IEEE Virtual Reality 2006

Alexandria, Virginia, USA
March 25 - 29, 2006

Workshop Proceedings

Emerging Display Technologies - New Systems and Applications: From Images to Sensing, Interaction and Enhancement

Organized by
Andreas Simon
Greg Welch
Mark Bolas

SPONSORED BY IEEE COMPUTER SOCIETY VISUALIZATION AND GRAPHICS TECHNICAL COMMITTEE

EDT 2006

Sunday, March 26, 2006

Emerging Display Technologies—New Systems and Applications From Images to Sensing, Interaction and Enhancement

Andreas Simon (University of Applied Sciences Northwestern Switzerland)
Greg Welch (University of North Carolina at Chapel Hill)
Mark Bolas (University of Southern California and Fakespace Labs)

The recent flurry of display technology development has produced families of technologies that make fixed and projected pixels cheaper, faster, more flexible, and of higher quality. These advances enable "smart pixels" and enable a number of burgeoning applications ranging from displays being used for better and more flexible images, to user interaction, scene sensing, and environment enhancement. Some example workshop submission topics include:

- multiview, multifocal, or high dynamic range displays;
- omnistereo projection systems;
- *ad hoc* or "poor man's" projection systems;
- ultra wide field of view HMD optics;
- ultra fast displays;
- head-worn or hand-held (mobile) paradigms;
- hybrid display systems and applications;
- adaptive projector display systems;
- extended color gamut or color matching displays;
- projector-based user/device tracking, interaction, or Mixed Reality reconstruction;
- embedded pixels for Spatially-Augmented Reality; and
- rendering techniques associated with the above.

EDT 2006 will be the second "Emerging Display Technologies" workshop. It should provide an opportunity to expand attendee thinking about ways to use contemporary display devices in VR systems and applications.

Workshop Organizers

Andreas Simon is a Senior Researcher and a Lecturer at the Academy of Art and Design of the University of Applied Sciences Northwestern Switzerland. His current research interests are the development and evaluation of rendering and interaction techniques for immersive virtual environments. Andreas Simon graduated with a M.S. in Computer Science from the Technical University in Berlin. He is a member of IEEE and of the German Gesellschaft fr Informatik. He has developed interaction devices and display systems like the YoYo and the i-Cone. He can be reached at `ansimon@gmail.com`.

Greg Welch is a Research Associate Professor in the Computer Science Department at the University of North Carolina at Chapel Hill. His research interests include Virtual/Augmented Environment tracking systems and Telepresence. Welch graduated with highest distinction from Purdue University with a degree in Electrical Technology in 1986, and received a Ph.D. in Computer Science from UNC-Chapel Hill in 1996. Prior to coming to UNC, Welch worked at NASAs Jet Propulsion Laboratory on the Voyager Project, and Northrop-Grumman's Defense Systems Division on the AN/ALQ-135 electronic countermeasures system. Welch is a member of the IEEE Computer Society and ACM. He can be reached at `welch@cs.unc.edu.`

Mark Bolas is a Visiting Associate Professor in the Interactive Media Department at the University of Southern California, and is Chairman of Fakespace Labs, Inc. His research focuses on creating technologies and content for fluent interaction and perception of synthetic environments. Bolas graduated from the University of California, San Diego with a B.A. in Physics and studies in Music and Signal Processing, and from Stanford University with an M.S. in Mechanical Engineering, Product Design. Bolas previously worked at Stanford University as a Consulting Assistant Professor and helped to found Polaris Road and Sonicbox Inc., as well as Fakespace Labs and Fakespace Systems. He can be reached at `bolas@well.com`.

Message from the Workshop Organizers

Dear EDT 2006 attendees,

So, how many computers are in the room with you now? How many screens and keyboards surround you? Think about your office, your lab, your mobile phone, your PDA, ...

As our culture and society become increasingly enveloped by computational technologies, we are quietly entering the world of a continuous cyberspace as real as that predicted by science fiction writers. The only difference is that the connection jack that was to have been implanted in the back of our skulls is in fact turning out to be our collections of display and interface devices.

We feel it best to embrace this inevitable path by inventing, studying, discussing, and understanding such devices, and thus—this workshop.

Andreas Simon, Greg Welch, and Mark Bolas

Program and Table of Contents

Sunday, March 26, 2006

8:00–8:20 Welcome and Introduction

8:20–10:00 Session I

10:00–10:30 Coffee Break

10:30–12:00 Session II

12:00–13:00 Lunch

13:00–13:50 Session III

14:00–15:00 Mini Tutorial

15:00–15:30 Coffee Break

15:30–17:10 Session IV

17:10–17:30 Closing

8:20–10:00 Session I

A Large Scale Interactive Holographic Display

Tibor Agócs, Tibor Balogh, and Tamás Forgács
Holografika, Hungary*

Fabio Bettio, Enrico Gobbetti, and Gianluigi Zanetti
CRS4, Italy†

Eric Bouvier
CS Communication & Systèmes, France‡

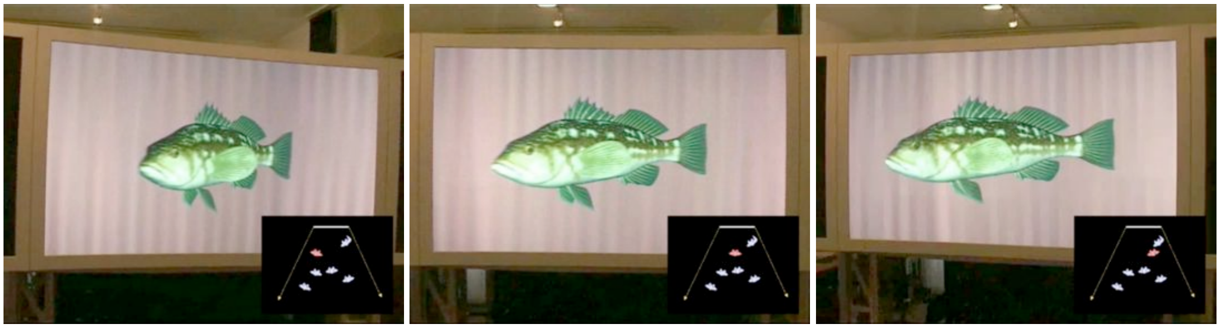


Figure 1: **Holographic display example.** The images that were taken from different positions in front of the display.

Abstract

Our work focuses on the development of interactive multi-user holographic displays that allow freely moving naked eye participants to share a three dimensional scene with fully continuous, observer independent, parallax. Our approach is based on a scalable design that exploits a specially arranged array of projectors and a holographic screen. The feasibility of such an approach has already been demonstrated with a working hardware and software 7.4M pixel prototype driven at 10-15Hz by two DVI streams. In this short contribution, we illustrate our progress, presenting a 50M pixel display prototype driven by a dedicated cluster hosting multiple consumer level graphic cards.

CR Categories: B.4.2 [Input/Output and Data Communications]: Input/Output Devices—Image display

Keywords: holographic displays, 3D displays

1 Short overview

We present a large scale interactive multi-user holographic display that allows freely moving naked eye participants to share a three dimensional scene with fully continuous, observer independent, parallax. The display is an instance of a scalable holographic system design based on a specially arranged array of projectors and a holographic screen. The feasibility of such an approach has already been demonstrated with a working hardware and software 7.4M pixel prototype driven at 10-15Hz by two DVI streams[1]. In this short contribution we illustrate our progress, presenting a large scale holographic display prototype with 50M pixel overall resolution. The display is driven by a dedicated cluster hosting multiple consumer level graphic cards.

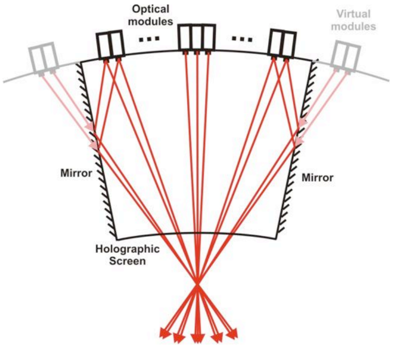


Figure 2: Schematic diagram. A large number of light beams can create a spatial point.

Display concept. Our 50Mpixel display prototype uses a specially arranged array of 64 XGA commodity projectors and a holographic screen with a diagonal of 1.8m. The projection modules project their specific image onto the holographic screen to build up the 3D scene. The applied distributed image organization makes it fundamentally different from other known multi-view solutions. The module views are not associated with specific view directions. Instead, the light beams to be emitted by the projection modules, i.e., the module images that are generated by the projectors, are determined by geometry. Each module emits light beams toward a subset of the points of the holographic screen. At the same time, each point of the holographic screen is hit by more light beams arriving from different modules. The light beams propagate to address fixed spatial positions that are independent from the viewer's position. Many modules contribute to each view of the 3D image, thus no sharp boundary occurs between views, and the display offers continuous and smooth change at different image areas, result-

---

*e-mail: {t.agocs|t.balogh|t.forgacs}@holografika.com

†e-mail: {Fabio.Bettio|Enrico.Gobbetti|Gianluigi.Zanetti}@crs4.it

‡e-mail: eric.bouvier@c-s.fr

Figure 3: **Holographic display example.** Images that were taken from different positions in front of the display. Objects appear at fixed physical positions.

ing in a truly 3D experience with continuous horizontal parallax. Multiple projectors illuminate each pixel on the screen, and the optical modules can be seen under different angles by looking from the pixel's point of view. The holographic screen transforms the incident light beams into an asymmetrical pyramidal form. The cut of this light distribution is a long rectangle, where the vertical size of the rectangle is the vertical field of view, while the horizontal size corresponds to the neighboring emitting directions. This configuration corresponds to the horizontal-only-parallax capability of the current prototype display. The principles on which the display is based would make it possible to provide vertical parallax. Doing so, would require, however, another order of magnitude increase in data size, rendering times, and system complexity, for little gain in the visual performance in standard settings. The horizontal light diffusion characteristic of the screen is the critical parameter influencing the angular resolution of the system, which is very precisely set in accordance with the system geometry. In that sense, it acts as a special asymmetrical diffuser. However, with standard diffusers and lenticulars it would be difficult, if not impossible, to produce the shape of the required angular characteristics. The screen is a holographically recorded, randomized surface relief structure that enables high transmission efficiency and controlled angular distribution profile. These fully randomized (nonperiodic) structures are non-wavelength dependent and eliminate moiré, without chromatic effects. The precise surface relief structures provide controlled angular light divergence. The angular light distribution profile introduced by the holographic screen, with a wide plateau and steep Gaussian slopes precisely overlapped in a narrow region, results in a highly selective, low scatter hat-shaped diffuse characteristics. The result is a homogeneous light distribution and continuous 3D view with no visible crosstalk within the field of depth determined by the angular resolution (see figure 2).

Driving the display. The display is driven in parallel by 64 DVI streams, one for each XGA projector. These streams are generated by an array of 16 PCs, connected to the display through four DVI connections each. Each PC runs a server that controls a graphics frame buffer. The server is responsible for generating images associated to a fixed subset of the display rendering modules. In the 50M pixels prototype, each PC generates 4 XGA images using two double head NVIDIA boards controlling a 4096x768 frame buffer. In order to support legacy graphics programs and to simplify the development of new holographic applications, we have implemented an OpenGL compatible front-end. The front-end runs on a client PC and looks to applications like an ordinary OpenGL library which, in addition to executing local OpenGL commands, also transparently broadcasts the graphics command stream to the dedicated cluster driving the holographic display. The client PC is connected to the cluster through dual Gbit ethernet links. The back-end servers listen to the network and decode the stream of multicast graphics commands coming from the client. Once decoded the commands are interpreted and sent to the local graphics renderer. The interpretation of the graphics commands involves modifying the way they are generated according to parameters available from the local configuration service, in order to transform the original central view into the view associated with each of the associated optical modules. For each of the optical module views, the graphics commands of the current frame are re-executed, with the following modifications: the original perspective matrix is replaced with a matrix that matches the module's specific position and viewing frustum; a geometrical calibration is performed, to correct nonlinearities in the display/optical geometry; a light calibration is performed to correct the intensity and contrast differences response of the optical modules; an angular resolution correction is performed for depth dependent anti-aliasing. The parameters required for each of these transformations are defined at configuration time.

Implementation and results. We have implemented a prototype hardware and software system based on the design discussed in this paper. The developed large scale prototype display is already capable of visualizing 50M pixels by composing optical module images generated by 64 XGA commodity projectors. The display provides continuous horizontal parallax with a approximately 45 degrees horizontal field-of-view. The luminance is over 5000 lumen (10000 lumen in high brightness mode) and allows the display to work under almost any kind of ambient lighting conditions.

The rendering library's front-end runs on either Linux or Windows operating systems, and currently implements most features of OpenGL 1.1. The library back-end, which drives the optical modules, is currently running on an array of 16 Linux PCs.

It is obviously impossible to fully convey the impression provided by the display on paper or video. As a simple illustration of the display capabilities, figure 3 presents photographs that were taken from different points of view in front of the large scale display. An accompanying video shows sequences of scenes recorded live using a moving camera.

Conclusions and future work. Our display allows freely moving naked eye participants to share a three dimensional scene with fully continuous, observer independent, parallax. The image quality of our prototype is comparable to nowadays projector-based Geowall displays, with the additional advantages of not requiring users to wear any kind of viewing device. The display looks like an ideal solution for high end multi-user applications. We are currently working on exploiting it for large scale model visualization.

Acknowledgments. This research is partially supported by the COHERENT project (EU-FP6-510166), funded under the European FP6/IST program.

REFERENCES

[1] Tibor Balogh, Tamás Forgács, Tibor Agocs, Olivier Balet, Eric Bouvier, Fabio Bettio, Enrico Gobbetti, and Gianluigi Zanetti. A scalable hardware and software system for the holographic display of interactive graphics applications. In *EUROGRAPHICS 2005 Short Papers Proceedings*, Conference Held in Dublin, Ireland, August 2005, 2005.

Xphere: A PC Cluster based Hemispherical Display System

Dongsik Jo[*] Hyun Kang[†] Gun A. Lee[‡] Wookho Son[⨍]

Digital Content Research Division, Electronics and Telecommunications Research Institute

ABSTRACT

Visual information has the greatest effect among five senses of human. Therefore, the satisfaction of visual information for representing virtual environments is necessary for good results in information acquisition, virtual training, virtual prototyping, etc. Recently, although there have been a large variety of display systems, there is no such a fully immersive and clearly projected display with high-resolution. In this paper, we describe a hemispherical display which supports a fully immersive experience and high-resolution images. In our display system for generating high-resolution images, a virtual scene image is divided into several pieces those are rendered by a PC cluster and projected with multiple projectors. In this paper, we also describe the PC cluster and projectors designed for optimized performance and a convenient control.

CR Categories: H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.
Keywords: Virtual realities, distributed system, process control, real time

1 INTRODUCTION

In virtual environments, visual displays play key roles. Therefore, many users have made requests for an innovative display technology such as head mounted displays and large projection displays (e.g., CAVE [1]). Immersive displays needs to support not only large-scale images but also high-resolution images to maximize visualization effects for various applications (e.g., a design review of a car interior or exterior, a virtual training and simulation of work arrangements). There are many cases of development of immersive display systems for user's needs, but these systems still lack of full immersion, stereoscopy, high-resolution and scalability. In this paper, we propose an immersive display with a screen in spherical shape, which is called "Xphere" (eXtensible Platform for spHErical REndering). We use a back-projection hemispherical screen with multiple projectors for visualizing a high-resolution seamless image in wide field of view. Also, we construct a PC cluster for processing distributed photorealistic rendering and for generating real-time visualization of 3D scene.

2 RELATED WORK

A variety of display systems for presenting virtual environments have been developed from wearing device such as HMD to fully immersive display system such as CAVE [1]. Tiled display from Fraunhofer IGD is a tiled-screen for supporting not only a large-scale image but high-resolution images [3]. A single seamless high-resolution image is composed of multiple images using tiling technique. The Vision station from Elumens is the most well known display product for its screen with spherical surface [7]. The screen has 1.5m diameter providing 160 degrees of field of view. In comparison, our Xphere differs from these with completely acceptable next characteristics:

☐Provides higher resolutions
☐Immersive VR environment
☐Covers full FOV(filed of view) of humans
☐Projection image occlusion avoided by back projection
☐3D stereoscopic view: less ghost effects
☐Seamless images: continuous projection surface
☐Scalable system for generating higher resolution
☐Supports multiple participants
☐Supports convenient control

3 IMPLEMENTATION

3.1 Multi-projection based hemispherical display

The Xphere have consists of three main hardware components: a curved screen of hemispherical shape, a set of multiple projectors, and a PC cluster. First, the screen has a shape of a partial sphere with 3m diameter horizontally and with 2m height vertically. Second, we used multiple projectors for generating high-resolution images with 16 projectors. And, we arranged multiple projectors to project 4x2 tiled images on the rear-projection hemispherical screen surface with stereoscopic view (figure 1). But, because geometrical distortions are unavoidable when projecting images on a non-planar surface with projectors, we have a process for calibrating 3D position and orientation of each projector relative to the screen before rendering the virtual scene in order to correct these distortions (figure 2).

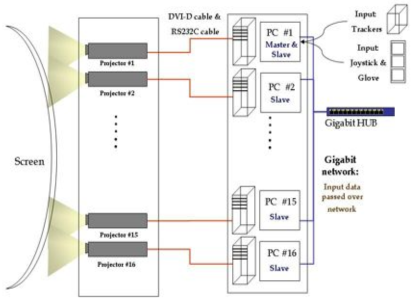


Figure 1: Xphere's configuration

---

[*] e-mail: dongsik@etri.re.kr
[†] e-mail: hkang@etri.re.kr
[‡] e-mail: endovert@etri.re.kr
[⨍] e-mail: whson@etri.re.kr

Figure 2: Geometry and color correction on spherical surface: The left picture shows four projection images overlapped on the hemispherical screen surface. In the middle, geometrical distortions are corrected by projective texture rendering technique [4] which first renders the virtual scene into a texture image with normal rendering pipeline, and then renders the final image by projecting the rendered texture image onto a geometrical surface. The right picture shows the seamless image produced by color correction. We use alpha blending technique by blending a color map image with the rendered image so that two images would blend smoothly.

Finally, a PC cluster for distributing the visual output of standard application is established. Each cluster nodes are connected through a gigabit Ethernet network for high speed data communication and synchronization.

The installed Xphere systems provide users with 4096 x 1536 resolutions including stereoscopy and 180° x 90° of field of view. These cover the whole sight of the user and are adequate to full capacity of visual information (figure 3).

Figure 3: Xphere in use for reviewing a car interior design

3.2 The design of PC cluster and performance test

The Xphere's PC cluster for executing 3D visual contents visualizes high-resolution images through a real-time process. Generally, we can classify the design of PC cluster by transmission differences of data structures between master and slave.

Distribute geometry is that primitives partitioned and sent to slaves. On the other hand, distribute control is that each slave node runs full application and master synchronizes running application [6]. We compared each method using a same polygonal data under same platforms and was able to aware that the rendering speed of distribute control is faster than distribute geometry (figure 4). This is due to store of 3D model data on all slave node and exchange of only synchronization information through the configured network.

Therefore, the 3D graphics distribution software for visualizing Xphere's contents is based on an open source library, OpenSG [5], which has been selected by considering rendering quality and performance.

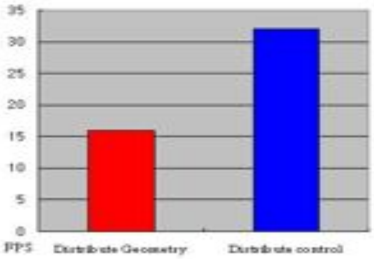


Figure 4: This picture shows the result of rendering speed (frame per second) between distribute geometry (red bar) and distribute control (blue bar). Distribute control has double than distribute geometry. As a software platform, we use OpenSG [5] library for distribute control and Chromium (http://chromium.sourceforge.net/) for distribute geometry on Microsoft Windows XP operating system.

3.3 Centralized control of PC Cluster and multiple projectors

The scalable display wall systems (e.g., tiled display, CAVE) such as Xphere consist of multiple resources including PC cluster, multiple projectors, software and user input. Therefore, these systems need to have management toolkits which provide easy and fast operations [2]. For this purpose, we developed the Xphere's control software for PC cluster and multiple projectors. Provided functions include program execution at the same time, selection of various contents, rebooting and shutdown of cluster nodes. The control software also provides control of multiple projectors such as power on/off and selection of the input source (figure 5).

