inside a lot of things around us. As ubiquitous computing environment and/or pervasive computing environment are spread, cyber space becomes important to our lives. Cyber space created by computers is being developed as a complementary space to the real world, where people can work and spend time.

'Tangible Space' is the concept of connecting the cyber space and real world seamlessly so that people can extend their capability over the limitations of each space. The main limitations of the cyber space are naturalness and utilization problem. We cannot feel natural with computer generated cyber space, and utilization of it is not efficient, either. One of the sources of the limitation is the interface with the cyber space.

We define 'Tangible Interface' as a surface where people meet the cyber space. Then, the Tangible Interface should provide seamless integration of real world and cyber space. The Tangible Interface has to give reality and natural feeling to users while they are using it. It should be efficient to utilize, too.

As computers are getting smaller and smaller, and are disappearing from our sight, we need a new type of interaction mechanism with the computers and the cyber space generated by them. In this environment, we don't care where the computers are located, but only confirm the result of our commands through some visual interfaces like a large screen display. For reality and naturalness of a visual interface, it is necessary to provide immersion feeling and natural, efficient interaction with displayed objects. Thus, this paper focuses on an interactive and immersive display as a Tangible Interface.

This paper presents a method to build a tiled display as an immersive large scale display. Our approach is to use off-the-shelf graphics boards instead of extremely expensive special hardwares. This paper presents a new method for natural and efficient interaction with a large scale display, too. The main and unique feature of the method is the use of live avatar of user. In this way, a user can feel some natural feeling as if he is in the cyber space, and accomplish efficient interaction with computer.

2. Related Research

In the immersive visual interface researches, there have been two main directions: 3D displays and large displays with high resolution. The 3D displays are classified into two categories, eye-glasses type and the others. Commonly used in commercial field is the eye-glasses type 3D display. The most well known eye-glasses type 3D display is the HMD(Head Mounted Display), which includes displays in itself. These days there are many commercial HMDs, but with limited display resolution. Another eye-glasses type 3D display uses a flat display and filter glasses. They, in turn, is classified into passive and active types along the existence of shuttering action synchronized with images on the display. The most commonly used passive filter glasses are utilizing polarizing filters rather than shutters. Non-glasses type 3D displays contain parallax barrier display, lenticular display, and holographic 3D displays. However, they are on the research stage until now.

The large scale display to cover up the field of view is another important factor for the immersion feeling. As an FPD(Flat Panel Display), 50~60 inch display is the largest one that can be obtained commercially these days. In order to get a large image, the projection system is one alternative to it. However, it is also limited to increase the image size because of the resolution per unit area. Currently projectors with UXGA (1600x1200) resolution are available in the market, but most expensive.

Another approach to implement the large scale display is tiling displays[1,2]. The tiled displays are classified into three types: room type, dome type, and open type. The CAVE[3] is the most well known room type tiled displays. There are 4-, 5- and 6- sided CAVE implementations as reported[4]. Current implementations of the CAVE are limited to one projector displayed on each wall without intensity blending. The dome type displays contain commercially available 'V-dome' of Trimension Inc.[5] and 'StarRider' of E&S[6]. The open type displays contain wall displays and reconfigurable displays. Commercially available are 'Reality Room' of Trimension Inc.[5], 'GVR' series of Panoram technologies[7], and 'RAVE' of Fakespace systems[8]. These systems are using 3-6 projectors. In the research field, immersive systems with more projectors have been studied. Argonne National Lab. and Univ. of Chicago have developed the ActiveMural and the u-Mural[2]. Princeton university has implemented the Scalable Display Wall[9]. And Univ. of Minnesota has made the PowerWall system. The OOTF(Office Of The Future) group of Univ. of North Carolina has been trying to incorporate immersive displays into usual offices or rooms[10].