Figure 5: Xphere's control software: this picture shows a functionality of exchanging the command and executing the program simultaneously for all cluster nodes.

REFERENCE

[1] C. Cruz-Neira, D. J. Sadin, and T. A. Defanti, "Surround-screen projection-based virtual reality: the design and implementation of the CAVE", in Proc. SIGGRAPH '93 (1993), 135-142.
[2] Chen, H., Wallace, G., Gupta, A., Li, K., Funkhouser, T., and Cook, P. "Experiences with scalability of display walls", Proc. IPT/VR 2002 (2002).
[3] HEyeWall. http://www.heyewall.com/
[4] M. Segal, C. Korobkin, R. van Widenfelt, J. Foran, and P. Haeberli, "Fast Shadows and Lighting Effects Using Texture Mapping", Computer Graphics 26-2 (1992), 249-252
[5] OpenSG, http://www.opensg.org
[6] Paul Rajlich. Building and using tiled display walls.
[7] TOPDOME. http://www.elumens.com/

The HYPERREAL Design System

Masaru Hisada* Osaka University
Kazuhiko Takase* Osaka University
Keiko Yamamoto* Osaka University
Ichiroh Kanaya* Osaka University, PRESTO, JST
Kosuke Sato† Osaka University

ABSTRACT

This paper presents a novel mixed reality (MR) system for virtually modifying (e.g., denting, engraving, swelling) shape of real objects by using projection of computer-generated shade. Users of this system, which we call HYPERREAL, perceive as if the real object is actually being deformed when they operate the system to modify the shape of the object while only the illumination pattern of the real object is changed. The authors are aiming to apply this technology to product designing field: designers would be able to evaluate and modify form of their product more efficiently and effectively in an intuitive manner using HYPERREAL than conventional design process (typically, computer aided design, or CAD, systems and solid mock-ups) since the system is able to provide users with actuality/presence of real mock-up and flexibility of shape data on a computer system, such as CAD system, all at once.

CR Categories: J.6 [Computer-Aided Engineering]: Computer-Aided Design (CAD)

Keywords: mixed reality, form design, visual perception

1 INTRODUCTION

This paper aims a new design supporting system by virtually controlling the shape of the real object without damaging 3-D presence and realistic touches of the object.

While computer aided design (CAD) systems are widely used for product design and development currently, designers often use real object like clay model or mock-up of the final product (e.g. automobile) when they evaluate appearance of 3-D surface of the product. Use of fully designed mock-up has been seemed unavoidable since it is hard for us to get a sensibility of realistic touches of the object from object's model rendered by computer graphics (CG) especially on a 2-D screen. However, we cannot control shape of the real object freely unlike we can on a CAD systems. Thus in design process designers must copy the shape of the first mock-up into a CAD model, after modifying its shape on the CAD, designers must rebuild its mock-up by kinds of 3-D printer to evaluate impression of its design. This process goes over and over until the final shape of the product is fixed. The design process using the real object costs long time and disturbs creative activities of designers. The authors aim to solve these problems using recent development of mixed reality (MR) technology.

There are several MR systems superimposing virtual texture image onto the real object using the light projection that gives users high spatial and stereoscopic effect watching the real object [1-5]. In these systems and other related researches, modifying the object shape has not been considered so that the real objects need to have exactly the same shape with virtual objects. The authors propose a projector-based MR system which can apply to any shape. We call this HYPERREAL design system.

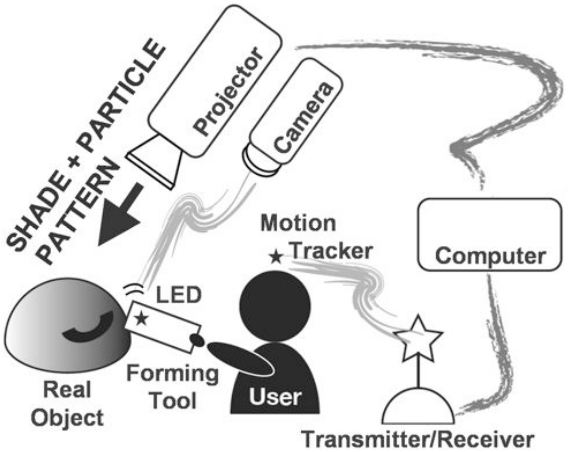


Figure 1: The system overview. A real object onto which computer-generated shade and motion particles are projected provides user with illusion as if the object is actually deformed. The computer keeps tracking the position of the user and the motion of the forming device and regenerates the projection image so as to reflect the deformation done by the user with coherence of user's vision retained.

2 METHOD

In this paper, the authors put emphasis on solution to the problem of how to display the illuminated object with 3-D presence and realistic touches. Projector-based MR systems allow us to easily modify the appearance of object's surface as we like by illuminating them with rendered images. On the contrary, the shape of the object itself seems, indeed, not modifiable.

Though we can perceive depth information roughly with binocular stereopsis, strict depth information is perceived by a surface texture such as shade and shadow, or brightness and darkness [6]. The authors denote these factors of depth perception psychological factor. For example, we cannot perceive bumpy surface of the golf ball without psychological factor on the surface of the ball. Shade and shadow on the surface does create an effect of bump on the surface of the golf ball. With psychological factor, it is possible to express high-definition spatial effect by using CG technique such as shading and shadowing. The authors realize virtual forming of real object by illuminating real object with psychologically adjusted CG images.

---

*email: {his, takase, kei, kanaya}@sens.sys.es.osaka-u.ac.jp
†email: sato@sys.es.osaka-u.es.jp

So as to enhance the effectiveness of shading and shadowing, the authors also explore the possibility of using particle rendering combined with conventional surface rendering for HYPERREAL system. Motion of particles has been used to illustrate specific information on shape of the static 3-D object intuitively and effectively on the 2-D display [7]. However, the use of motion as a mean of changing shape of the real object has been a rather unexplored area. In this paper we present how the flow of moving particles coupled with shading and shadowing along the object's surface can be used to provide the user with supplemental perception of object's shape and structure.

It is important to note that virtual forming creates the appearance of the surface, but does not affect the shape of the surface itself. The shape of the real object retains its original form. We must consider the position of the user's head every time the user moved his/her head, for creating the appropriate rendered images, since the shade and shadow move on the surface as the viewpoint changes.

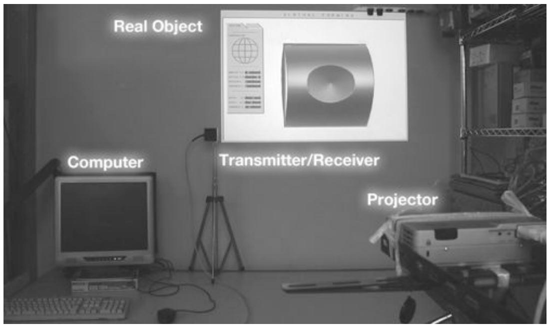


Figure 2: Outlook of the HYPERREAL design system

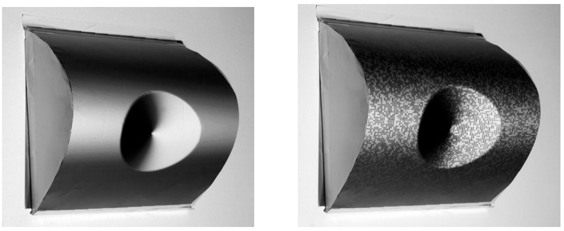
3 HYPERREAL – A VIRTUAL FORMING SYSTEM

The overview of the HYPERREAL design system is shown in Figure 1. The outlook of the experimental setup is shown in Figure 2. This system runs on a PC workstation (Intel Pentium-4). A six-dof position tracker (Polhemus Fastrack) is used to detect the position of the user's head. The authors use a projector (Plus V1100) to render images directly on white (natural, unmodified) object used as a screen.

Virtual forming process is done as follows. First, we set a neutral object of known shape within the range of projector's illumination. Before virtual forming, the 3-D shape model of the real object is captured on the computer. Changing the shape of the object, or virtual forming, is done in the virtual space on the computer. Then, the HYPERREAL system simulates appearance of the object from a certain viewpoint on the virtual space using the CG technique to produce the appearance of the object as it ought to be after virtual forming. The photometric pattern obtained from the simulation is used to illuminate the neutral object and give an illusion of the different surface shape. These sequences of processes go on in real-time.

Users can modify the surface of the object by directly touching the surface with special hand piece whose position is tracked by the HYPERREAL system. The system computes corresponding appearance of the object as the user intended and replaces the pattern of the illumination so that the user gets an illusion as if the object has actually been deformed.

Figure 3 shows examples of virtual forming. Figure 4 shows actually dented surface (for comparison) and virtually dented surface.



(a) (b)

Figure 3: (a) Object with virtual-shade projection
(b) Object with virtual-shade plus particles

Figure 4: Actually dented surface (left)
and virtually dented surface (right).

4 CONCLUSION

In this paper, the authors have presented a new design supporting system applying MR technology. This system has enabled to control the surface shape of real object without actually modifying it. While more work is needed, the authors also demonstrate the effectiveness of motion rendering combined with shading and shadowing as a mean of enhancing shape perception. The authors currently are developing user-friendly device to modify shape of objects in more intuitive way. The authors expect this system helps the designers to design 3-D forms intuitively and speedily without disturbing their creativity and imagination.

REFERERNCE

[1] O. Bimber, R. Rasker: *Spatial Augmented Reality – Merging Real and Virtual Worlds;* A K Peters, Ltd., 2005.

[2] R. Raskar, G. Welch, W. C. Chen: Table-Top Spatially-Augmented Reality: Bringing Physical Models to Life with Projected Imagery; *Proc. 2nd Int'l Workshop on Augmented Reality,* pp. 64-71, 1999.

[3] D. Bandyopadhyay, R. Raskar, H. Fuchs: Dynamic Shader Lamps: Painting on Movable Object; *Proc. IEEE and ACM International Symposium on Augmented Reality,* pp. 207-216, 2001.

[4] R. Raskar, R. Ziegler, T. Willwacher: Cartoon Dioramas in Motion; *Proc. Int'l Symposium on Non-photorealistic Animation and Rndering (NPAR),* 2002.

[5] K. Sato, Y. Kusumoto, T. Uchida: Real Objects with Virtual Texture on An MR Table Projection Based Mixed Reality; *Proc. IEEE VR Research Demonstration,* 2001.

[6] James J. Gibson: *The Ecological Approach to Visual Perception;* Houghton Mifflin Company, 1979.

[7] E.B. Lum, A. Stompel, K. L. Ma: Using Motion to Illustrate Static 3D Shape – Kinetic Visualization; *IEEE Trans. on Visualization And Computer Graphics,* Vol. 9, No. 2, pp. 115-126, 2003.

Long Visualization Depth Autostereoscopic Display using Light Field Rendering based Integral Videography

Hongen Liao, Keisuke Nomura, Takeyoshi Dohi

The University of Tokyo

ABSTRACT

We developed an autostereoscopic display for distant viewing of a three-dimensional (3D) computer graphics (CG) image without using special viewing glasses or tracking devices. The images are created by employing a light field rendering and pixel distribution algorithm for integral photography (IP) / integral videography (IV) imaging. How to enhance the image depth, especial the long visualization depth, is a challenge work. In this study, the images are rendered from a referential viewing area for each viewpoint and the elemental images are reconstructed by pixel redistribution method. The corresponding result images are projected to a screen that is separated from the lens array by a referential viewing distance as the setting of image rendering. A photographic film is used to record the elemental image through each lens with a photograph-taken method. Our photograph-taken based IV display enables precise 3D images with long visualization depth to be displayed at long viewing distances without any influence from deviated or distorted lenses in a lens array. We created an autostereoscopic display that appears to have three-dimensionality even when viewed from a distance, with an image depth of 5.7 m or more in front of the display. To the best of our knowledge, the presented long-distance IV display is technically unique as it is the first report of generating an autostereoscopic image with such a long viewing distance in the field of computer graphics.

CR Categories and Subject Descriptors: B.4.2 Input/Output Devices – Image Display

Keywords: autostereoscopic display, integral photography, integral videography, long visualization depth

1 INTRODUCTION

Computer graphics (CG) is the field of visual computing, which utilizes computers both to generate visual images synthetically and to integrate or alter visual and spatial information sampled from the real world. The images can be photographs, drawings, movies, or simulations -- pictures of things which do not yet exist and maybe could never exist. Images from places we cannot see directly, such as medical images from inside body, can also be reproduced by CG. Most of CG-generated three-dimensional (3D) image are displayed on two-dimensional (2D) screen by use of 3D object rendering algorithm, such as shading, lighting, mapping, color creating, ray tracing at al. The motion parallax and the image depth can not be reproduced when the 3D image displayed on 2D screen.

Traditional stereoscopic display consists of creating a 3-D illusion starting from a pair of 2-D images. It creates depth perception in the brain by providing the eyes of the viewer two different images, representing two perspectives of the same object, with a minor deviation similar to the perspectives that both eyes naturally receive in binocular vision.

The binocular stereoscopic technique in combination with the CG has been popularly applied in field of virtual reality. Most of that research focused on the development of stereoscopic display multiple perspective views inherently requires a very high display resolution. The binocular stereoscopic display reproduces the depth of projected objects by using fixed binoculars; because the images for the left and right eyes are formed separately, there is a disparity in the reproduced image. Therefore, different viewers can have inconsistent depth perception.

The use of autostereoscopic display for CG displaying is a technological improvement because it augments the image depth for 3D display without use of any supplementary glasses or tracking devices. Integral photography (IP) [1] is an ideal way to display 3D autostereoscopic images—spatially formatted visible images can be formed from an arbitrary viewpoint. This autostereoscopic technique is also called “fly's eye lens photography” because of its use of an array of tiny lenses and a film for capturing and displaying images. A number of elemental images are recorded on the film. When the film is placed at the same position relative to the lens array and is irradiated with a diffused light, the rays retrace the original routes and reproduce the image at the same position as the located object. IP uses both horizontally and vertically varying directional information and thus produce a full parallax image, making it a promising method for creating 3D autostereoscopic displays. Integral videography (IV) is an animated extension of the IP.

Most of the reported studies have focused on widening the viewing angle and improving the viewing resolution of IP/IV image. These devices only have an image depth of several centimeters to several decade centimeters. To the best of our knowledge, there has been no report about producing a 3D autostereoscopic image with a depth of several meters for viewing with the naked eye.

In this paper, we present a 3D autostereoscopic display for distant viewing of a 3D CG image with an image depth of several meters in front of and behind the display. The images are calculated from a referential viewing area for each viewpoint and the elemental images are reconstructed by pixel redistribution. We proposed a photograph-taken method for recording the corresponding result images to a photographic film through each lens. The images displayed have three-dimensionality even when viewed from a distance.

2 LIGHT FIELD RENDERING BASED IV IMAGE GENERATION

2.1 Referential Viewing Area and Light Field Rendering

A referential viewing area based elemental images rendering technique is proposed to increase the depth of the IV image. The setting of the referential viewing area will be also used in recording the IV elemental images for each lens. The images can be created by CG software or captured by cameras from different viewing directions corresponding to the divided referential

---

7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
E-mail: liao@miki.pe.u-tokyo.ac.jp

viewing area (Fig. 1). The viewing distance and the setting for image rendering is the same distance used in recording and viewing the IV image. Since the image rendering and recording of our 3D imaging system is based on a standard viewing distance and viewing area, the developed autostereoscopic display could be free from the distortion of the lens array and the image, as will be described in section 2.5.

As a method of rendering the images for each viewpoint, light-field rendering could be used to render arbitrary view of a static 3D scene [2] [3]. Light fields were designed as a method for rendering a fast 3D image. We can simulate the function of IV display with the process of reproducing light rays by CG.

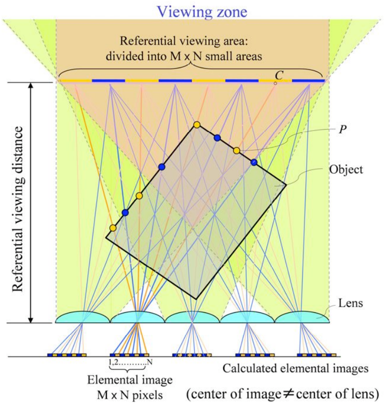


Figure 1. IV elemental image generation using referential viewing area based ray tracking algorithm.

2.2 Computer Graphics based Image Generation

The elemental IV image is one of rendered result of special CG technique. Any kind of CG rendering method can be used for IV image generation. Figure 1 shows the computing principle of 3D images for convex lens of a lens array in a 3D display device. We set the referential viewing distance as the distance from the location of the elemental image behind the convex lenses to a plane (a referential viewing area) located at the position where combined elemental images can be viewed as a 3D IV image.

The quality of IV image will be improved when the resolution of the rendered image is increased. We can synthesize a high quality image by interpolation of the contiguous elemental images [4]. The viewing area is decided by the optical characteristic of the lens and the corresponding position of the covered elemental image.

The elemental images are calculated according to the following procedure. First, the referential viewing area is divided into a 2-D M by N matrix of small areas in accordance with the resolution of the 3D image to be reproduced. For each small area, the image viewed from each viewpoint is rendered by CG software. The point at which the ray becomes incident with the object is determined to be a surface point of the object. Brightness and color are calculated based on surface color of the object and the lighting conditions, as if the object were viewed from the center of the small area. This procedure is repeated for all M by N areas.

The viewing direction of virtual camera from each viewing point is as the line connected the viewpoint and the center of the screen. The light filed of the image taken by virtual camera covers all of area of the lenses. Since the center of the screen will not be changed when we render the image, the 3D object can be moved forward of backward with different image depth.

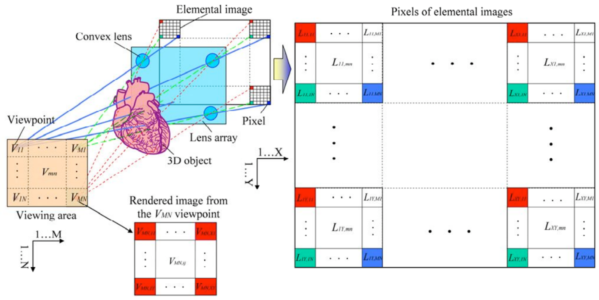


Figure 2. CG based image generation for each viewpoint and elemental image corresponding to each lens.

2.3 Pixel Distribution for Elemental Image

The principle of CG-based image creating method for elemental IV image is shown in Fig. 2. For each small viewing area, one image is taken. The resolution of the image should be greater then the total number of lenses. To achieve a high-quality IV image, we first rendered the image with a 2×2 times pixels number of the total lenses number. Corresponding pixels are extracted and synthesized to create IV elemental images. The rendered M×N images are distributed to the IV elemental images. Each elemental image is saved as a single image file in PC.

The pixel distribution calculation describes how to distribute the pixel for each elemental image as follows. We set the coordinate of X axis as horizontal and Y axis as vertical. The definition of viewing point and the lens number are shown below.

$ViewPoint(m, n)$; ($m = 1, 2, \ldots, M$; $n = 1, 2, \ldots, N$).

$Lens(i, j)$; ($i = 1, 2, \ldots, X$; $j = 1, 2, \ldots, Y$).

The M and the N are the numbers of divided viewing area in horizontal direction and vertical direction. The viewing area number is the same as the pixel number of the elemental image corresponding to each lens. The X and the Y are the total lens number of lens array in horizontal and vertical direction. The plane image of the object obtained with *ViewPoint(m,n)* is definite as $V_{mn}[X][Y]$. The elemental image behind the *Lens(i,j)* is $L_{ij}[M][N]$

Consider of the possible pixel viewed from the viewpoint is recorded as the intersection of the screen and the straight line that connects the center of lens and the viewpoint. The pixel of $V_{mn,}[i][j]$ is corresponding to the position of line $[i]$ and low $[j]$ behind the lens as shown in Fig.4. The entire pixel V_{ji} in the rendered image is distributed to the corresponding lenses as below.

$$V_{mn,}[i][j] = L_{ij,}[M-m+1][N-n+1]$$

By repeating this processing about the image V_{mn} obtained with all viewpoint (m, n), the whole elemental images can be calculated. This method also solved the problem of two-step IP proposed by Ives [5] using computer generated IP.

2.4 Long Visualization Depth Autostereoscopic Display

To reproduce a 3D image precisely by utilizing a conventional lens array, the lens arrangement and the geometry of each lens should be absolutely flawless. In practice, it is very difficult to obtain a flawless lens array. We succeeded in creating an image display that appears to have three-dimensionality even when viewed from a distance. The present method solves the above-mentioned problems, thereby letting a display device present virtually deviation-free and distortion-free 3D image. Perfect 3D images are obtained by minimizing the deviations of light rays in the observing areas, even if some lenses in the array are deviated and/or distorted.

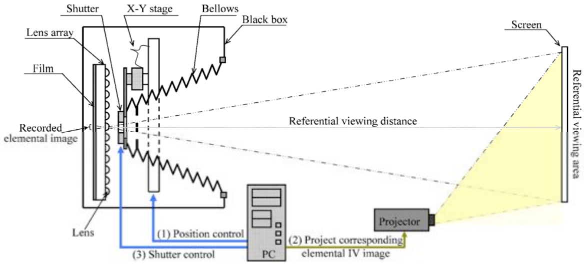


Figure 3. Configuration of IV image generating device.

Figure 3 shows a configuration of the 3-D image generating device. The device comprises a lens array having a predetermined number of convex lenses arranged in a matrix and a photographic film arranged behind the lens array almost on the focal plane. The lens array and the film are placed in a black box. The 3-D image generating device also includes an X-Y stage shutter moving device holding a shutter and a bellows. Except for the lens directly behind the shutter, the other remaining convex lenses are covered by the black box and the bellows so that portions of the film corresponding to the remaining lenses are not exposed. The X-Y stage and the shutter are connected to a computer. A projector is connected to the computer so that the calculated elemental images can be projected on the screen. The screen is separated from the lens array by the referential viewing distance of the 3-D image generating device [6].

Elemental images on the film are generated according to the following steps. The photographic film is placed near to the focal plane behind the lens array frame. We control the X-Y stage and move the shutter to the position in front of the desired convex lens. The corresponding image to the selected lens is read from the PC memory and is projected on the screen. The projected images and the exposure time are automatically controlled by the computer. In this procedure, only a portion of the recording means behind the desired convex lens is exposed by opening the shutter. Except for the desired lens, the other remaining convex lenses are covered by the black box and the bellows so that portions of the recording means corresponding to the remaining lenses are not exposed.

2.5 Long Visualization Depth Autostereoscopic Display

The same lens array used for generating images is also used for displaying the generated images. Despite that the generated image will deviate from the intended position if the lenses are deviated, the displayed image will be reproduced at virtually the right position on the referential viewing plane, since the generated deviated image is compensated by the same deviated lenses, which returns the deviated light rays to designed positions in the referential viewing area. Since the light rays are scarcely deviated in the vicinity of the referential viewing area, the presented method can reproduce a 3D image with the desired position and geometrical shape.

3 LONG VISUALIZATION DEPTH IV DISPLAY

3.1 Display and lens array

The lens array is composed of 105×70 lenslets. Each lenslet element was made of glass and has a uniform base size of phi 6 mm. The focal length of each lenslet is 13 mm. Taking into account the spherical aberration, diffraction, the focal length error of the lenslet, the iris diaphragm of phi 3 mm was chosen and the lens array was arranged in the shape of hexagon with a lens pitch of 10 mm. The lens array frame is made of aluminum with high precision manufacture technique. The lens array is covered by an iris array.

We used a positive glossy film (Konica, QAC7-TDO) to record the image in these experiments. The resolution of film was measured by using a test chart with a set of white and red binary stripes. The results show that the resolution of the film is over 65cycle/mm (pixel pitch: 0.00769mm), which is satisfied with the requirement of recorded IV elemental image (pixel pitch of 0.0127mm, 3.6mm / 284 viewpoint, where 3.6 mm is the size of each elemental image). Each lens covered 280×210 pixels, the elemental images were calculated by using the referential viewing area based light field rendering algorithm.

3.2 CG-based images and distributed elemental image

We evaluated the image depth using 3D objects that were placed at different depths in front of and behind the lens array. The structure of displayed objects include a club with 2 m in front of the screen and 4 m behind the screen, a ball at 1 m in front the screen, a cube at 0.5 m behind the screen, letter "IP" 0.5 m in front of the screen, "ATRE" at the lens display point and "3D" 1.5 m behind the screen (Fig.4). The object was rendered from different viewing point by use of CG software (Shade, Expression Tools Inc.). The lens array is combined by 105 (H) × 70 (V) lenses. Each lens covers an elemental image with a resolution of 280×210 pixels. Some of selected distributed images are shown in Fig.5.

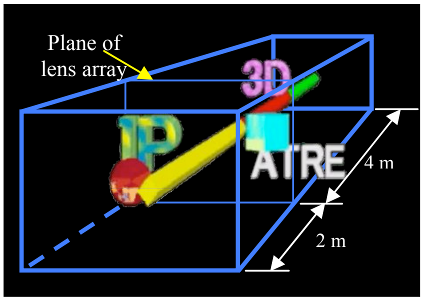


Figure 4. 3D model structure of displayed image.

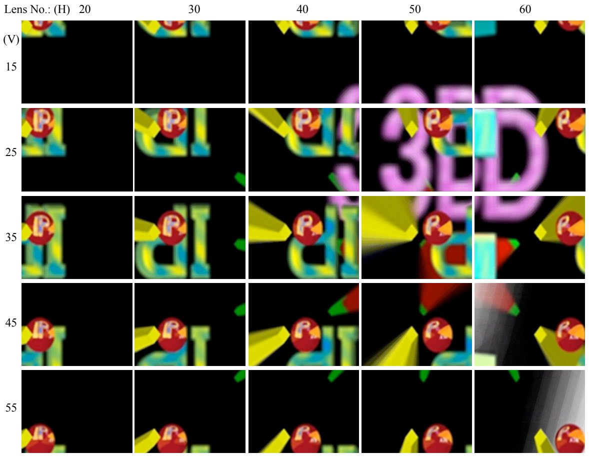


Figure 5. Selected reconstructed elemental images corresponding to the lenses at the position of line 15, 25, 35, 45, 55 and law 20, 30, 40, 50, 60.

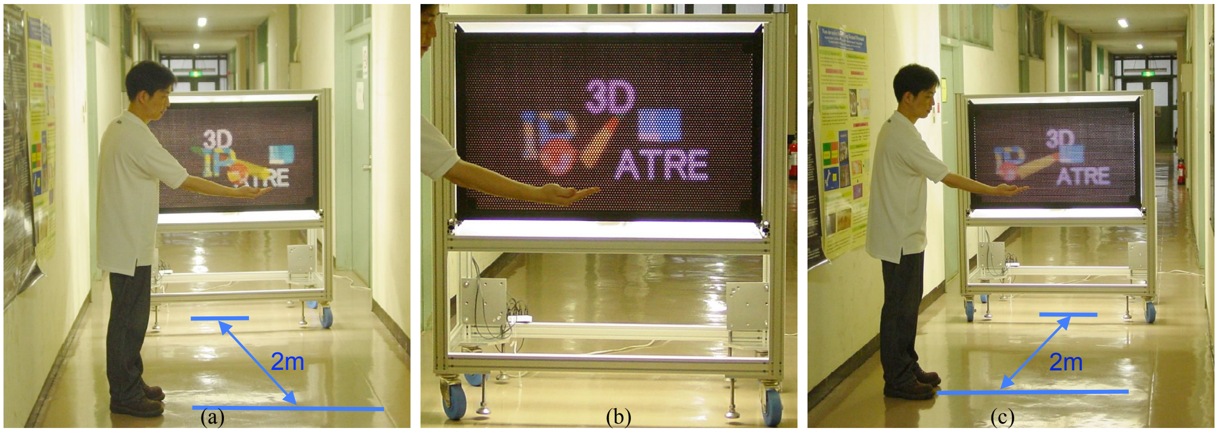


Figure 6. A 3D computer graphics image produced by the IV autostereoscopic display for distant viewing, viewed from the left to right. The club stands out in mid-air on the palm of the viewer's hand about two meters away (yellow) and four meters inside the display.

3.3 Manufactured long distance IV display

The manufactured long visualization depth IV display is showed in Fig.6b. At a viewing distance of five meter, the yellow part of club jumps out about two meter from the screen and the green part of club appears about four meters inside the screen (Fig.6). The viewer can move up or down, right or left, and the image will track as fixed in the position of the hand in the figure shows (Fig.6a, 6c); the club appears to be directly on the palm. The projected IV images were taken using a digital camera (Nikon D1X). The focal length and the F-number of camera lens were 50 mm and 16, respectively. The pupil diameter of the camera iris was about 3 mm, which is similar to the pupil diameter of the human eye in an ordinary lighted indoor environment. The viewing angle was 13.6 ° in horizontal and 10.2 ° in vertical.

The motion parallax of the projected IV image was continuously observed. The image could be observed clearly from different directions. The palm was placed two meter in front of the display and was adjusted to the position under the yellow club. The reconstructed IV images (in this case club) are on the same position their relative position does not change from any viewing angle.

3.4 Manufactured long distance IV display

We also evaluated the maximum spatial resolution of the displayed images by using a set of black and white strips. The IV images of the projected stripes were taken using a digital camera. The focal length and the F-number of camera lens were 50 mm and 16, respectively. The pupil diameter of the camera iris was about 3 mm, which is similar to the pupil diameter of the human eye in an ordinary lighted indoor environment. The resolution was determined by the minimum width at which the stripes could be clearly observed. Experiments were conducted in which five observers each made five viewing attempts and these revealed that with this apparatus, images could be displayed in a range from 5.7 meters in front of the screen to 3.5 meters into the screen.

4 SUMMARY AND FUTURE WORKS

Our 3D images obtained by the image-generating device were designed to be observed from an apparent area in the vicinity of the referential viewing distance, the positional errors of the light rays in the apparent area were so small that no deviations in the reproduced images could be perceived.

Since our algorithm generates the IV elemental image for each lens using pixel distribution method, it is not necessary to perform special hidden surface processing in IV image generation. This method addressed the issues of IV/IP image rendering and made it free from the pseudoscopic image problems. Furthermore, the common CG rendering method like translucence, reflection, peculiar surface processing can be directly used in IV image generation.

We proposed a CG-based IV image generation method for high quality IV image generation. Since the images are calculated based on respective points in the referential viewing area, light rays in the vicinity of the referential viewing area are calculated with high precision. The display can be viewed by multiple peoples from a distance of 5 m or more. The image depth is over 5.7 m in front of the display that could be observed by multiple peoples.

Our future works include developing a dynamic long visualization depth IV display and improving the image quality of the reproduced IV image. The CG based IV rendering algorithm should be improved and a more computational power should be enforced in the future work.

The work was supported in part by Grant-in-Aid for Scientific Research of the MEXT, Japan (17680037, 17100008) and Terumo Life Science Foundation.

REFERERNCE

[1] M.G. Lippmann, “preuves reversibles donnant la sensation du relief, ” J. de Phys., vol. 7, pp.821-825, 1908.

[2] M. Levoy and P. Hanrahan, “Light field rendering,” in SIGGRAPH’96, pp.31-42, 1996.

[3] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, “The Lumigraph,” in SIGGRAPH’ 96, pp. 43-54, 1996.

[4] T. Naemura, T. Yoshida and H. Harashima, “3-D computer graphics based on integral photography,” Opt. Exp., vol. 8, no. 2, pp.255-262, 2001.

[5] H.E. Ives, “Optical properties of a lippmann lenticulated sheet,” J. O. S. A. Vol.21, pp170-176, 1931.

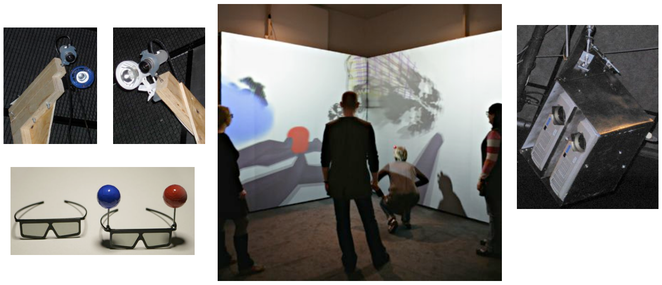
[6] H. Liao, M. Iwahara, Y. Katayama, N. Hata, T. Dohi: 3-D from a Distance using Integral Photography, Opt. Let., Vol.30 No.6, pp.613-615, 2005

10:30–12:00 Session II

Upponurkka: An Inexpensive Immersive Display for Public VR Installations

Tapio Lokki* Tommi Ilmonen† Wille Mäkelä‡ Tapio Takala§

Helsinki University of Technology
Telecommunications Software and Multimedia Laboratory



Abstract

Upponurkka is an inexpensive immersive display system for public virtual reality installations. It is designed for public places to be robust, wireless, and light-weight. The system consists of two passive stereo screens and optical tracking. In this paper the hardware and software of Upponurkka are briefly presented. In addition, experiences of using the system in a public art exhibition are reported.

CR Categories: B.4.2 [Hardware]: Input/Output Devices—Image display; J.5 [Computer Applications]: Arts and Humanities—Fine arts

Keywords: virtual reality, immersive display, VR for public

1 Introduction

Upponurkka (literally "a submersible corner" in English) is a low-cost immersive display, especially designed for public VR installations. Like a cave, it is a multi-user display, although only one viewer is tracked so that he/she sees the correct perspective. Special attention in design was paid on the robustness of the system. As a starting point we made a policy decision that from the user perspective the system should be wireless and light-weight.

The motivation for Upponurkka was to build a cost-effective 3D display which could be used in an exposition of immersive art in Kiasma, Museum of Contemporary Art in Helsinki, Finland. The presented immersive art is created by drawing into the air 3D images with free-hand style in a cave-like environment [4, 5]. The exhibited artworks are the result of series of studies in which we are seeking an immersive free-hand medium for artists. The immersive art is painted with a special drawing device resembling a data glove, presented in more detail in previous papers [3, 6].

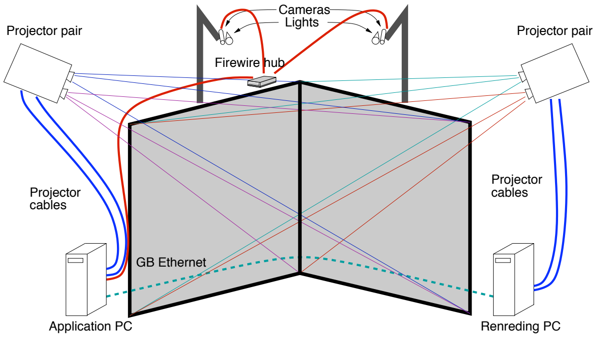


Figure 1: The components of Upponurkka.

2 Equipment

The equipment, both hardware and software, of Upponurkka is completely self-made and constructed from off-the-shelf components. The total cost of all equipment was about 13,000 €.

2.1 Hardware

The schematic illustration of the hardware is depicted in Fig. 1. The projection system consists of four off-the-self LCD projectors equipped with passive polarization filters in front of their lenses, see Fig. 2. In fact, the emitted light from LCD projectors is always polarized. However, in this case the polarization angle was 45 degrees from vertical/horizontal direction and it was practical to rotate the polarizations with the filters to vertical and horizontal direction. Of course, we could have tilted the projectors and manage without the filters (and obtain more luminance), but then the projector rack construction and image adjustment to the screens would have been more complicated. In the current setup, the projectors produce passive stereoscopic images to two 3x3m polarization preserving screens which are mounted in right-angle. The projectors are positioned slightly off-center to minimize reflection of the projected light from one screen to the other.

The images are rendered with two PCs, named the application PC and the rendering PC in Fig. 1. Each computer has a single graphics card with dual outputs and both are responsible for render-

---

*e-mail:Tapio.Lokki@tkk.fi
†e-mail:Tommi.Ilmonen@hiit.fi
‡e-mail:wmakela@tml.hut.fi
§e-mail:Tapio.Takala@tkk.fi

Figure 2: Projectors are mounted to the rack that is hanging from the ceiling.

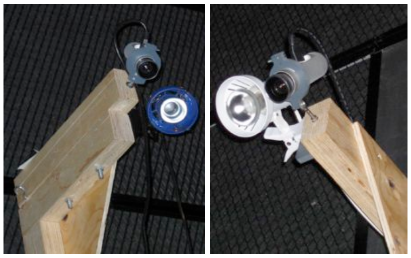


Figure 3: Cameras for motion tracking. To work properly cameras need extra light, produced by ordinary spotlights.

ing 3D content to one pair of projectors. The PCs have integrated 1 GB Ethernet controllers that take care of the communication between PCs.

The tracking of the head of the user is based on custom made optical tracking. The tracking hardware consists of two Apple iSight FireWire cameras that are equipped with wide-angle converters. They are connected to the application PC via the FireWire hub. In addition, for good performance the tracking needs some extra light which is produced with normal spotlights, mounted as close to the cameras as possible, see Fig. 3. The tracked user is wearing passive stereo-glasses which are equipped with color balls that are covered with retroreflective material, see Fig. 4. Such solution is very inexpensive and robust in public environment.

2.2 Software

Upponurkka is run by in-house software on top of the Linux operating system. The graphics to be rendered is distributed to the rendering PC with the Broadcast GL (BGL) [2]. It is an alternative approach for managing graphics clusters. The application using BGL broadcasts binary encoded OpenGL API calls to the rendering slaves over a UDP/IP socket. In Upponurkka the application PC runs the main application and both the application PC and the rendering PC operate as independent rendering slaves. The BGL is very efficient in data-intensive applications that require good scalability to multiple displays.

2.2.1 Optical Tracking

Maybe the most interesting innovation in Upponurkka is a cheap optical tracking system. The software is based on the FLUID camera tracking code [1] and it tracks two color balls with two cameras. The tracking is performed with 30 Hz sampling frequency and it works very well in the viewing area which is about 3x3m in front of the screens.

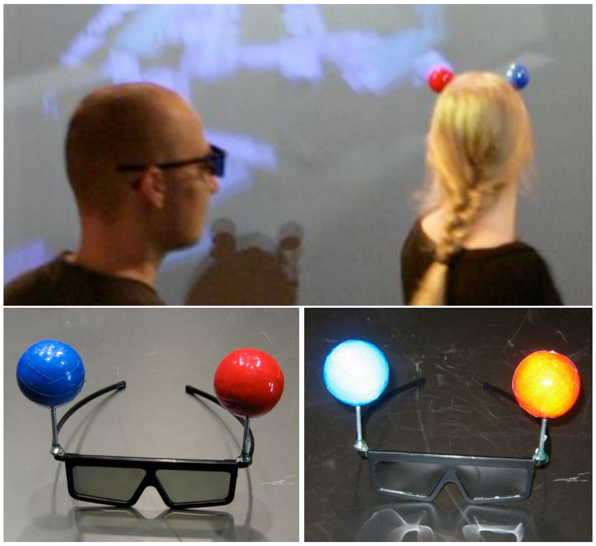


Figure 4: *Top:* Polarizing stereo glasses in use, one equipped with tracking markers. *Bottom:* Blue and red retroreflective markers without (on the left) and with (on the right) spotlight illumination.

For optical tracking we needed markers that would not be confused with other objects in the room. At first we considered using markers with rare color that would seldom appear in clothes of users. In practice such colors do not exist, since the museum visitors may have any kind of clothing. As a next step we tested how a retroreflective material would work. As seen in Fig. 4 retroreflective markers appear very bright when the spotlight illuminates them. Such illumination is easily obtained by mounting spotlights close to the cameras. As a result the markers can be detected based on two features. First the markers have a well-defined color and second the markers are brighter than any normal objects. In exhibition settings in addition to the markers the only bright objects are the ones with glossy reflection that hits the camera (leather handbags etc.). Such reflections are typically white (regardless of the color of the material) and easy to filter out.

The applied Apple iSight FireWire cameras don't, unfortunately, support the control of the color balance. Therefore, color drift occurs as the cameras adjust their color profile automatically. This problem was overcome by using monochromatic markers (red and blue) — a pure red is pure red, at least green and blue values are close to zero, regardless of the color balance of the camera.