As the interaction channel of the interactive reality imaging interface, we are considering haptic devices and/or the PUI (Perceptual User Interface)[11]. The PUI is characterized by interaction techniques that combine multimodal computer I/O devices and machine perception of human behavior such as face and gesture recognition. With these capabilities, the interactive reality imaging interface could provide people with "Tangible Space" at hand. Actually an interactive space named 'KidsRoom' was built by MIT Media Laboratory[12]. It was equipped with large screens and computer vision based interaction system. The KidsRoom is similar to our system with the PUI functions. However, while the goal of the KidsRoom is to create an automated and interactive playspace, our goal is to create an interface to the "Tangible Space" with reality and naturalness.

3. Immersive Display

We have developed a tiled display system using multiple projectors to build the immersive large scale display. In order to build a tiled display, we have to solve three problems: realtime image warping, image blending, and large format image generation.

We need image warping to stitch several images and to make them into an image. Further, it should be done in realtime for large format videos. Since the realtime image warping requires heavy computation and high speed image processing, tiled displays previously have been provided by special hardwares at extremely high cost. However, there has been great improvement in the performance of GPUs (Graphics Processing Units), the core of commercial graphics cards these days. Thus, it is considered that the GPUs can handle the realtime warping function.

We need a tool to control the high performance GPU, and to make a computation using it. The typical tools are the DirectX of Microsoft and the OpenGL of SGI. They are high level APIs (Application Program Interfaces) for graphics programming. Nvidia has developed Cg (C for Graphics) language for low level programming of GPUs, too[13]. We decided to use the DirectX to control GPUs because we can use rendered screen as a texture with it. Using a texturing function, we can implement image warping of rendered screen in realtime. Fig. 1 shows the flow of image warping process.

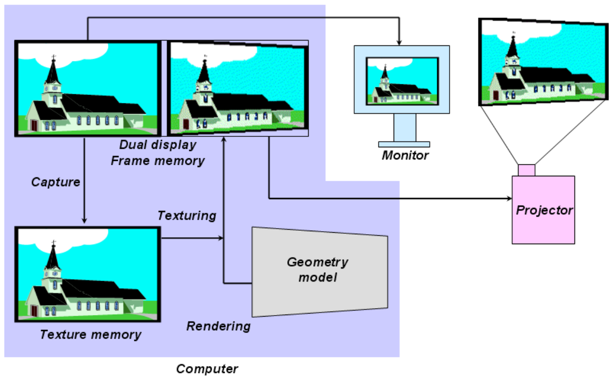


Fig. 1 The flow of image warping process.

In this process, we use dual display function of graphics cards. We displays on the first screen an image that we want to warp. Then, we warp the rendered image on the first screen and display it on the second screen. The second screen is projected so that it consists of a part of the tiled display.

Using the DirectX, we let an image generation program render images at the first half of frame memory. The content of the frame memory is the main screen that is shown to the user. The first half of the frame memory is captured as a texture into the texture memory. The texture memory is not shown on the screen. The second half of the frame memory is reserved for warped image. The texture is used to render a geometrical model at the second half of the frame memory. The geometry of the model is controlled by users. The model is set to have a specific geometry so as to result in a desired image warping when it is textured and rendered. It should be determined according to the relation among the locations of screen, projectors and viewer. If the screen is not flat, the screen geometry should be considered when we determine the geometry of the model. The second half of the frame memory is projected on the display. Fig. 1 shows this process. For every tile of the large display, we use a computer with dual display function, and go through the previous process. For instance, if we want to make a display with 4 tiles, 4 computers are needed.

When we stitch multiple image tiles to make them a large image as a whole, we need to find how much they should be warped. Though this can be done manually, it should be tedious job. Thus, we have developed a program for automatic computation of warping parameters. It is a similar approach to the previous researches using computer vision technology[14]. The program detects displayed region for each tile using computer vision based feature detection, and computes warping parameters that determine how much the image tile is warped. Fig. 2 shows a screen shot of the program. The program captures empty screen of each display by controlling a camera. We use a camera that is attached to a pan-tilt control mechanism. Then, it detects region of each display, and generates warping parameters to make a rectangle image region as a whole.