The complete motion tracking pipeline is:

1. Capture images.
2. Detect the markers from the images.
3. Calculate the ray from camera to the marker.
4. Calculate the marker locations from the ray information.
5. Calculate head location from the marker locations.
6. Perform Kalman filtering on the location data to obtain smooth input for graphics rendering.

At each stage we encounter noisy data that needs to be dealt with. For this reason the system implements several ad-hoc error correction/recovery methods on stages 2–4.

Table 1: Distribution of exhibition visitors, grouped with time spent watching artworks.

user type	wear glasses	wear tracked glasses	pick a set	main focus: interface	main focus: artworks	spent time	rough % of visitors
A	-	-	-	+	-	10 sec	10
B	+	-	-	+	-	20 sec	10
C	+	+	+	+	-	1 min	10
D	+	+	+	+	-	2-5 min	40
E	+	+	+	-	+	10-20 min	20
F	?	?	?	?	?	?	10

3 USE CASE: IMMERSIVE ART EXHIBITION

Upponurkka was mounted to a separate room in Kiasma, the Museum of Contemporary Art in Helsinki, for a period of three weeks in November 2005. During this time, a few thousand visitors experienced the immersive art exhibition, ”Painted into air”, by Wille Mäkelä and ten guest professionals of traditional fine art. Fig. 5 is taken in the museum and in Fig. 6 a few sketches are presented. The total of 25 artworks were first painted in the cave-like environment, most of it with the first version of the painting software [4, 5]. In Kiasma, the 3D paintings were presented automatically one by one, for about a minute each which is close to a normal museum visitor's time per an art work. The paintings were shown in three separate series, following the steps of the painting tool development. When a set ended, the tracked user could pick a new set by walking against one of the three symbol columns. Besides the picking of a set, the user had no other interaction options but walking in the immersive artworks. With such user interface no action buttons or other interaction devices, which attracts the user to game behavior, were needed. Besides, such assessories tend to break down in public VR installations. Naturally, an user guide and a thumbnail map of the artworks were presented on the side walls.

To find out the approximate behavior of the visitors, we saved log data, such as tracker data, security videos, message interface, and notes. Some people who came to the room when it was empty did not wear either the tracked stereo glasses or the usual ones, just shortly looked at the user interface (stuck columns) on the screen and went out. Some others learned intuitively to use the tracked glasses and saw every single artwork in a row.

Table 1 explains the visitors action in the art exhibition. User types A and B are alone and do not find any support. Type C user has more guts and gets a very interesting and disturbing new experience, withdrawing immediately to smelt it. User in category D is enjoying the interface with an amusement park attitude, walking and diving around, trying to fool the computer with fast runs, very often attracting more people to gather around with normal stereo glasses, even 20 by the same time. Finally, type E user is alone or with a small group, focusing on an aesthetic experience. Approximately, people found it intuitive and very interesting to use the system. Apparently a new user learned to use the system easier if the user came to the room at a moment when somebody was already using the tracked glasses.

The room was mostly without a guarding guide. Still after the three weeks, only five of the 20 polarization glasses were broken, including one pair of target glasses. In spite of little kids jumping against the walls, the silver screens were still in one piece, only a little worn out at the low edge.

3.1 Discussion

Scrutinizing the user types A-F suggests that a better support should be given to types A and B. Upponurkka was located in the isolated space in the museum, therefore an individual did not often see any others to use the system. Thus, tacit knowledge about the use of the installation did not flow in the best way. Next time, a more open exhibition space should be considered. An ideal place might be a corner of a large exhibition hall, letting the spectators to have a gradual approach from the periphery to the stereo glass zone and then to the hot spot of the tracked user. For museum personnel, the open solution would ease the task of guiding and guarding. An important condition would then be that all the other installations in the hall should also use dim lights.

People should have understanding about what harms the system. Basically, written warnings are not the way to make clear security rules. The best way is the tacit knowledge coming from other people, either users or museum personnel. A foot list could though be adjusted to protect the low edge of the screens.

4 CONCLUSION

The inexpensive and robust to use immersive display system for VR applications is presented. Both applied hardware and software are overviewed and user experiences are reported. The presented system has been applied in public art exhibition without technical problems. In the exhibition the system run smoothly, without any interruption, for three weeks and people found out intuitively how to operate with an immersive display and deal with other users. This proves that VR applications can be shown also in public places.

REFERENCES

[1] T. Ilmonen and J. Kontkanen. Software architecture for multimodal user input – FLUID. In *The Proceedings of the 7th ERCIM Workshop*, pages 223–242, Chantilly, France, October 2002.

[2] T. Ilmonen, M. Reunanen, and P. Kontio. Broadcast GL: An alternative method for distributing OpenGL API calls to multiple rendering slaves. *The Journal of WSCG*, 13(2):65–72, 2005.

[3] W. Mäkelä. Working 3D meshes and particles with finger tips: Towards an immersive artists' interface. In *New Directions in 3D User Interfaces Workshop, IEEE Virtual Reality Conference*, pages 77–80, Bonn, Germany, March 12 2005.

[4] W. Mäkelä and T. Ilmonen. Drawing, painting and sculpting in the air: Development studies about an immersive free-hand interface for artists. In *Beyond Wand and Glove Based Interaction Workshop, IEEE Virtual Reality Conference*, pages 89–92, Chicago, IL, USA, March 28 2004.

[5] W. Mäkelä, M. Reunanen, T. Takala, and T. Ilmonen. Possibilities and limitations of immersive free-hand expression: a case study with professional artists. In *ACM Multimedia*, pages 504–507, New York, NY, USA, October 10-16 2004.

[6] M. Reunanen, K. Palovuori, T. Ilmonen, and W. Mäkelä. Affordable fingertip tracking with ultrasound. In *Virtual Environments 2005, Eurographics/ACM SIGGRAPH Symposium Proceedings*, pages 51–58, Aalborg, Denmark, October 6-7 2005.

Figure 5: Upponurkka in action in Kiasma, Museum of Contemporary Art in Helsinki.

Figure 6: A few example sketches painted in the air by Wille Mäkelä.

La Cueva Grande: a 43-Megapixel Immersive System

Curt Canada
Los Alamos National Lab
cvc@lanl.gov

Tim Harrington
Los Alamos National Lab
toh@lanl.gov

Robert Kares
Los Alamos National Lab
rjk@lanl.gov

Dave Modl
Los Alamos National Lab
digem@lanl.gov

Laura Monroe
Los Alamos National Lab
lmonroe@lanl.gov

Steve Stringer
Los Alamos National Lab
stringer@lanl.gov

Bob Tomlinson
Los Alamos National Lab
bob@lanl.gov

ABSTRACT

Los Alamos National Laboratory (LANL) has deployed a 43-megapixel multi-panel immersive environment, La Cueva Grande (LCG), to be used in visualizing the terabytes of data produced by simulations. This paper briefly discusses some of the technical challenges encountered and overcome during the deployment of a 43-million pixel immersive visualization environment.

CR Categories and Subject Descriptors: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism – Virtual Reality; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – Artificial, augmented and virtual reality.

1 INTRODUCTION

Designed by Fakespace Systems and LANL personnel, LCG is the highest resolution stereo immersive facility in the world. In this paper, we discuss the requirements and design of this facility, its implementation, and its software infrastructure.

2 MOTIVATION AND BACKGROUND

Today, ASC scientists have the capability to perform a single production calculation that generates more data than is contained in the Library of Congress print collection. Currently, a large three-dimensional problem may have 500 million cells with tens of variables per cell. A single time-step dump from such a problem is on average about 150 gigabytes. This means that a single time-step dump from such a calculation is about 100 times larger than the complete time sequence saved from a typical 2-dimensional production run done in 1997. Usually, 350 to 400 of these time steps are saved for a single physics calculation. A comprehensive visualization infrastructure is needed to do post-processing analysis of such large amounts of data.

LANL scientists have had access to visualization facilities ranging from high-end desktops to small multi-panel displays, to a Fakespace RAVE™ (FLEX) immersive system, to a PowerWall Theater, a large multi-panel stereo display seating more than 80. The new LCG was designed as a very high-resolution immersive room for the use of ASC scientists.

3 REQUIREMENTS

Requirements for this facility were developed from past user experiences with the LANL RAVE immersive system and other LANL visualization systems.

Members of the development team also reviewed publications [1, 3, 7, 10, 12] and visited facilities at the Beckman Institute at the University of Illinois at Urbana [6], where a six-sided immersive system has been deployed, and at the Electronic Visualization Lab at the University of Illinois at Chicago [4], where the first CAVE®, was developed.

1a) Room size. Requirements dictated by the size of the available room forced several design decisions. This room measured 30x30 ft, was two stories high, and this constraint precluded many possible configurations of the system.

1b) Room location. The available room is located on the second floor of the building. This had implications for the weight load that could be sustained.

2) Projector brightness and dynamic range. Digital projectors were required rather than the old CRT projectors for this facility, since digital projectors are much brighter, and because digital projectors provide excellent dynamic range of colors, making it possible to discriminate between slight differences in color in the visualizations of the users' simulations.

3) Resolution and pixel density. Good resolution is a necessity for a high-quality exploration of scientific data. [10] High resolution and high pixel density were important requirements for this facility for the following three reasons: 1) the complexity of user data demands extremely high resolution, so that small, yet important, features can be clearly seen, and 2) individual pixels are much more obvious in a digital projector than in a conventional CRT projector.

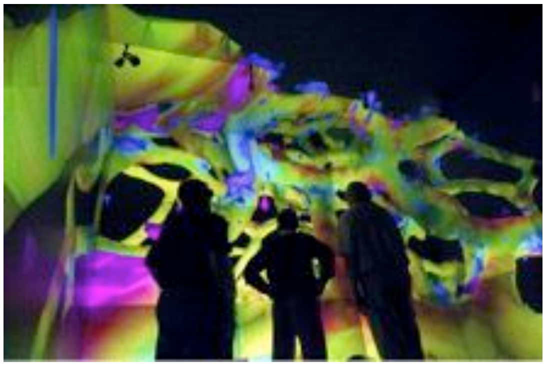


Figure 1. Several LANL scientists look up at data from a foam crush simulation visualized in La Cueva Grande. Data courtesy of Scott Bardenhagen, LANL. Photograph by Presley Salaz, LANL.

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness. This paper is published under LA-UR 05-9199.

4) Video formats. Each screen was required to have square pixels, and the same resolution as all other screens, since otherwise, it would have been impossible to genlock the graphics cards.
5) Tracking. The tripping hazard in the small space made wireless tracking a requirement. Wireless tracking also has the advantage of greater resolution than magnetic tracking.
6) Sonics. The environment was required to be relatively free of system noise. It was also necessary to minimize echo.
7) Conformance to the LANL Visualization Corridor. The system had to conform to the standards of the LANL Visualization Corridor. This required integration of the computing platform with the video distribution system, and required consistency of the visualization software environment with that on other systems.

4 DESIGN

4.1 Facility Design

The new immersive display had to meet the space limitation of the room. It also had to meet the demanding needs of pixel density, to compare with the other multi-panel facilities in the visualization corridor of the SCC building. Another design question was the shape and number of surfaces in the facility, and whether all surfaces should be rear-projected.

4.1.1 Facility geometry

La Cueva Grande uses a rear-projected floor. This design choice was made based on prior experience with the LANL RAVE. The RAVE uses a front-projected floor where the projection is onto a durable reflective material, but all the walls are rear-projected onto glass. The display colors thus can never be perfectly matched between these two projection surfaces. The shadows of the users are also highly visible on the floor, because of the brightness and great depth of field of the digital projectors. So, to enable color matching and to eliminate shadows, we chose rear projection.

Because a rear-projected floor had been chosen, it was possible to have a ceiling. Configurations having a ceiling included a five-sided room and a six-sided, totally enclosed room. However, the sixth side was less necessary to LANL needs, and had the disadvantage of producing greater echo than a five-sided space. A permanent non-immersive observation space outside the display area proper was also desired. Thus, a six-sided space was rejected.

4.1.2 User-accessible area

The user area inside of the display space is a volume 15 ft across by 12 ft high by 10 ft deep, which leaves enough room for the light paths to fill the 5 ft wide by 4 ft tall panels. The observation platform outside of LCG proper is approximately 12 ft deep and 25 ft across.

4.1.3 Resolution, projectors and screens

High resolution was desired for this facility. There were, of course, other considerations, including room size and overall cost.

LANL personnel and Fakespace system engineers found that arrays of 5x4-ft panels could be used on each of the five surfaces to give a front wall that was 15 ft wide by 12 ft tall. This front wall uses a 3x3 array of panels/projectors. The side walls were designed to match up with the front wall in 3x2 and 2x3 arrays of panels/projectors for the floor, ceiling, and the two side walls. The front wall uses a direct light path, while the side walls, the floor, and the ceiling displays all use mirrors to get the light to the screens. This maximized the screen surface that would fit into the room. There are 33 panels (and projectors) that make up the full display system, and the pixel density in the facility is 21 pixels per inch.

Human visual acuity, when measured as the ability to resolve two distinct point targets, is about an arc-minute. [12] At the 21 pixels/inch provided in the above scenario, a person would no longer be able to make out distinct pixels when standing about 13.6 feet away, which is a few feet outside of LCG, in the user area platform. Of course, in a facility of this nature, the closer one can approach, the better. For example, for the pixels to be indistinguishable at 4 feet (the middle of LCG) the pixel density would have to be to 71 pixels/inch, which would nearly triple the number of projectors needed to fill the display space, unfortunately leading to much greater cost and also an increased heat load and increased projector noise. In the end, a balance was reached between resolution, cost, and size.

Christie Mirage 2000 projectors were chosen for LCG. The screens are Draper™ screens that fit on top of structural clear Plexiglas screens to maintain the 'flatness' of the walls. The floor uses special 3-in-thick clear Plexiglas in each panel to allow users to safely walk on the display floor. The infrastructure installed in the room was calculated to be able to handle the weight of the display system. The total weight of the display system is 60,000 lbs., and the floor was estimated to be able to handle a dead weight load of approximately 100,000 lbs. Originally, the superstructure of the system, was specified to be steel, but this added too much weight to the system, so fiberglass beams were used to lighten the load.

4.1.4 Tracking system

The Vicon optical tracking is composed of eight high-resolution cameras, which together cover the entire display space, and much of the observation area. Objects to be tracked, such as stereo glasses, gloves, or other interaction devices, are marked with reflective markers, and infrared strobes flash the space at 60 times per second. The cameras capture images of the markers attached to the tracked objects. This tracking system has high spatial resolution, at .5 mm, and also has good temporal resolution.

The cameras are mounted on the ceiling of the facility. Because of the need for coverage, it was necessary to place four of the cameras in front of the LCG ceiling screens. Also, the camera strobes are in the visible spectrum, in contrast to the usual Vicon strobes, which are infrared. This was necessary because the infrared strobes interfered with the infrared emitters that synchronize the stereo glasses with the screens. In practice, the cameras and strobes are not found to be too intrusive.

Two more cameras are deployed on the ceiling just outside the LCG, on each side. The final two are situated on the ceiling in the back of the user area, about 11 feet from the front of the LCG. This placement gives excellent coverage of the LCG screen area, and much of the user platform is covered as well.

The cameras are mounted from behind, so there is no occlusion from the mounting system. The eight cameras provide enough redundancy so that user occlusion is not in general an issue.

A major advantage of this type of optical system lies in its lack of wires and other tethers. This enables a great deal of freedom of motion. Another advantage is the freedom it provides to define gestural and other interfaces.

4.2 Computing System

The 33 projectors and the console designed into La Cueva Grande required new computing power not available at LANL prior to this design effort. SGI had just developed a follow-on SMP system that could address our needs - run CEI's EnSight suite of visualization and analysis tools right out of the box and support the 34 graphics devices we wished to drive. The new SGI system is an Onyx4 3900 with a modular design that was configured with 34 graphics engines (or pipes) based on an ATI FireGL X1. SGI integrated the cards into the architecture so that the cards were genlockable. The system was delivered with 17 GBricks each containing 2 graphics pipes and 20 CBricks (computer modules) for a total of 80 processors and 80 gigabytes of memory with a fiber-channel-based disk system with more than 1 terabyte of storage capacity.

5 CONSTRUCTION OF THE FACILITY

5.1 Construction Sequence

This section comprises a sequential series of illustrations that show the construction of LCG.

Construction Phase 1: The contractor installed the major structural elements and did as much as possible to complete dust-generating activities prior to equipment installation. Major subelements in this phase included the steel floor grid, the user platform and stairway, rough-in of HVAC, electrical, fire suppression, and security subsystems, and a 1-ton ceiling bridge crane.

Installation Phase 1: Fakespace installed the major structural elements, including the four-tier fiberglass mezzanines, the steel and aluminum screen support assembly, the acrylic screens, projector stands, and mirror stands.

Construction Phase 2: The contractor returned to finish installing all infrastructure, including electrical conduit, fiber wireways, HVAC ducting, fire suppression pipes, and sprinklers.

Installation Phase 2: Fakespace returned to install the electronic equipment and wiring (projectors, controller, etc), and do final alignment and color matching of the display images.

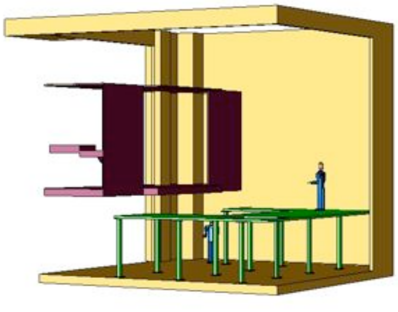


Figure 2. The fiberglass first level is installed.

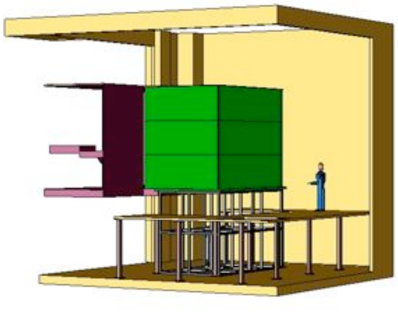


Figure 3. The acrylic screens are set into place over the steel and aluminum supports.

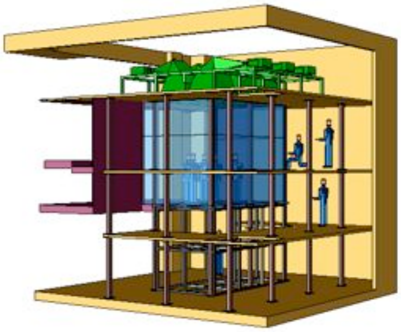


Figure 4. The first projectors and mirrors are loaded into the room.

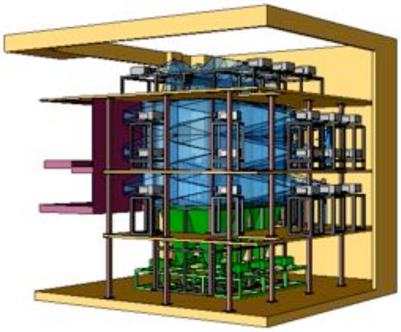


Figure 5. Final assembly.

5.2 Challenging Elements and Lessons Learned

1. Integrating the expectations and practices of two very different trades. Ordinarily, the architectural and construction trades execute as-built tolerances that can be measured in inches. However, the immersive visualization “trade” has significantly tighter requirements for fidelity to design, measured in 1/32-in. or even 1/64-in.increments. This is due to the necessity of keeping the highly sensitive human eye from seeing misalignments in images displayed across multiple screens.

Lessons Learned: Effort should be made to impose strict discipline in this respect during the construction rough-in.

2. Load Management. The steel superstructure initially designed by the supplier was found to be too heavy for the design load of the second-floor slab, so fiberglass was used instead.

Lessons Learned: It would simplify and cut costs if any future installation of this scale were executed on a ground floor level.

3. Life/Fire Safety. Due to the extensive use of fiberglass and acrylic materials in the installation, the Laboratory Fire Marshal was compelled to impose extraordinary requirements to ensure the safety of staff and to minimize risk of damage to other facilities.

Lessons Learned: The need for fiberglass in place of steel for the mezzanine structure greatly increased the flame-spread and the smoke-generation characteristics of the installation. It would simplify such an installation to be executed on a ground floor.