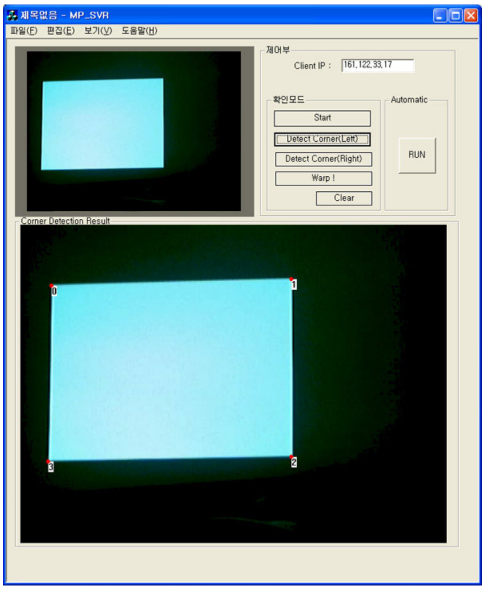


Fig. 2 A screen shot of warping parameter generation program.

Fig. 3 shows the basic configuration of the system, where the center camera captures each of the image tiles one by one. Once the warping parameters are generated by a main computer, they are distributed to the other computer automatically to be used for image warping.

Once the image warping is done, the next process is image blending. When we stitch multiple image tiles, there occur some overlapped regions between displays inevitably. In the overlapped regions, image is brighter than the other parts and colors are different. In order to make a seamless tiled display, we have to detect the

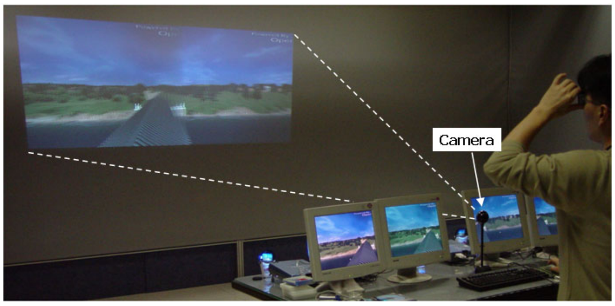


Fig. 3 Basic configuration of image warping and blending system. The center camera is used for generation of warping parameter and blending mask.

region and adjust the brightness and colors. This is the work to be done in the image blending process. In this process, we use multi-texturing technology for adjusting the brightness and colors. Since most of GPU's support multi-texturing functions, it can be done in realtime. Since warping parameter generation program can detect overlapped regions, we added to it a texture mask generation function for image blending. The texture mask is applied to image rendering, too. The texture mask is dependent on computer and projector. Especially it depends heavily on the properties of projector such as color gamut and color temperature. In our approach, we examine the properties of projector and determine color adjusting function assuming static overlapped region in offline process. The color adjusting function is a decreasing function that decreases from the start of overlapped region to be zero at the end of the image. Then, we generate texture masks along the varying size of overlapped region using the color adjusting function.

Since we use multiple computers to generate a single tiled display, we need to synchronize image generation programs among the computers. Currently, we are using NAVER platform to generate images and to synchronize them. The NAVER is a visualization platform for a computer cluster, and has been developed by another team in the TSI project. The NAVER platform is based on the OpenGL performer library of SGI, and provides image rendering functions. It provides software based synchronization mechanism among cluster computers, too. Thus, the NAVER is quite appropriate for tiled display system. Even the NAVER platform provides the function for asymmetric view frustum, which is quite needed to generate images of asymmetric field of view of virtual camera. This function is one of requirements to generate 3D image of virtual space in a tiled display system.

Now, all the three components for tiled display are prepared, image warping, image blending, and image generation. Fig. 4 shows a shot of experiment for generating a large display of 4 tiles. Even though the experiment was done with 4 tiles, our system is scalable, and it can be expanded to any number of tiles by increasing the number of computers and projectors.

Actually we are building a tiled display with 8 projectors side by side to cover up all the wall of a room.

Fig. 4 Experiment for generating a large display of 4 tiles.

4. Interaction with a Display

This paper presents a new mechanism of interaction with a large display using live video avatar of user and computer vision based hand gesture recognition[15]. The proposed system enables a user to appear on the screen as a video avatar, and to walk and navigate across the screen. The system also allows the user to interact with items or icons in the screen by touching them with a hand. In order to build the system, we have developed live video composition, active IR vision-based hand gesture recognition systems, and incorporated voice recognition system, too. Later we added 3D human body tracking to the system for 3D interaction.

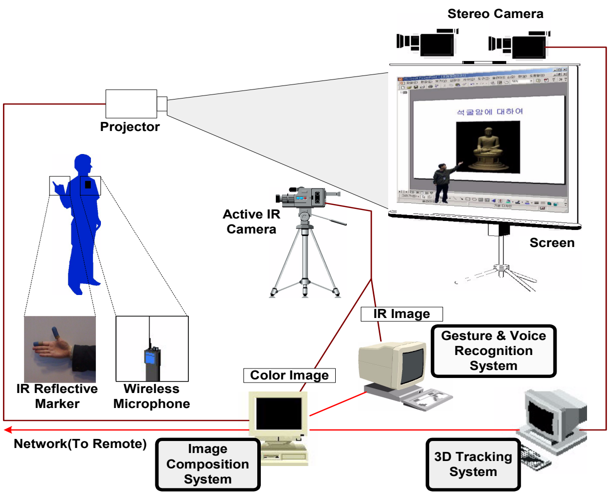


Fig. 5 The configuration of interaction system

The proposed system consists of three subsystems: *Image composition subsystem, Gesture & voice recognition subsystem,* and *3D Tracking subsystem*. Fig. 5 shows the configuration of the system, and three computers represent the subsystems. The user interacts with the large screen in front of a camera while looking at the screen. The image composition subsystem extracts the user image and overlays it on the screen. So, the user sees himself as a video avatar on the screen during interaction. The user can interact with icons and items within the screen. He can use hand gesture and voice to control them. For robust hand gesture recognition, we used an IR (Infra Red) camera and IR reflective markers. The 3D tracking subsystem uses a stereo camera and tracks the user's position. The user can use this subsystem for 3D interaction. As a whole, this proposed system gives us a more intuitive interaction method with a large screen.

In order to make a live video avatar, the image composition subsystem captures the images of a user with video camera. In Fig. 5, we can see that an active IR camera is capturing the images of the user. This active IR camera is a combination of an IR camera, an IR light source, and color video camera. The structure of the active IR camera is shown in Fig. 6.

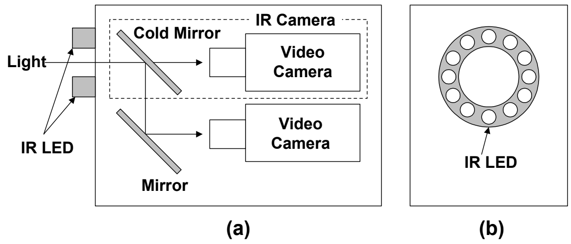


Fig. 6 The structure of the active IR camera

The IR camera and IR light source are used in the gesture & voice recognition subsystem. The image composition subsystem uses the output of the color video camera. The image composition subsystem extracts the image of a user from the background, and uses it as a live video avatar. Background subtraction or chroma-keying method can be used in this process. The video avatar is overlaid on the screen. The composition of video avatar can be done with a video processing hardware or a chroma-keying hardware. However, we implemented it in realtime by software. The user can control the position of the video avatar by walking around the stage as long as he remains within the field of view of the video camera. Fig. 7 shows the image composition mechanism of video avatar.