4. Ensuring Serviceable Utility. During design, attention was paid to ensure that construction did not interfere with the servicing of the display equipment. This effort resulted in the early detection of instances where planned infrastructure, such as HVAC ducting and fire suppression pipes, made access to projectors problematic. The supplier pointed these instances out in a timely way, and design corrections followed. It is estimated that about 90% of such potential flaws were caught early in the design cycle.

Lessons Learned: The decision to build in multiple phases logically meant that critical infrastructure would be in place prior to display equipment. It is highly recommended that this approach be taken in the deployment of any similar facility.

5. Quality Assurance, and Fidelity to Original Intent. The pixel count is a principal and defining characteristic of LCG. The pixel count requirement ultimately dictated nearly every design parameter, as it set the requirement for projectors, for screen size, and therefore for the mezzanine structure, power, and cooling.

6 SOFTWARE INTEGRATION

An important part of the post-deployment effort was the integration of the software with the rest of the LANL Visualization Corridor. The visualization corridor was designed so that any user could use any system, once he or she had become familiar with the basic components of the corridor. The LCG had to meet this requirement to maximize ease of use.

6.1 CEI's EnSight software

Computational Engineering International (CEI) EnSight software is the standard scientific visualization tool in use at LANL. It runs on all the ASC architectures on both shared and distributed-memory rendering platforms, and can be used in any of the LANL ASC display environments. Every ASC user code is interfaced to this tool. This software is fully integrated into the multi-panel, non-planar immersive facility. Users can come into the facility, start EnSight the same way that they normally would at their desktops.

6.2 Tracking software

The images gathered by the tracking cameras of the Vicon system are sent to an image processing system that uses the Vicon Tarsus software. Tarsus extracts from the images positional and rotation data corresponding to the objects being tracked. However, the EnSight software does not use the native Tarsus format, but instead uses the VRCO Trackd interface for tracked input devices.

LANL personnel have written a translator between the Tarsus and Trackd formats, so that EnSight could make use of the tracked data. Included is an interface to permit users to define and add their own tracked devices easily, thereby allowing them to take advantage of the flexibility and freedom provided by Vicon.

6.3 Other software

Other software in use in this facility includes StereoMoviePlayer, a LANL movie player, as well as CEI's EnLiten geometry browser and EnVideo movie tool.

7 FACILITY USE

La Cueva Grande is a very attractive and intimate working space. The high resolution, the small pixel size, the close-up view, and the brightness of the projectors provide an excellent sense of presence and give the objects in the space solidity.

Once the facility was deployed, it became clear that the perceived constraint of small room size was actually an advantage. Many users have remarked upon its usability as a working space and this is due in part to the intimacy of the environment, combined with the resolution of the facility.

An emergent mode of use is when a group allows one person to interact with the visualization, while others observe and discuss. When a change in position is desired, the group asks the driver to make the change, thus putting the user interface at one remove. This was an unexpected mode of interaction, and it is currently the mode of choice by the users. The scientists who are observing are not distracted by the need to interact, and are free to concentrate on the simulation. The need to learn these interactions impedes use, and as more natural interactions are developed, we anticipate greater success.

The LCG is about half as deep as it is wide, so is shallow but non-planar, having side walls at right angles to the front wall. It affords use as a PowerWall with side views to infinity. The viewer has a view as if he or she were looking out a window into the space. Because this view goes to infinity, any object placed in front of the viewer can be seen in its entirety, which is not true of any planar PowerWall.

8 ACKNOWLEDGEMENTS

Many people took part in the design and deployment of LCG.

At LANL, Tom Wyant, Chuck Wilder, and Jerry Antos provided infrastructure support, and Georgia Pedicini addressed security. Steve Hodson, Dave Pugmire, Katharine Chartrand and Bob Greene wrote software and brought user data into LCG.

At Fakespace, Steve Fine, Bruce Martin, and Jeff Salasek were responsible for the design of the superstructure installation. Charles Fraresso and Chad Kickbush were project managers, and Doug Boyers and Ryan Torrey led the on-site installation team.

The firm of DMJMH+N, Inc. executed the architectural design. Hensel Phelps Construction Company carried out the construction. Terri Bednar, Jack Tapie, John Baillie, Rusty Brown, Michel Castejon, Keith Rich, Connie Griffiths, Don Pickett, Michael Skowvron, and Susan Bechly of SGI were instrumental in the Origin 3900 stand-up. Anders Grimsrud and Daniel Schikore of CEI developed the multi-planar display support capabilities of the EnSight software. Jason Hunter of Vicon set up the Vicon system. Brian Nilles and Jon Damush also contributed to the Vicon deployment.

Many thanks to Hank Kaczmarski, Camille Goudeseune, and Ben Schaeffer at the Beckman Institute at the University of Illinois and to Tom DeFanti, Dan Sandin, and Greg Dawe at the Electronic Visualization Laboratory at the University of Illinois at Chicago, who kindly allowed us to view their immersive facilities while we were planning ours.

REFERENCES

[1] D. Bowman, E. Kruijff, J. LaViola, I. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison-Wesley, 2005.

[2] Computational Engineering International, http://www.ceintl.com/.

[3] C. Cruz-Neira, D. Sandin, T. DeFanti, J. Hart. “The CAVE Automatic Virtual Environment”. *Communications of the ACM, 35(6)*, pp. 64-72, 1992.

[4] Electronic Visualization Lab at the University of Illinois at Chicago: http://www.evl.uic.edu/intro.php3

[5] Fakespace Systems Inc: http://www.fakespace.com/.

[6] Integrated Systems Laboratory, Beckman Institute at UIUC. http://www.isl.uiuc.edu/.

[7] M. Robinson, J. Laurence, A. Hogue, J. E. Zacher, A. German and M. Jenkin. “IVY: Basic Design and Construction Details”. *ICAT 2002*, 2002.

[8] VRCO: http://www.vrco.com/.

[9] VICON Motion Capture. http://www.vicon.com/.

[10] A. van Dam, A. Forsberg, D. Laidlaw, J. LaViola, and R. Simpson. "Immersive VR for Scientific Visualization: A Progress Report", IEEE Computer Graphics and Applications, Vol. 20, No. 6, Nov/Dec, 2000.

[11] C. Ware, *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers, Inc., San Francisco, California, 2000.

[12] T. Yamada, M. Hirose, and Y. Isda. “Development of complete immersive display: COSMOS”, *Proceedings of the VSMM* ‘98, pp. 522-527, 1998.

Environmental and Immersive Display Research at the University of Southern California

Mark Bolas*
Interactive Media Division
University of Southern California

Jarrell Pair† Kip Haynes‡
Institute for Creative Technologies
University of Southern California

Ian McDowall‡‡
Fakespace Labs, Inc.
Mountain View, California

ABSTRACT

The University of Southern California and its collaborative research partner, Fakespace Labs, are participating in a number of research programs to invent and implement new forms of display technologies for immersive and semi-immersive applications. This paper briefly describes three of these technologies and highlights a few emerging results from those efforts.

The first system is a rear projected 300 degree field of view cylindrical display. It is driven by 11 projectors with geometry correction and edge blending hardware. A full scale prototype will be completed in March 2006.

The second system is a 14 screen projected panoramic room environment used as an advanced teaching and meeting space. It can be driven by a cluster of personal computers or low-cost DVD players, or driven by a single personal computer.

The third is a prototype stereoscopic head mounted display that can be worn in a fashion similar to standard dust protection goggles. It provides a field of view in excess of 150 degrees.

1 CYLINDRICAL DISPLAY SYSTEM

In late 2004, the University of Southern California Institute for Creative Technologies was asked to design and deploy a display system for training soldiers in close air support (CAS) tasks. CAS requires a ground-based soldier to guide an aircraft to a target using radio communication. The project called for a minimum 270 degree horizontal field of view and the ability to view aircraft directly overhead. The physical dimensions of the system had to fit within a former vehicle repair garage. Initially, a front projection dome system was considered. However this approach was not feasible due to equipment cost restrictions and long term maintenance concerns.

The final design was a rear projected cylinder with a ceiling screen as illustrated in Figure 1. Our initial specifications provided a twenty foot radius ceiling projection covering an eight foot high vertical screen. These dimensions provided a 300 degree horizontal field of view exceeding the application requirements.

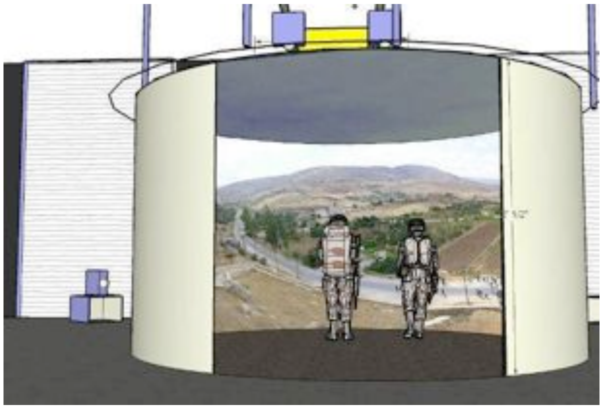


Figure 1: Rendering of final cylindrical display system to be deployed March 2006

1.1 Implementation Process

A significant project challenge involved producing the highest display resolution possible within a somewhat limited budget. To accomplish this goal, we looked to the work of University of North Carolina at Chapel Hill researchers who in the late 1990s developed high resolution displays by blending and warping imagery from multiple projectors [1]. Commercial hardware incorporating these techniques have emerged in the past several years.

Leveraging the availability of these products, we designed our final system to use eleven 1400x1050 resolution DLP projectors. Each projector is equipped with built in geometry correction and edge blending hardware. Seven projection channels provide imagery for the vertical screen and the remaining four channels provide the ceiling sky projection. Each channel connects to a single PC image generator (IG). Imagery on all IG nodes is synchronized via a master graphics client broadcasting global event changes. We chose rear projection to prevent user shadows on the large interior display surface.

The display was developed by first fabricating a small scale prototype, five foot high with an approximately 180 degree field of view (Figures 2-3). This system informed the construction of

---

*e-mail:bolas@well.com
†e-mail: pair@ict.usc.edu
‡e-mail: haynes@ict.usc.edu
‡‡e-mail: ian@well.com

the final version by allowing us to evaluate projection throws, geometry correction, blending performance, and screen materials.

The rear projected cylinder required a highly specialized frame design. We employed the services of a mechanical design contractor to design a unique solution utilizing a rigid steel frame in combination with suspension cables positioned outside the vast projector throw areas.

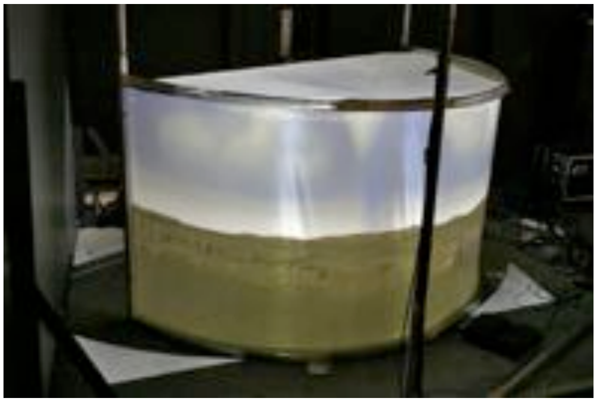


Figure 2: Rear view of small scale cylindrical display prototype with suspension cables and rigid steel frame

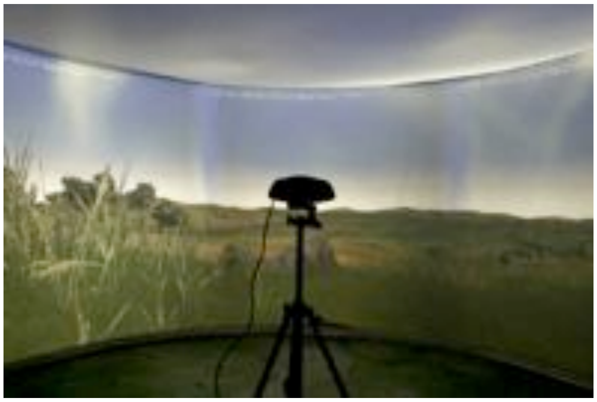


Figure 3: Interior view of small scale prototype

1.2 Project Findings

This effort highlighted the clear advantage of developing prototypes when designing new display systems. With our small scale system we immediately learned that an eight foot screen height provided an inadequate level of presence. We revised our final design by raising the screen height to twelve feet, which completely fills the user's vertical field of view when looking straight ahead.

A key project hurdle involved selecting an appropriate screen material. In addition to covering a large surface area, the screen could not be prohibitively expensive. Initial tests revealed the obvious need for screens with a relatively low gain factor. It was difficult to find a rear projection material of the size we needed without fusing together multiple vinyl segments. This process creates distracting seams in the displayed imagery. After testing several materials and contacting multiple manufacturers, we were ultimately able to identify an affordable vinyl material that could accommodate our twelve foot height and 300 degree field of view.

The construction of the full scale system will be completed in March 2006, and soldiers will begin using the system for close air support training soon afterwards. At that time, we plan to solicit user feedback on the display's effectiveness to guide the design of our future immersive systems.

2 LARGE AREA VISUALIZATION DISPLAY FROM A SINGLE COMPUTER

It is generally agreed that increasing display screen real estate increases a user's productivity and benefits collaboration and communication in classroom and meeting room environments. Until recently, it was impractical - except in very special installations - to provide more than a few screens to most computer users due to the size and cost of the displays themselves.

Given recent technological and commercial advances in flat panel and projection displays however, it is suddenly becoming both financially affordable and physically possible. As such, providing large numbers of displays for group presentation environments, and even for individual users, is close to being within reach for general adoption. One limitation remains: driving multiple displays is problematic from both a hardware and application software perspective.

Figure 4: Large area visualization display at USC's Zemeckis Media Laboratory

Some systems, such as NASA Ames' "hyperwall," [2] utilize one computer per display screen, thereby overcoming the problem of bandwidth limitations. Using one computer per display screen, however, is obviously expensive and also requires that custom application software be developed to synchronize a cluster of computers.

In trying to overcome this challenge, it is instructive to realize that much of the data on a computer display changes less often than 60 times a second, thus the video channel has redundancy that can be reduced to drive more displays without increasing the channel bandwidth. For example, there is little need to update the visual data for fairly static images, such as spreadsheets, text documents, or the like, at 60 Hz. In addition, a user looking at a bank of computer displays typically concentrates on a single display at a time, thus offering an opportunity to further reduce the overall required bandwidth.

2.1 Implementation:

Our system (Figures 5-6) synchronizes the video output from a single personal computer using a freeze-frame buffer that is a standard feature of the 14 DLP projectors (NEC Model LT260K) we use. A PC routes its single video output signal to a distribution amplifier that relays the signal to the projectors. The PC then uses a USB port to convey commands to our custom-developed 'freeze box' to either freeze or un-freeze the projected video on selected projectors. In this way, a single PC can drive multiple projectors. The freeze box consists of a simple micro-controller that interfaces to the projectors via 14 remotely located infra-red LEDs which it drives to mimic standard remote-control commands - including 'freeze frame'.

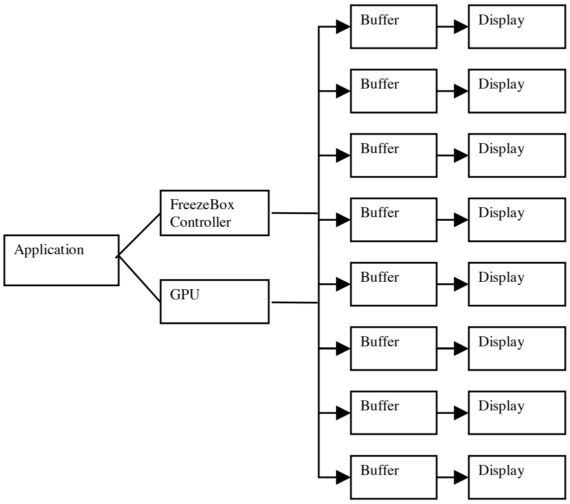


Figure 5: System architecture

For example, the PC can be programmed to sequentially display 14 full screen JPEG images, with each image displayed for 2 seconds before cycling to the next image. As it cycles between the images, a different projector is commanded to freeze its image. In this way, 14 unique images will be displayed at the end of the cycle. The 2 second delay is required to ensure proper timing given the projectors' characteristics and the open-loop control architecture.

2.2 Discussion

The system has been used for a number of applications including presenting panoramic imagery [3] and allowing for massive amounts of imagery to be simultaneously viewed. A proof of concept test with extremely detailed imagery from Clifford Ross's R1 camera system [4] was implemented and highlighted the need for such tiled systems to meet the demand of the human eye for resolution. The room is often used for collaborative presentations in which a presenter and the audience work together to add visual content to the many screens [5].

Although it would be interesting to pursue alternative architectures that allow for multiple applications and GPUs to directly address each display's buffer, we have been struck by the value of limiting the system to a single PC and GPU. The ease with which applications can be modified to work with the approach is quite valuable. It requires only that software be modified to drive multiple displays from a single application, rather than trying to synchronize multiple instances of an application distributed across a cluster.

3 FAST ACTION WIDE FIELD OF VIEW HEAD MOUNTED DISPLAY

Humans use peripheral vision for spatial orientation and motion cues, while central vision is useful for detailed imagery and color perception. In many ways, central vision can be thought of as telling one 'what' is in an environment, while peripheral vision informs the user about the 'where' of an environment. Early HMDs focused on providing a sense of presence through the power of peripheral vision. In the quest for crisp central imagery, many modern HMDs have reduced their fields of view to the point of eliminating their ability to provide a feeling of immersion, or have grown bulky to the point of interfering with natural head motions.

Figure 6: Prototype Fast Action HMD (FAHMD)

The Virtual Technologies and Environments program at the Office of Naval Research sponsored the development of novel approaches for the design and development of Head Mounted Displays that can meet the requirements of training for close quarters battle (CQB) requirements. Such a display system must be capable of providing spatial and situational context to the soldier, thus it requires a very wide field of view that can display imagery to fill the user's peripheral vision. Additionally, CQB requires rapid head and body motions, thus demanding a very lightweight display solution and innovative mounting configuration.

Fakespace Labs and its STTR research partner USC have recently developed an HMD to meet these requirements and have just begun to qualify the capabilities and experiences that can be created with the prototype HMD called the Fast Action HMD (FAHMD).

3.1 Discussion

The FAHMD employs a canted optical approach similar to the Cyberface II [6] to achieve a 160 degree field of view. It provides a generous exit pupil that eliminates the need to adjust the IPD for different users and is light enough to allow for reasonably rapid head motions.

While we have only recently completed the prototype system, our initial experiences were significant, although strictly qualitative. The handful of users that have tried the system typically had over 10 years experience with immersive display systems and technologies. They reported a sense of presence on par or beyond that of most previous immersive system experiences.

Designing environments for the FAHMD has proven to be more difficult than expected, as the users are surprisingly finicky about what looks “right” in the environment. In many ways, we feel that the field of view is providing a sense of presence strong enough to begin to evoke reactions similar to those described by Masahiro Mori in his theory "The Uncanny Valley" [7]. As the environments get closer to “real”, users begin to relate to the virtual environment in a natural fashion, thus deviations from natural become increasingly objectionable.