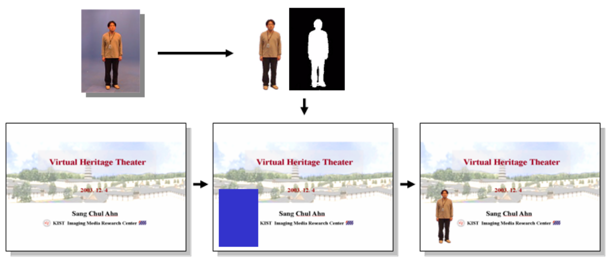
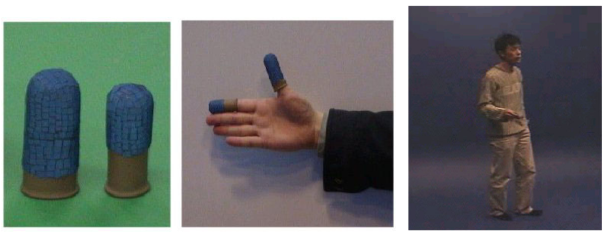


Fig. 7 The image composition mechanism of video avatar.

The proposed system uses hand gesture and voice as the main interaction tools. The gesture & voice recognition subsystem enables the user interaction by vision-based hand gesture recognition and voice recognition. For instance, a user can issue a command to play music by saying “Music Start” to a microphone. By hand gesture “Up / Down / Left / Right”, he can also move his video avatar across the screen to select or play an item. This is useful when it is too far from him to reach by walking within the field of view of camera. We implemented 7 hand gesture commands and 19 voice commands. In our implementation, hand gesture commands were included in the voice commands. However, we found that hand gesture commands were more useful for 2D/3D motion than voice commands. Voice commands were more useful for discrete actions such as start or end of some action.

Although computer vision-based hand gesture recognition has been widely studied, it also inherits the shortcoming of most computer vision algorithms: sensitiveness to lighting condition. Thus, we adopted active IR based recognition method for robust recognition. As mentioned before, an active IR(Infra Red) camera is used for capturing a user’s hand gesture. The IR camera can be made with a normal video camera and an IR filter. In Fig. 6, the upper part composes the IR camera. The cold mirror is an IR filter that absorbs IR rays while reflecting visible rays. We used a cold mirror that absorbs the rays of above 800nm in wave length. Additionally we made two IR reflective thimbles for user’s hand. The IR reflective thimbles were made with retro-reflective material so that they can be viewed best from the camera with IR light source. Fig. 8 shows the IR reflective thimbles and a user’s hand wearing them. Since thumb and index finger are mostly used for selecting and pointing, we decided to use the thimbles for thumb and index finger.



(a) (b) (c)

Fig. 8. IR reflective thimbles, (a) thimbles, (b) a hand wearing thimbles, (c) user interacting by hand gesture.

Since the IR reflective thimbles look white in an IR image, we can extract the regions by simple thresholding. Then, we apply the labeling operation to the regions, and find thimble regions of thumb and index finger by size. The hand gesture is recognized by relative position and direction of thimbles. Once the hand gesture is recognized, it is converted to a command and sent to image composition subsystem by network. Then, image composition subsystem applies an appropriate action to the screen. Here we have to make the IR image have the same coordinates to that of normal video image. That’s why we use the active IR camera structure as in Fig. 6.

In the proposed system, a user can use 3D interaction. The 3D tracking subsystem extracts a user out of the images from stereo camera, and gets his 3D position. The information is sent to image composition subsystem so that it can use the 3D information, for instance, to overlay a video avatar at appropriate 3D position.

We applied the proposed system to build an intelligent room called “Smart Studio”. In the Smart Studio we could interact with a large screen and use an embedded computer in the environment. We could enjoy music and navigate internet as we want. We could enjoy some games with whole body action, too. From our experiments, we noticed that the proposed system provides efficient interaction mechanism in space. Fig. 9 shows some of the screen shots of various applications.