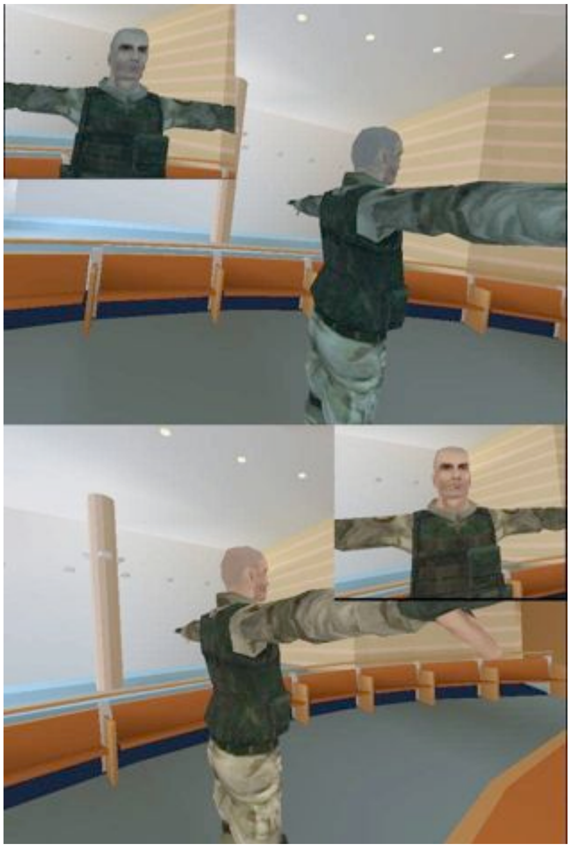


Figure 7: Virtual human models used in room lighting tests

An anecdotal example of this concerned the color of virtual human models in a room environment (Figure 7). Users commented that the color of a particular character's skin appeared too gray, and as a result the model looked like a corpse. When a graphical mask was used to crop the field of view to 60 degrees diagonal, the virtual human's tint did not immediately appear wrong. We theorize that the larger field of view was allowing the user to consider the overall lighting model of the room, with respect to which the virtual human's tint was incorrect. With the 60 degree field of view, it was not possible to be close enough to the human model and simultaneously explore the room's lighting conditions to make such an observation

We expect to begin testing how this ability to visually determine context affects a user's experience of color, size, and motion, and hope to then explore the effect on presence.

4 ACKNOWLEDGEMENTS

The cylindrical display project described in this paper was developed with funds of the Department of the Army under contract number DAAD 19-99-D-0046. The FAHMD project was developed under ONR Contract N00014-03-C-0481. Any opinions, findings and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsors.

The authors wish to thank Dr. James Templeton and Linda Siebert for their advice and support, and Perry Hoberman for his assistance implementing the squeeze box.

REFERENCES

[1] Raskar, R., Welch, G., Fuchs, G. “Seamless Projection Overlaps using Image Warping and Intensity Blending.” Proceedings of the 4th International Conference on Virtual Systems and Multimedia, Gifu, Japan, November 1998.

[2] Sandstrom, T.A., Henze, C., Levit, C. “The Hyperwall.” In Proc. IEEE CMV 2003 Conference on Coordinated Multiple Views, Information Visualization 2003.

[3] Fisher, S., Anderson, S., Ruiz, S., Naimark, M., Hoberman, P., Weinberg, R. "Experiments in Interactive Panoramic Cinema", Stereoscopic Displays and Virtual Reality Systems IX, Proc. SPIE 5664 (2005).

[4] Salamon, J., "Tom Swift's New Camera, Ready for Space and Spies: An Artist Turns Inventor", The New York Times, pp. E1, Dec. 9, 2005.

[5] Hall, J., Fisher, S., "Experiments in Backchannel: Collaborative Presentations Using Social Software, Google Jockeys, and Immersive Environments". Forthcoming CHI 2006, April 22-27, 2006.

[6] Howlett, E., “Cyberface II applications note.” LEEP Systems/Pop-Optix Labs, Waltham, MA, 1990.

[7] Mori, M., “Bukimi No Tani (The Uncanny Valley).” Energy, 7, 33–35, 1970.

13:00–13:50 Session III

RANSAC-Assisted Display Model Reconstruction for Projective Display

Patrick Quirk, Tyler Johnson, Rick Skarbez, Herman Towles, Florian Gyarfas and Henry Fuchs

Department of Computer Science
University of North Carolina at Chapel Hill

ABSTRACT

Using projectors to create perspectively correct imagery on arbitrary display surfaces requires geometric knowledge of the display surface shape, the projector calibration, and the user's position in a common coordinate system. Prior solutions have most commonly modeled the display surface as a tessellated mesh derived from the 3D-point cloud acquired during system calibration.

In this paper we describe a method for functional reconstruction of the display surface, which takes advantage of the knowledge that most interior display spaces (e.g. walls, floors, ceilings, building columns) are piecewise planar. Using a RANSAC algorithm to recursively fit planes to a 3D-point cloud sampling of the surface, followed by a conversion of the plane definitions into simple planar polygon descriptions, we are able to create a geometric model which is less complex than a dense tessellated mesh and offers a simple method for accurately modeling the corners of rooms. Planar models also eliminate subtle, but irritating, texture distortion often seen in tessellated mesh approximations to planar surfaces.

CR Categories: I.3.2 [Graphics System]: Distributed/network graphics— [I.3.5]: Computational Geometry and Object Modeling— [I.3.7]: Three-Dimensional Graphics and Realism—Virtual reality I.4.8 [Scene Analysis]: Range data, Stereo, Surface fitting— [I.6.7]: Simulation Support Systems—Environments

Keywords: functional reconstruction, RANSAC plane fitting, 3D-point cloud simplification, surface fitting, projective display

1 INTRODUCTION

To create undistorted imagery on a display surface using casually placed projectors requires the projected imagery to be pre-warped so that the view of the imagery from the user's perspective will appear correct.

If the display surface is a simple plane, the warping function can be described with a simple homography that re-maps the *ideal* image from the display plane back into the projector's image plane. This homography, which can be represented as a collineation matrix, can be concatenated into the graphics transformation stack to render the pre-warped image at no additional computational expense [9, 13, 2].

If the display surface illuminated by a given projector is more complex (e.g. a multi-plane corner or even an arbitrarily shaped curtain), then we and others [12, 10, 15, 1] use a 2-pass rendering algorithm to re-map the desired image so it looks perspectively correct from a given viewpoint. The actual *warping* is achieved in the second pass by using the graphics hardware to projectively texture the ideal image rendered in the first pass onto a 3D model of the display surface.

Many 2-pass display implementations use a calibrated, stereo camera pair in an up-front calibration step to first evaluate 3D points on the display surface based on features that are actively projected. Many types of patterns and codings are commonly used to make feature extraction and the stereo correspondence problem as simple as possible. The resulting 3D-point cloud representation of the display surface is commonly triangulated to produce a tessellated mesh surface description used in the aforementioned 2-pass rendering algorithm.

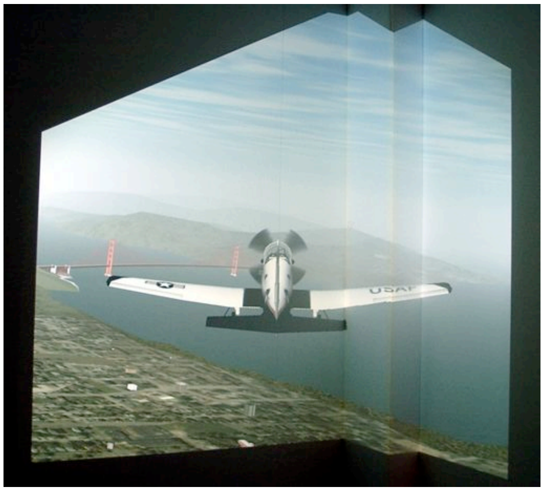


Figure 1: FlightGear displayed into a complex room corner without distortion.

The challenge with these reconstruction methods is producing an accurate surface model. If sampling is too sparse, the piecewise approximation will not accurately model the display surface shape. This is commonly the case when trying to find the corners of a room. If the sampling is too dense and the tessellated mesh is not simplified, corners may be found with reasonable accuracy but the cost of rendering such a model in the second pass may be become overly expensive. In addition, the noise inherent in all image-based stereo reconstruction processes produces 3D points with error. These resulting noisy surface models can produce subtle, but noticeable re-mapping errors in many rendered scenes.

To avoid these difficulties, we have pursued another approach to the task of accurate display surface fitting. It is based on the notion of function reconstruction [6], which turns the task of surface fitting into one of mapping common surface functions (planes, cylinders, spheres) to the surface domain represented by the point cloud samples. In fact, function fitting has been applied by Raskar [11] to reconstruct quadric surfaces for curved-screen projective displays.

In this paper, we describe a method for functional reconstruction of the display surface that takes advantage of the knowledge that most interior display spaces are in fact piecewise planar. Using a RANSAC plane fitting algorithm, we first *recursively* extract individual plane definitions from a complex point sample set originally derived using standard stereo reconstruction methods. The plane definitions are then processed and converted into simple planar polygon descriptions. The resulting geometric model is significantly less complex than a dense tessellated mesh and more accurately models the corners of rooms, which are computed from the intersection of planes. Using the smooth planar models in the 2-pass projective texture rendering process also eliminates the subtle texture distortion often seen in tessellated mesh approximations to planar surfaces in tiled projective display applications.

In Section 2, we describe our method, in Section 3 we present and discuss some results, and in Section 4 we summarize and offer thoughts on future work.

2 DISPLAY SURFACE MODELING

To re-iterate, display surface modeling along with projector calibration are required in order to pre-warp and thus create perspectively correct imagery for a given user's view. In this section, we detail the three steps our new procedure uses to create an accurate display surface model of complex room geometry.

This room model is then used in the 2-pass application rendering process to achieve the necessary pre-warp of the projected images to create imagery that can wrap around the walls of the room, providing an undistorted immersive experience for the user.

2.1 3D-Point Cloud Generation

A calibrated, stereo camera pair is used to reconstruct the display surface and create the initial 3D-point cloud representation. The basic steps in stereo reconstruction are well established:

- Extract feature points from the stereo camera images.
- Establish pairwise feature point correspondences.
- Triangulate to find the 3D point representing the ray intersection of each corresponding (x,y) feature point pair.

Since ideal display surfaces are without texture, it is most convenient to use the projectors themselves to create the image features needed for surface extraction. Features outside the illuminated region are not needed as it is only necessary in the rendering process to re-map the ideal view image into the image space of each projector.

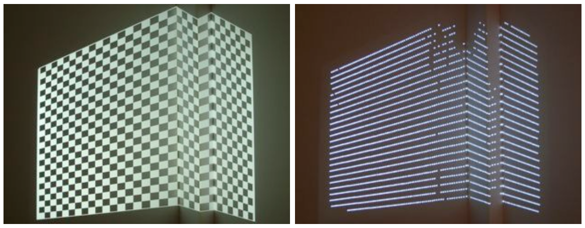


Figure 2: (a) Checkerboard feature pattern. (b) Reprojected 3D feature points.

Only a rather sparse sampling of the piecewise planar surfaces is required for RANSAC plane fitting, so we currently project a checkerboard pattern to establish robust features. The Open Computer Vision library (a SourceForge project) is used to extract the corner features in the stereo images. If the piecewise planar segments of the display surface are small, such as the narrow corner column visible in Figure 2a, we can shift the projected checkerboard pattern horizontally and/or vertically to create more feature points.

To make the task of identifying corresponding features simple and robust between the stereo image pair, we also project a series of images with binary-coded markers that are centered on the checkerboard corners that give each corner feature a unique identification code. Figure 2b shows the resulting 3D points being rendered and reprojected onto the display surface. Missing points may result from checkerboard corner extraction failures or feature point identification decode errors that are generally associated with low image contrast.

Given the list of 2D matching image points and the camera matrix P for each camera, we use a simple linear triangulation method [4] to find the corresponding 3D points. For each point, this involves solving for four homogenous unknowns in four equations using a singular value decomposition method of the DLT algorithm.

2.2 RANSAC Plane Fitting for Recursive Plane Extraction

Since the point cloud generated in our system calibration process could represent multiple planar surfaces, it would be necessary to first segment the data into plane-related point clusters if one were to use a standard iterative minimization technique (e.g. least squares) for plane fitting. This segmentation process would basically classify points as inliers and outliers with respect to a given plane fit operation.

To avoid this complex segmentation step, we use a Random Sample Consensus (RANSAC) algorithm [3] to fit planes to the point cloud data. RANSAC is designed to work in the presence of many data outliers. The algorithm begins by fitting a plane to a set of 3 points randomly selected from the input point cloud. These 3 points are then classified as inliers. All other points are then compared to a cost function (distance from the plane) and are reclassified as inliers if they are within an epsilon distance of the plane. These two steps are repeated until the number of inliers converges to a maximum or a count limit is reached. A least-squares fit of the consensus inliers then defines the output plane definition.

To fit planes to all the data in the the point cloud description, the RANSAC plane fitting process is repeated with the candidate points for the next pass coming from the outliers of the previous pass. We terminate the plane generation loop after either a user-specified number of planes are evaluated or the number of outliers remaining to be processed falls below a heuristic threshold.

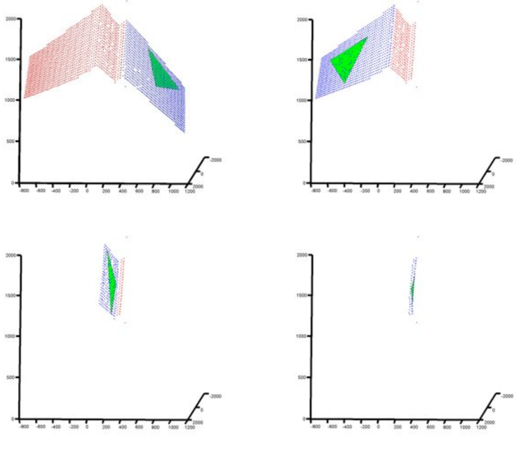


Figure 3: Matlab rendering of successively reconstructed planes.

Figure 3 shows the extraction of four successive planes beginning with a combined 3D-point cloud generated for a display surface looking into a room corner that contains a rectangular column. The green triangle in each image represents the plane extracted on that pass, while the blue points show the inliers and the red points the outliers for that plane. Notice that the number of points to process in each subsequent pass diminishes as inliers for each fitted plane are removed.

Our current plane fitting implementation is in Matlab and we utilize the ransacfitplane.m function authored by Kovesi [7].

2.3 Plane to Polygon Conversion

Given the set of plane definitions generated from the RANSAC plane fitting step, we next create a polygonal representation of the display surface for use in the projective texturing step of our 2-pass rendering algorithm.

The general solution to this task can be a complex issue of which planes to intersect in order to created bounded polygons. To simplify this decision, we currently restrict the display to vertical wall surfaces. This allows us to create a 2D *floor plan* representation of the point cloud data by projecting the inliers of each extracted plane down onto a virtual floor plane of the room, thus creating a line segment for each plane whose endpoints span the projected inliers.

In general, the line segments representing the walls of the room may not overlap due to a lack of point cloud data in the corners of the room. To determine which line segments correspond to contiguous walls, we extend the endpoints of each line segment in order to form intersecting corners. To avoid erroneous corner detection, we consider only those intersections which require the least amount of line segment extension to become intersecting corners.

After extending the line segments in the floor plan to form the corners of the room, it is trivial to create a polygonal display model from the result. We merely extend each line segment in the up-direction by some desired height to turn the line segments into quadrilaterals, forming a simple piecewise planar display model.

3 RESULTS

The ability of our plane fitting method to automatically extract accurate piecewise-planar display surface models is illustrated in Figures 1 and 4b. These images show two different application programs displayed onto a room corner with no perceived distortion. Without an accurate surface model the 2-pass rendering system would simply not be able to create such precise, pre-warped projected imagery in which straight lines remain straight. As a measure of the amount of pre-warping required, Figure 4a shows the actual pre-warped output that projected to create the scene captured in Figure 4b.

Figure 4: Piazza San Marco, Venice. (a) Pre-warped frame buffer output. (b) Undistorted projected view.

Figure 5a shows the actual room corner *without* projector illumination. It is a complex 4-plane structure that includes a 5" x 12" offset column. Figure 5b shows the reconstructed computer model of this corner with room-aligned coordinate system axes. Beginning with 3618 reconstructed vertices, our RANSAC plane-fitting, surface reconstruction algorithm produced an output display model represented by just 4 quadrilaterals defined by ten 3D vertices. The plane fit tolerance specified was 5 millimeters.

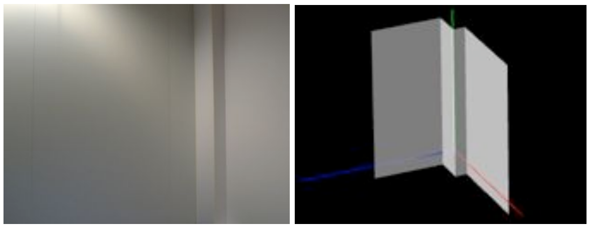


Figure 5: (a) Real-world room corner. (b) Reconstructed display surface model.

One measure of comparing the real room with the extracted model is to compare the dimensions of the offset column. In doing so, we found that the offset column feature was modeled with less than 2.3% dimensional error.

As another measure of the overall display system modeling accuracy, we can also render the display model with each plane distinctly colorized. If the model and projector matrix computation are accurate, then the projected colorized planes will map back precisely onto the their respective planes in the real-world. Figure 6 shows just such a re-projection. Close examination shows a very accurate remapping with color transitions occurring as expected at the surface intersections with one minor exception. In the lower part of the image, the yellow plane overlaps onto the real-world red plane a few millimeters.

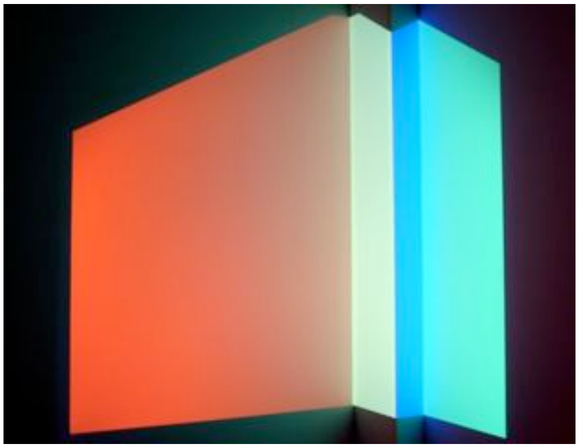


Figure 6: Extracted display model re-projected onto the real-world corner.

Our testing to date has involved configurations of one and two projectors (ProjectionDesign F1 SXGA) illuminating a corner of a room. We have explored two different corner configurations - a standard 2-plane corner and the more complex 4-plane corner. The material presented in this paper is based on a single projector configuration illuminating the complex wall corner.

Our calibration toolset is a combination of C/C++ code and Matlab scripts running on Windows XP. Beginning with a pre-calibrated stereo camera pair, the script-driven process runs automatically taking less than 5 minutes to complete the full rendering system calibration from checkerboard image capture to display model polygons and projector matrices out.

4 SUMMARY AND FUTURE WORK

We have demonstrated a new method for functional reconstruction of projective display surfaces that takes advantage of the knowledge that most interior display spaces are piecewise planar. Using a RANSAC algorithm to recursively fit planes to a 3D-point cloud sampling of the surface, this new solution clearly generates geometric models that are both simple and accurate in their representation of the display surface shape. As a result, we are now able to more precisely create geometrically correct and pleasing immersive imagery on the walls of any room with casually placed projectors than previously possible when modeling the surface as a complex tesselated mesh.

There are a number of areas for future research including:

- Scaling up to handle more walls and projectors and thereby create a more immersive, wide-area visual display system.
- Improve the robustness of the current plane to polygon conversion algorithm to handle without restrictions an arbitrary configuration of extracted plane definitions.
- Implement the optimal triangulation algorithm [4] to replace the less precise homogeneous DLT method used now.
- More fully review the literature for the applicability of other surface fitting strategies that can model a wider variety of shapes and sharp features. Certainly the research of [5, 14, 8] and others is of relevance. Also, the multiRANSAC algorithm [17] for detecting multiple planar homographies in parallel is of importance.

Longer term, we are especially interested in developing techniques whereby the calibration process is not just a pre-process step to rendering, but one [16] that continuously and successively refines and improves display quality while the user application is running.

ACKNOWLEDGEMENTS

The authors would like to especially thank Chris Ashworth for his early contributions in prototyping many of the ideas in this paper, and to our research colleagues Greg Welch and Marc Pollefeys for their collaborative support. This research was primarily supported by the DARPA DARWARS Training Superiority and DARPA VIRTE (Virtual Technologies and Environments) programs under the Office of Naval Research award number N00014-03-1-0589.

REFERENCES

[1] M.S. Brown and W.B. Seales. A practical and flexible tiled display system. In *Proceedings of the 10th Pacific Conference on Computer Graphics and Applications*, pages 194–203, New York, NY, USA, 2002. IEEE.

[2] Han Chen, R. Sukthankar, G. Wallace, and Kai Li. Scalable alignment of large-format multi-projector displays using camera homography trees. In *Proceedings of Visualization 2002*, pages 339–346, New York, NY, USA, 2002. IEEE.