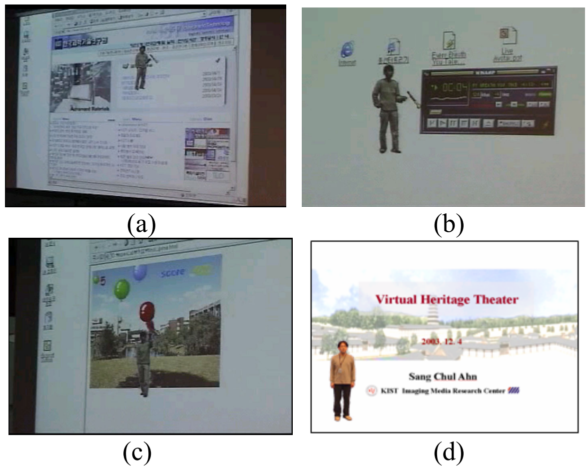


Fig. 9. Screen shots of various applications, (a) internet surfing (b) music control (c) game (d) Powerpoint presentation.

5. Conclusion

In the near future society, where computer and information technology play a more important role, virtual environment will create a new paradigm of daily life, which we never experienced before. The gap between reality and virtual space will be more and more reduced and many innovative technologies will be developed for it. Developing tangible interface technology to effectively access to the cyber world is so future oriented and challengeable in that sense. Here in this paper we have presented a new method to build an interactive immersive display as a visual interface of Tangible Interface. We used commercial graphics boards to build a large scale display instead of expensive special hardwares. We have also proposed a live avatar based interaction mechanism with the large scale display. It provides a natural and efficient interaction with display. We think that this type of interaction with display can prevail in the near future when the computers get smaller and smaller, are embedded in our living space, and disappear from our sight.

References

1. R. Raskar, G. Welch, and H. Fuchs, "Seamless Projection Overlaps Using Image Warping and Intensity Blending," in Fourth International Conference on Virtual Systems and Multimedia(Gifu, Japan), Nov. 1998.
2. M. Herald, I.R. Judson, and R.L. Stevens, "Introduction to Building Projection-based Tiled Display Systems," IEEE Computer Graphics and Applications, pp.22-28, July/August 2000.
3. C. Cruz-Neira, D. Sandin, and T. DeFanti, "Surround-screen Projection-based Virtual Reality: The Design and Implementation of the CAVE," in SIGGRAPH 93 Conference Proceedings, Aug. 1993.
4. TAN ProjectionTechnology, http://www.tan.de/.
5. Trimension Systems Ltd. http://www.trimension-inc.com/.
6. Evans & Sutherland(E&S), http://www.es.com/.
7. Panoram Technologies, Inc. http://www.panoramtech.com/.
8. Fakespace Systems, Inc, http://www.fakespacesystems.com/.
9. K. Li, H. Chen, Y. Chen, D.W. Clark, P. Cook, S. Damianakis, G. Essl, A. Finkelstein, T. Funkhouser, T. Housel, A. Klein, Z. Liu, E. Praun, R. samanta, B. Shedd, J.P. Singh, G. Tzanetakis, and J. Zheng, "Building and Using A Scalable Display Wall System," IEEE Computer Graphics and Applications, pp.29-37, July/August 2000.
10. R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs, "The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays," in SIGGRAPH 98 Conference Proceedings, pp.179-188, July 1998.
11. M. Turk, G. Robertson, "Perceptual User Interfaces," Communications of the ACM, vol.43, no.3, pp.33-34, Mar. 2000.
12. A. Bobick, S.S. Intille, J.W. Davis, F. Baird, C.S. Pinhanez, L.W. Campbell, Y.A. Ivanov, A. Schutte, and A. Wilson, "The KidsRoom," Communications of the ACM, vol.43, no.3, pp.60-61, Mar. 2000.
13. William R. Mark, R. Steven Glanville, Kurt Akeley, Mark J. Kilgard. 2003. "Cg: A System for Programming Graphics Hardware in a C-like Language". In Proceedings of SIGGRAPH 2003.
14. Han Chen, Rahul Sukthankar,Grant Wallace, Kai Li, "Scalable Alignment of Large-Format Multi-Projector Displays Using Camera Homography Trees", IEEE Visualization 2002.
15. Sang C. Ahn, Tae S. Lee, Ig J. Kim, Yong M. Kwon, Hyoung G. Kim, "Large Display Interaction using Video Avatar and Hand Gesture Recognition," Image Analysis and Recognition, LNCS3211, Oct. 2004.