[3] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.

[4] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in computer vision*. Cambridge University Press, first edition, 2002.

[5] Hugues Hoppe, Tony DeRose, Tom Duchamp, Mark Halstead, Hubert Jin, John McDonald, Jean Schweitzer, and Werner Stuetzle. Piecewise smooth surface reconstruction. In *SIGGRAPH '94: Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 295–302, New York, NY, USA, 1994. ACM Press.

[6] Hugues Hoppe, Tony DeRose, Tom Duchamp, John McDonald, and Werner Stuetzle. Surface reconstruction from unorganized points. In *SIGGRAPH '92: Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, pages 71–78, New York, NY, USA, 1992. ACM Press.

[7] Peter Kovesi. Ransacfitplane - a robust matlab function for fitting planes to 3d data points. http://www.csse.uwa.edu.au/*(tilde)*pk/research/matlabfns/robust/ransacfitplane.m, June 2003.

[8] Maxime Lhuillier and Long Quan. A quasi-dense approach to surface reconstruction from uncalibrated images. *Transactions on Pattern Analysis and Machine Intelligence, IEEE*, 27(3):418–433, 2005.

[9] Ramesh Raskar. Immersive planar display using roughly aligned projectors. In *Proceedings of Virtual Reality 2000*, pages 109–115, New York, NY, USA, 2000. IEEE.

[10] Ramesh Raskar, M. Brown, R. Yang, W.-C. Chen, G. Welch, H. Towles, B. Seales, and H. Fuchs. Multi-projector displays using camera-based registration. In *Proceedings of Visualization '99*, pages 161–168, 522, New York, NY, USA, 1999. IEEE.

[11] Ramesh Raskar, Jeroen van Baar, Thomas Willwacher, and Srinivas Rao. Quadric transfer for immersive curved screen displays. *Computer Graphics Forum*, 23(3):451–460, 2004.

[12] Ramesh 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: Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 179–188, New York, NY, USA, 1998. ACM Press.

[13] Rahul Sukthankar, Robert Stockton, and Matthew Mullin. Smarter presentations: exploiting homography in camera-projector systems. In *Proceedings of International Conference on Computer Vision, 2001*, 2001.

[14] George Vosselman, B.G.H. Gorte, G. Sithole, and T. Rabbani. Recognising structure in laser scanner point clouds. *International Archives of Photgrammetry, Remote Sensing and Spatial Information Sciences*, 46(8/W2):33–38, 2004.

[15] Ruigang Yang, D. Gotz, H. Towles J. Hensley, and M. Brown. Pixelflex: a reconfigurable multi-projector display system. In *Proceedings of Visualization 2001*, pages 167–174, 554, New York, NY, USA, 2001. IEEE.

[16] Ruigang Yang and Greg Welch. Automatic projector display surface estimation using every-day imagery. In *9th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2001*, 2001.

[17] M. Zuliani, C. S. Kenney, and B. S. Manjunath. The multiransac algorithm and its application to detect planar homographies. In *IEEE International Conference on Image Processing*, Sep 2005.

Limpid Desk: Transparentizing Documents on Real Desk in Projection-Based Mixed Reality

Daisuke Iwai* Osaka University
Sawako Hanatani† Osaka University
Chinatsu Horii‡ Setsunan University
Kosuke Sato§ Osaka University

ABSTRACT

Searching of spread documents on a real desk is an arduous task. The chaos on the desk makes users' searching of the desired documents difficult. In this paper, we propose Limpid Desk which supports a document search on a desk by transparentizing the upper layer of a document stack in projection-based mixed reality environments. In the system, the special pattern light which is calculated to compensate the appearances of the upper layer documents as if they are transparent is projected to the stack, and as a result, users can visually access to the lower layer document. We propose a touch sensing method using a thermal image for the input interface of the system. The method realizes that users' touch areas on real documents can be detected with no worn or hold devices. Hence, users can intuitively select the stack, which they would like to transparentize, by their simple touch gestures. We present three intuitive interaction techniques which allow users to figure out what the lower layer document is without physically removing its upper documents. Since the searching space, the manipulating space, and the display space are completely unified onto a real desk, users can directly access to the lower documents without PC monitor and interface. We claim that this kind of spatial consistency of the manipulation is the key for realizing the intuitive interaction.

CR Categories: H.5.1 [INFORMATION INTERFACES AND PRESENTATION (e.g., HCI)]: Multimedia Information Systems—Artificial, augmented, and virtual realities

Keywords: Projection-Based Mixed Reality, Radiometric Compensation, Smart Desk, Transparentized Documents, Thermal Image, Touch Sensing

1 INTRODUCTION

It is difficult to search a desired document on a real desk where many documents are multiply stacked. In contrast, in a PC's desktop, users can easily find the window which they would like to edit or see from all opening windows. For example, Application Switcher in Apple's Mac OS X allows users to access to the desired applications only by Command-Tab key. Furthermore, Mac OS X offers an intuitive window search interface, Exposé, by which users can see all opening windows immediately and find the desired one only by their one action. And, Microsoft announces that the glass-like interface elements that users can see the lower layer windows through its upper ones is planned to implement on the next OS, Windows Vista. We apply these methods to the real document search task.

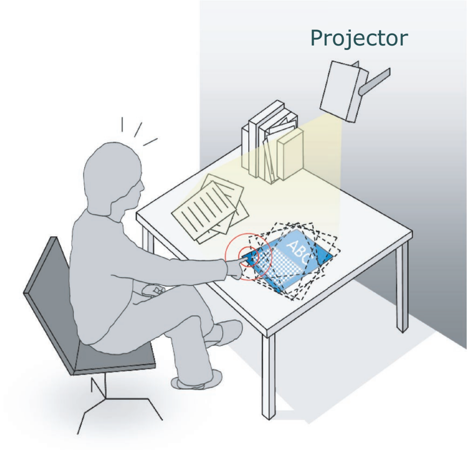


Figure 1: The design concept of Limpid Desk

In this paper, we propose an intuitive document search system, Limpid Desk, which utilizes a projection-based mixed reality (Fig. 1). In the system, a radiometric compensation technique is applied to transparentize the upper layer documents of the stacks, and users can visually access to the lower layer document without physically removing its upper ones. We also propose a touch sensing method, by which users' touch areas on real documents can be detected, using a thermal image for the input interface of the system. And then, we describe three intuitive interaction techniques such as users can browse and direct the whole stacked documents which they would like to transparentize or see only by their simple touch gestures.

2 TRANSPARENTIZING DOCUMENTS IN PROJECTION-BASED MIXED REALITY

Some researchers used projector to transparentize real objects [1, 2]. Their target objects are suitable for projection because they are high quality screens like a refrigerator's simple white door and a retro-reflective material. But, real documents like magazines and photos have spatially varying reflectance properties which modulate the appearance of the projected image. There are some works on controlling real objects' color appearances using the radiometric compensation method of a projection-based mixed reality technology to solve this problem [3, 4]. We apply these methods to transparentize upper layer documents which have complex textures.

Figure 2 shows the result of the transparentizing of the upper layer document by the radiometric compensation method. Figure 2a and 2b shows the document stack and the lower layer document before projection. Figure 2c and 2d shows the projection result when the original and compensated image of the lower layer document is projected to the upper layer document. It is confirmed that the appearance of the lower layer document is not reproduced in Fig. 2c because of the spatially varying reflectance properties of

---

*e-mail:iwai@sens.sys.es.osaka-u.ac.jp
†e-mail:hana@sens.sys.es.osaka-u.ac.jp
‡e-mail:chinatsu@kjo.setsunan.ac.jp
§e-mail:sato@sys.es.osaka-u.ac.jp

(a) Document Stack (b) The lower layer document

(c) After projecting the original image (d) After projecting the compensation image

Figure 2: Transparentizing the upper document with projection of compensating light

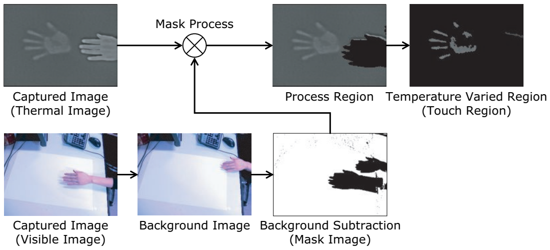


Figure 3: Flow of the touch area sensing

the upper layer document, but in contrast, it is almost reproduced in Fig. 2d. This result indicates that the projection-based mixed reality technology is able to almost transparentize real documents in limited applications of human interface.

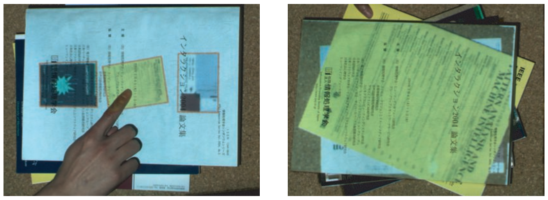
3 Touch Sensing Method

It is important to construct an input interface of the system which can detect users' touch areas on real documents without any user-worn or -hold devices, to realize seamless and consistent interactions as the users browse and direct stacked documents only by touching them. But, it has been a difficult problem to detect users' touch areas on real objects. In this paper, we propose a touch sensing way to realize such interface suitable for desktop environments using thermography.

The thermographic method uses the heat storage phenomena on users' touch areas on real objects. For example, when a user touches an object, his/her body heat is transferred to it. And then, when the user releases his/her hand from it, the heat remains at its surface for a while. The method measures this heat storage on the real object through a thermal image to detect the user's touch area.

A temperature varied area in the thermal image denotes a user's touch area. Particularly, not only a touch area but also a user's body itself causes a temperature varied area in the thermal image. We involve a CCD camera to use a visible image to eliminate the user's body area from the thermal image. The process flow is shown in Fig. 3. First, the background subtraction eliminates the user's body in the visible image and the mask image is created. And then, the temperature varied areas extracted from the masked thermal image is obtained as the user's touch areas.

(a) Touching the upper layer document (b) Projecting the image of the lower layer document to the touch area



(c) Selecting the thumbnail image of the desired document (d) Virtual pop-up displaying by touching the periphery of the desired document

Figure 4: Implemented interaction techniques

4 Interaction Techniques

Limpid Desk consists of a video projector, a CCD camera, and a thermo infrared camera. We implemented three types of interaction techniques for the system (Fig. 4). The first one is "Touch Area Transparent Interaction" in which user's touch areas on an upper layer document are transparentized (Fig. 4a, 4b). And the second one is "Thumbnail Type Document Search Interaction"; a user touches the upper layer document, then the thumbnail images of the all lower layer documents are projected to the upper layer document. And then, the user touches the thumbnail image of the desired document (Fig. 4c). As a result, the upper layer documents are transparentized and the user can see the desired lower layer document. The third interaction technique is "Direct Select Interaction" in which a user can directly point at the document they would like to see in the stack by touching it (Fig. 4d).

5 Conclusion

In this paper, we proposed Limpid Desk in which users can visually access to the lower layer of document stacks on a real desk by transparentizing the upper layer documents in the projection-based mixed reality. We also proposed the input interface which can detect users' touch areas on real documents. We described about the system and three intuitive interaction techniques on it. In future work, we evaluate the intuitiveness of the proposed interaction techniques and the effectiveness of the document search through the system.

References

[1] L. Bonanni and C.H. Lee. The kitchen as a graphical user interface. In *Proceedings of SIGGRAPH Electronic Art and Animation Catalog*, pages 109–111, 2004.

[2] M. Inami, N. Kawakami, and S. Tachi. Optical camouflage using retro-reflective projection technology. In *Proceedings of ISMAR*, pages 348–349, 2003.

[3] S. K. Nayar, H. Peri, M. D. Grossberg, and P. N. Belhumeur. A projection system with radiometric compensation for screen imperfections. In *Proceedings of PROCAMS*, 2003.

[4] T. Yoshida, C. Horii, and K. Sato. A virtual color reconstruction system for real heritage with light projection. In *Proceedings of VSMM*, pages 161–168, 2003.

14:00–15:00 Mini Tutorial

Multi-Projector Techniques for Real-Time Visualizations in Everyday Environments

Oliver Bimber, Bauhaus-University Weimar

Introduction

Immersive and semi-immersive projection displays, such as CAVEs, walls, workbenches, cylinders, and domes are being used to support virtual reality applications in many professional domains. The visualization of data with these displays, however, requires dedicated rooms for setting up non-mobile screens, and allows the interaction with purely synthetic information only.

Current research activities investigate conceptual and technical possibilities for realizing such visualizations in real world environments. They strive for enabling immersive and semi-immersive virtual reality, as well as augmented reality experiences without the need for special display surfaces or permanent screen configurations. Enabling visualizations without projection-optimized screens but on complex everyday surfaces, however, leads to several difficulties: The projected images are geometrically warped, color-blended, and regionally defocused. Furthermore, the application of multiple projectors requires geometric registrations, as well as luminance and chrominance matching and intensity fading to display a single consistent image. Thus, projected images have to be corrected in real-time and on a pixel-precise basis. In addition, other surface properties, such as depth information and global illumination effects can be extracted and human perception factors can be taken into account for improving the final image quality as well as enabling new possibilities. Monoscopic and stereoscopic graphics make semi-immersive, immersive and augmented visualizations in everyday environments possible—without the need for special projection screens.

Topics

This tutorial will give a technical introduction in state-of-the-art real-time multi-projector techniques and automatic projector-camera calibration methods that enable correct visualizations on complex screen surfaces. In particular it presents **geometric warping** of images projected onto non-planar surfaces, **radiometric compensation** for projecting onto colored and textured surfaces, **multi-focal projection** for creating focused images on surfaces with a large depth range, **reverse radiosity** for compensating global light scattering, and image-based techniques for supporting **view-dependent stereoscopic visualizations** on complex surfaces. It will be explained how these techniques are implemented on the GPU to achieve real-time performance and pixel-precision.

Tutorial Presenter

Oliver Bimber is a Junior Professor for Augmented Reality at the Bauhaus University Weimar, Germany. He received a Ph.D. in Engineering at the Technical University of Darmstadt, Germany. From 2001 to 2002 Bimber worked as a senior researcher at the Fraunhofer Center for Research in Computer Graphics in Providence, RI/USA, and from 1998 to 2001 he was a scientist at the Fraunhofer Institute for Computer Graphics in Rostock, Germany. He received the degree of Dipl. Inform. (FH) in Scientific Computing from the University of Applied Science Giessen and a B.Sc. degree in Commercial Computing from the Dundalk Institute of Technology. In his career, Bimber received several scientific achievement awards and is author of more than forty technical papers and journal articles. He is author of the book "Spatial Augmented Reality" (together with Ramesh Raskar) and serves on the editorial board of the IEEE Computer magazine (graphics and multimedia editor). Bimber also gave a number of guest lectures at recognized institutions. Among them were Brown University, Princeton University, the IBM T.J. Watson Research Center, Osaka University, the DaimlerChrysler Virtual Reality Competence Center, and the Mitsubishi Electric Research Lab (MERL). His research interests include next-generation display technologies, real-time rendering and computer vision. Bimber is member of IEEE, ACM, ACM Siggraph, and Eurographics.

15:30–17:10 Session IV

Handheld Augmented Reality Displays

Daniel Wagner* and Dieter Schmalstieg

Graz University of Technology, Austria

Figure 1: Two examples of collaborative games using handheld augmented reality displays. Left: The Invisible Train, Middle and right: Mr. Virtuoso teaches arts history

1 INTRODUCTION

Augmented Reality (AR) can naturally complement mobile computing on wearable devices by providing an intuitive interface to a three-dimensional information space embedded within physical reality. However, existing AR systems like MARS [1] or Tinmith [2], which require a user to wear a notebook computer in a backpack and a head-mounted display (HMD) are expensive, fragile and inconvenient to wear. Thin-client approaches using a Tablet PC or Personal Digital Assistant (PDA) merely as a portable display [3][4] require a dedicated server infrastructure and limit mobility.

We believe there is a need for an unconstrained, infrastructure-independent AR display running to fill the gap in situations where traditional backpack systems are too costly and cumbersome, but thin client implementations exhibit inadequate deployability, scalability or interactive behavior. Particular examples include sporadic use over lengthy time spans, in between which devices must be stowed away, mixed indoor/outdoor use in wide-area environments, and massively multi-user application scenarios. This has motivated us to develop a state of the art AR framework targeting lightweight handheld displays.

2 TECHNOLOGY

Platform. Currently there are three distinct classes of commercially available wearable computers as potential candidates for a standalone handheld AR display: cellular phones, PDAs and Tablet PCs. All of these device designs make specific trade-offs between size, weight, computing power and cost. While cellular phones are extremely portable and widespread, their current lack of adequate processing power and local network connectivity renders them a suboptimal platform for rich and meaningful AR applications. Furthermore, their small display size and limited data input capabilities are less than ideal for three-dimensional user interfaces. Tablet PCs do not share the aforementioned drawbacks but are considerably more expensive and too heavyweight for single handed, or even prolonged two handed use. Taking into account all of these constraints, we chose the PDA as target platform for our handheld AR display. PDAs are a good compromise between processing power, size and weight; they are socially acceptable and their touch screen input mechanism is familiar, which we considered crucial for our planned deployment of AR applications to untrained users.

Unfortunately, the software development infrastructure available for embedded platforms such as Windows CE does not match the standards usually available for the developments of highly interactive 3D graphics applications on PCs. The scarce resources of the PDA make it necessary to resort to cross-development using a suitable cross platform abstraction layer, for example to access frame buffer, windowing system and camera devices. The latest version of our AR framework Studierstube facilitates such cross-platform development, and allows to perform the majority of the development work more conveniently and efficiently on a regular PC, leaving only final testing and performance optimization to be carried out on the target device.

3D rendering. Computer graphics researchers are accustomed to working with standardized low-level APIs such as OpenGL, but handheld devices provide hardware support only for the more constrained OpenGL ES standard to overcome this gap we have developed a library called Klimt that wraps an underlying OpenGL ES implementation and adds the most important missing OpenGL and WGL features such as floating-point data types, missing primitive types, long integer index types, etc. Using Klimt, many existing OpenGL applications require only minimal modifications in order to run hardware-accelerated on a PDA. Internally, Klimt is implemented as a collection of C++ classes with a C wrapper for OpenGL API compatibility. As a means to optimize runtime performance Klimt makes heavy use of templates and inline functions.

Scene graph. Our AR framework Studierstube does not directly interface with OpenGL, but instead builds on top of the open source scene graph library Coin. Open Inventor allows rapid prototyping of data-driven applications, which can be written in C++ or in a scripting language. We ported the Coin rendering library, which implements the Open Inventor API, to Windows CE, running on top of OpenGL/Klimt.

To make the use of a general purpose scene graph on a comparably low performance handheld device more satisfactory, a number of more advanced optimizations had to be applied. Many parts of a scene graph remain unchanged over a longer period,

---

* wagner@icg.tu-graz.ac.at

which is why most scene graph libraries use a so called primitive cache that stores data in a format optimal for the underlying graphics hardware. This primitive-cache was extended to support fixed-point formats and therefore provide most graphics data in the native format of mobile graphics accelerators. To save memory which is a scarce resource on embedded devices we removed several less used features of Coin which resulted in a significantly reduced memory footprint. Windows CE is very restricted with respect to virtual memory, and we therefore outfitted all Studierstube libraries including Coin with a custom memory manager that can allocate memory outside the process' memory slot and allows fine grained monitoring on how much memory is spent in which parts of an application.

Tracking. The tracking requirements of AR can be provided by vision tracking using the built-in camera of the handheld device. The well-known marker tracking library ARToolKit was used as the starting point for developing a significantly modified and optimized tracking solution running on both PCs and handheld devices. ARToolKitPlus can track at a top speed of 6.5ms per image on a current PDA, so self-contained vision tracking is no longer a major bottleneck for embedded devices. We replaced the original checkerboard template-matching with CRC-encoded markers in ARToolKitPlus, which are always detected faster and allow up to 4096 different targets.

To compensate for ever-changing lighting conditions when using mobile devices, we added a real-time automatic thresholding algorithm. Moreover, the software can compensate for radial brightness falloff which is typical for the low-quality lenses used in the mobile cameras.

3 APPLICATIONS

In contrast to AR applications with tangible user interfaces, such as the Magic Book [5], which focus on the use of fiducial markers as moveable, dynamic parts of the application, we sought to employ the handheld's tracked display itself as the tangible, dynamic interaction element. Therefore, we decided to focus on pen-based touch-screen input as the main interaction technique. A static arrangement of multiple fiducial markers in the environment enables the PDAs to perform self-tracking from a variety of angles and distances.

Another important requirement was that the application should be sufficiently spatially distributed to give an impression of the properties of our tracked display surface with respect to panning and zooming interactions — users should be required to move in closely with environment to discover important details, and to move the perspective away from the setting in order to gain an overview of the scene. This differs from other applications such as the magic book, which are designed to be fully visible within the field of view of the user, and therefore require no navigational actions from the user.

Usually, marker-based tracking techniques are sufficiently accurate for computing a marker's position and orientation relative to a camera, but not to perform inside-out tracking of the camera in relation to the environment. We therefore make heavy use of multi-marker tracking to overcome this limitation. Using all visible markers in the camera image provides improved stability in tracking of the device with respect to the environment.

The Invisible Train [6] (see left picture in Figure 1) is a simple multi-player game, in which players steer virtual trains on a real wooden miniature railroad track. Using the touch-screen players can influence the game by operating track switches and adjusting the speed of their trains. The aim of the game is to prevent the two trains from colliding as long a possible. The game correctly blends all virtual objects correctly with the real environment: virtual objects can occlude real ones and vice versa.

Virtuoso is a collaborative educational game for up to four players. The players have to sort a collection of artworks by date of creation along a virtual timeline on a real wall (middle picture in Figure 1). Players can take the 3d artworks from the wall an onto their PDAs to gather more information, then drop them again on another spot. The multi-lingual art history expert Mr. Virtuoso can be queried for advice by placing the artwork on his desk (right picture in Figure 1).

4 EXPERIENCES AND DISCUSSION

Comparing our observations from public trials with previous experiences using HMD + backpack setups confirmed our assumption that handheld AR displays are more accessible to a general public, and exhibit better learning curves than traditional mobile AR systems: We found that visitors had little to no reservations against using our system. Several participants figured out how to play the Invisible Train by simple trial and error, bystanders learned the gameplay by observation while waiting.

Consequently, our supervision overhead was considerably lower than expected. We frequently observed unsupervised user experiences as visitors passed around the PDAs while explaining the game to each other. Most participants played at least a single game (averaging roughly 60 seconds) before handing their PDA to the next visitor.

In contrast to backpack setups, we experienced almost no hardware related failures, except when users deliberately removed the plug-in camera. We conclude that wearable devices intended for a general public must be self-contained units without any loosely attached parts. They should also automatically restart in case of a provoked or accidental failure.

According to user feedback, our application was considered sufficiently responsive for the intended type of spatial interaction. Only a negligibly small fraction of players felt their PDA's display update rate and delay impairs their ability to play the game. For illustration, the latest version of Virtuoso runs at ~13Hz on a Dell Axim X51v.

Based on our experiences we believe that handheld AR displays are an important technology for making AR accessible to a large audience, and in situations that can truly be called mobile. If done right, AR can work very well on today's handheld devices. AR should therefore be very well positioned to become an important user interface style on everyday consumer equipment.

REFERENCES

[1] Höllerer, T., Feiner, S., Terauchi, T., Rashid, G., and Hallaway, D. Exploring MARS: Developing Indoor and Outdoor User Interfaces to a Mobile Augmented Reality System. *Computers and Graphics, 23(6), Elsevier Publishers, Dec. 1999*, pp. 779-785, 1999

[2] Piekarski, W. and Thomas, B. Tinmith evo5 - An Architecture for Supporting Mobile Augmented Reality Environments. *2nd Int'l Symposium on Augmented Reality*, pp 177-178, New York, Oct.2001

[3] Gausemeier, J., Fruend, J., Matysczok, C., Bruederlin, B., and Beier, D., Development of a real time image based object recognition method for mobile AR-devices. *Proceedings of the 2nd international conference on Computer graphics, Virtual Reality, visualisation and interaction in Africa*, pp. 133-139, 2003

[4] Ingram, D., Newman, J., Augmented Reality in a Wide Area Sentient Environment. *Proceedings of the 2nd IEEE and ACM International Symposium on Augmented Reality (ISAR 2001)*, October 2001, New York

[5] Billinghurst, M., Kato, H., Poupyrev, I., The MagicBook: A Transitional AR Interface. *Computers&Graphics 25* (2001) 745–753

[6] Wagner, D., Pintaric, T., Ledermann, F., Schmalstieg, D., Towards Massively Multi-User Augmented Reality on Handheld Devices, Proceedings of Third International Conference on Pervasive Computing, Pervasive 2005, May 2005, pp. 208-219

“Personal Practically Panoramic” Multimodal Interfaces

Kayoko Kanno, Newton Fernando, Alam Bolhassan, Sho Narita, & Michael Cohen

Spatial Media Group, University of Aizu
Aizu-Wakamatsu, Fukushima-ken 965-8580; Japan
{m5091104, mcohen}@u-aizu.ac.jp

ABSTRACT

We have developed second-generation prototypes of the Internet Chair, a novel internet appliance. The first generation explored using the chair as an input device; “$^{S}{}_{c}$ha$_{i}$r^{e},” the prototype employed here, is a pivot (swivel, rotating) chair deployed as an output device, a rotary motion-platform information appliance, dynamically aligning haptic display with wireless visual displays and spatial audio in rotation-invariant virtual spaces. As a haptic output modality, chairs with servomotors render kinesthetic and proprioceptive cues, twisting under networked control, to direct the attention of a seated subject orienting seated users like a “dark ride” amusement park attraction or under active user control, local or distributed. Using its audio display modality, “nearphones” embedded in the seat headrest, the system can present unencumbered binaural sound with soundscape stabilization for multichannel sound image localization. In groupware situations like teleconferencing, chat spaces, or multiplayer gaming, such orientation is also synchronized with panoramic or turnoramic displays or twisting iconic representations of the users, avatars in virtual spaces, enabling social situation awareness.

The $^{S}{}_{c}$ha$_{i}$r^{e}, manifesting as personal LBE (**l**ocation-**b**ased **e**ntertainment), can be used in both stand-alone and networked applications. We have developed several clients that exploit such “practically panoramic” capability, including simulators, games, and 360° browsers, providing sensory-integrated multimodal applications, variously including stereographic or mobile features.

Additional Keywords:
{augmented, enhanced, hybrid, mediated, mixed} reality/virtuality, haptic interface, information furniture, **l**ocation-**b**ased **e**ntertainment (LBE), motion platform, networked appliance, soundscape stabilization.

1 INTRODUCTION

There are more chairs in the world than windows, desks, computers, or telephones. According to a metric of person-hours used, and generalized to include couches, stools, benches and other seat furniture, the chair is the most popular tool on earth, with the possible exceptions of its cousin the bed, and eyewear. This research belongs to fields variously described as or associated with ambient computing, calm technology, and ubicomp.

We are developing second-generation prototypes of the Internet Chair (Koizumi et al., 2000; Cohen, 2003), a novel internet appliance. The first generation prototype explored using the chair as an input device. “$^{S}{}_{c}$ha$_{i}$r^{e}” (for ‘shared chair’), the prototype extension described here, is a pivot (swivel, rotating) chair deployed as an output device, a rotary motion-platform information appliance. Its haptic display modality is yaw, dynamically synchronizable with wireless visual displays and spatial audio in a rotation-invariant virtual space.

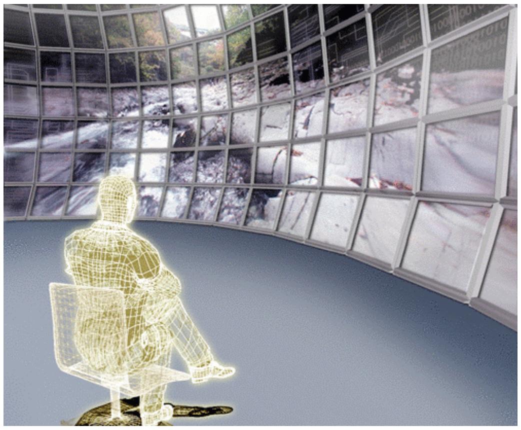


Figure 1: Inspiration (Graphic by “Eyes, Japan”)

In groupware situations— like teleconferencing, chatspaces, or multiplayer gaming— such orientation can also be used to twist iconic representations of a seated user, avatars in a virtual world, enabling social situation awareness.

As an alternative to transaural loudspeakers (providing crosstalk-cancelled binaural cues), speaker arrays, and normal headphones, we are using “nearphones,” external loudspeakers placed near but not on the ears, straddling the headrest of a chair: a hybrid of circumaural headphones (which block ambient sounds) and loudspeakers, as shown in Figure 2. Using its audio display modality, the system can present unencumbered binaural sound with soundscape stabilization for multichannel sound image localization.

As a haptic output modality, chairs with servomotors render kinesthetic and proprioceptive cues, twisting under networked control, to direct the attention of a seated subject, orienting seated users like a “dark ride” amusement park attraction or under active user control, local and/or distributed. The $^{S}{}_{c}$ha$_{i}$r^{e}, manifesting as personal LBE (**l**ocation-**b**ased **e**ntertainment), can be used in both stand-alone and networked applications.

We are developing various multimodal “personal practically panoramic” interfaces that exploit some unique features of this networked rotary motion-platform: spatial audio renderers; chromastereoptic and stereographic image-based and synthetic CG renderers, mobile phone interfaces, including both session- and individual widgets; SQTVR (**s**tereographic **QTVR**) panorama and turnorama browsers; a multiplayer shooting game; and a driving simulator with audio way-finding.

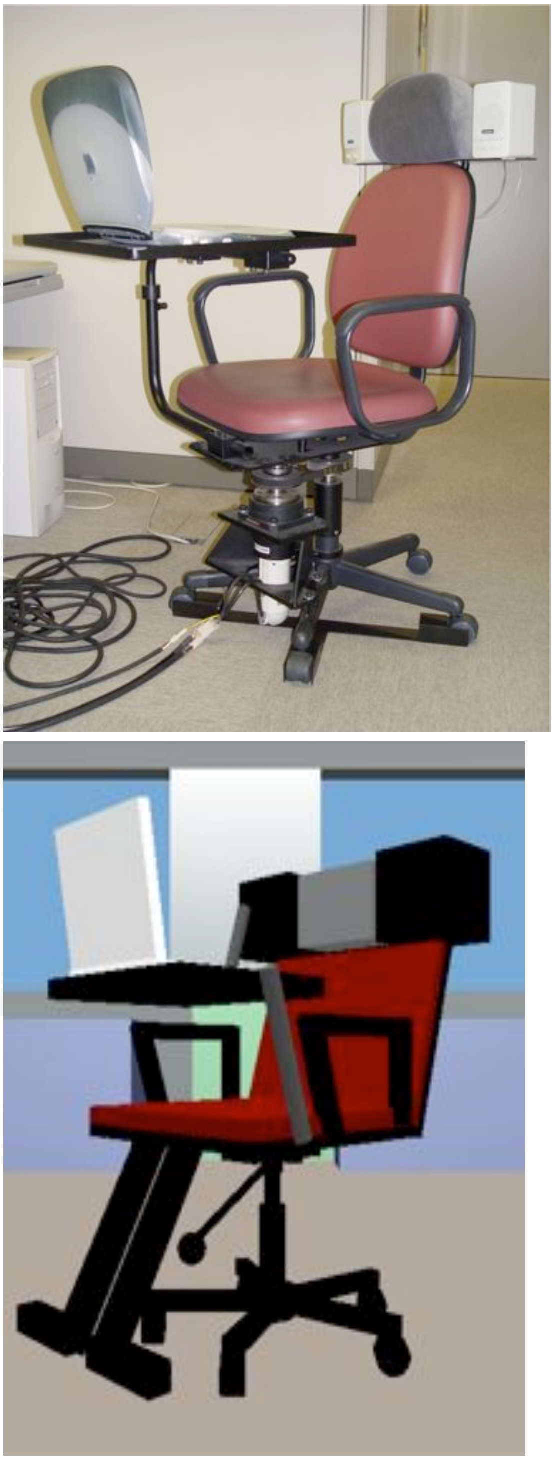


Figure 2: For its haptic output modality, servomotors render kinesthetic force display, rotating each $^{S}{}_{c}$ha$_{i}$r^{e} under networked control. Note the nearphones straddling the headrest.

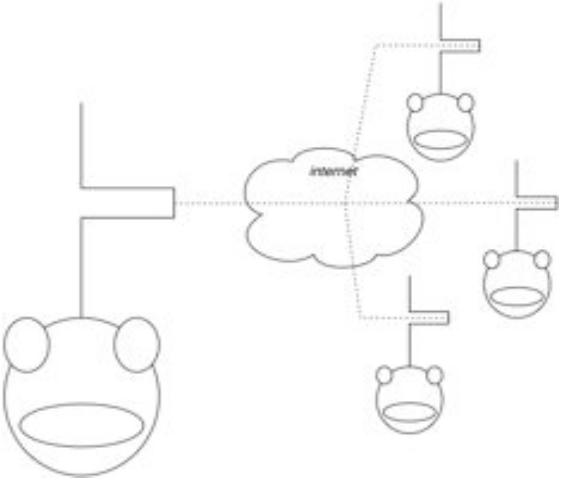


Figure 3: Simplified concept: nearphone-equipped rotating seats exchanging torque via networked mutual tractor beams

2 IMPLEMENTATION

2.1 Session-Integrated Multimodal I/O Clients

We have designed and implemented an architecture and framework to support collaborative virtual environments (CVEs) allowing distributed users to share multimodal virtual worlds. Our CVE architecture (as shown below in Figure 4) is based upon a client/server (C/S) model, and its main transaction shares the state of virtual objects and users (avatars) by effective multicast via replicated-unicast of position parameters (translation, rotation, and zoom) to client peers in a session. The client/server architecture integrates multimodal control/display capabilities and clients—including the haptic renderer, spiral visualizers and intensity panners, visual perspective and spatial sound renderers—described following.

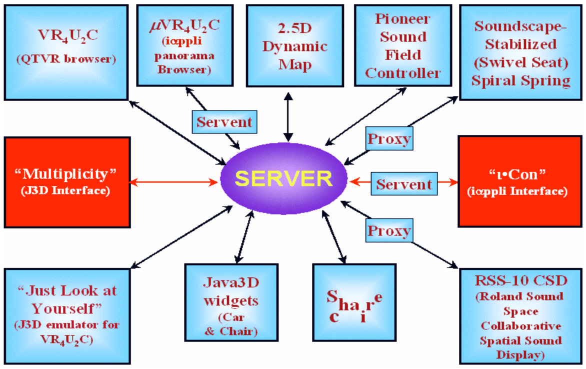


Figure 4: Collaborative Virtual Environment Architecture: Groupware Suite

2.2 Azimuth-display output modality

Our second-generation prototype— developed with partners Mechtec (www.mechtec.co.jp) in Kita-Kata, Eyes (www.aizu.com) in Aizu-Wakamatsu, and Yamagata University in Yonezawa (www.yz.yamagata-u.ac.jp)— features a powerful (with about 3–4 Newton·meters of torque, adjustable to limit maximum speed and force) servomotor for force display and computer-mediated rotation of the chair. Each chair stimulates a visceral sensation as it whirls around to direct the attention of a seated subject with adjustable insistence/forcefulness— imperatively rotating like a "dark ride" amusement park attraction, or subtly nudging the user in a particular direction. In practice, each $^{S}{}_{c}ha_{i}r^{e}$ uses two session channels, one to track its realtime orientation and one to anticipate its rotational target. The heterogeneous clients in our multimodal CVE groupware suite interoperate seamlessly. Of particular relevance is a computer graphic rendering of a space, allowing various camera positions, including endocentric (1st-person: from the point-of-view of the avatar), egocentric or tethered (2nd-person: attached to but separate from the avatar), and exocentric (3rd-person: totally detached from the avatar) perspectives, like that in the bottom of Figure 2. For cable-less operation necessitated by the spinning chair, these clients run on laptop computers networked via Wi-Fi. (In our lab we use various Mac iBooks and Powerbooks with the IEEE 802.11 AirPort wireless option.)

3 APPLICATIONS

3.1 S^6: Soundscape-Stabilized Swivel-Seat Spiral-Spring

A GUI that displays and controls the azimuth of the $^{S}{}_{c}ha_{i}r^{e}$ using a spiral spring metaphor, as shown Figure 6, and also allows positioning of audio sources, directionalized from resident sound files or captured in realtime from analog streams. To enable mobile control of the $^{S}{}_{c}ha_{i}r^{e}$, we have developed an equivalent graphical interface, shown in Figure 5.

Figure 5: "*i*•Con-s" interface on DoCoMo iαppli

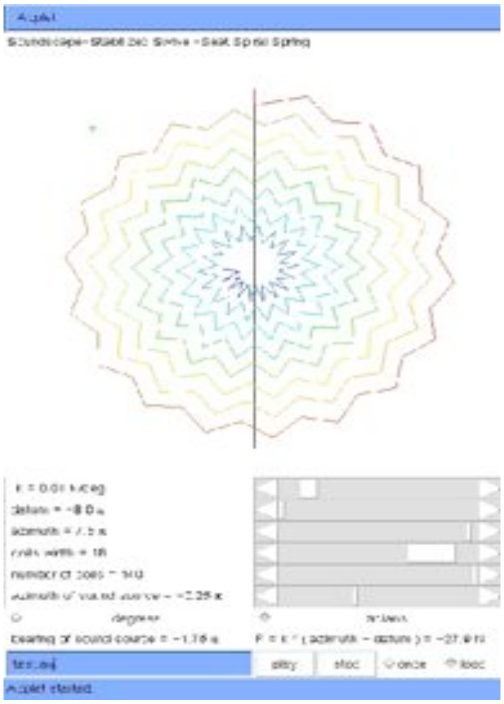


Figure 6: Soundscape-stabilized spiral-spring swivel seat interface. A spiral spring metaphor is used to display the yaw of the chair (as indicated by the total deflection of the spring) and azimuth of a virtual sound source (as indicated by the plus sign, set here on the upper left at -45°=NW). The mutual displacements are used to control intensity stereo panning for playback of audio files and streams.

3.2 Driving Simulator with Audio Way-Finding

We have developed a networked driving simulator (Adachi, Iwai, Yamada, & Cohen, 2005) as a virtual-reality based interface (control/display system) featuring integration with the $^{S}{}_{c}ha_{i}r^{e}$ rotary motion platform for azimuth-display, stereographic display for 3D graphics, and spatial audio (sound spatialization) way-finding cues, configured with appropriate controls, namely a brake and accelerator and battery-powered force-feedback USB driving controller. A 'simplex' mode couples the local control and display, while an alternative 'duplex' mode disables such immediacy, relying instead upon returned network events to update the visual and displays. This scheme accommodates network delays and client latency, synchronizing the multimodal display. For particular instance, the $^{S}{}_{c}ha_{i}r^{e}$ Internet Chair has significant sluggishness, a consequence of mechanical inertia (seatee payload) and user comfort.

3.3 VR$_4$U$_2$C Interface for SQTVR

We have integrated the $^{S}{}_{c}ha_{i}r^{e}$ with "VR$_4$U$_2$C" ("virtual reality for you to see") our QTVR image-based rendering client (Bolhassan, Cohen, & Martens, 2004), as shown in Figure 7. This multiuser multiperspective panoramic and object movie (turnorama) browser was developed using Apple's QuickTimeVR technology and the Java programming language with the support of the "QuickTime for Java" application programming interface (API). It allows coordinated display of multiple views of a virtual environment, limited practically only by the size and number of monitors or projectors available around users in various viewing locations. VR$_4$U$_2$C, can be used interactively to explore and examine detailed multidimensional, virtual environments (photorealistic or otherwise) using a computer and conventional input devices— including mouse, trackball, rotary controller, track pad, and keyboard— as well as more exotic interfaces— including speaker array spatial audio displays, mobile phones, and swivel chairs with servomotor control. Through a unique multinode dynamic traversal technique, VR$_4$U$_2$C provides an elegant solution to the challenge of interactively track- and dolly-able stereoscopic display of QTVR imagery, as shown in Figure 8.

Figure 7: VR4U2C Interface: panoramic and turnoramic browser

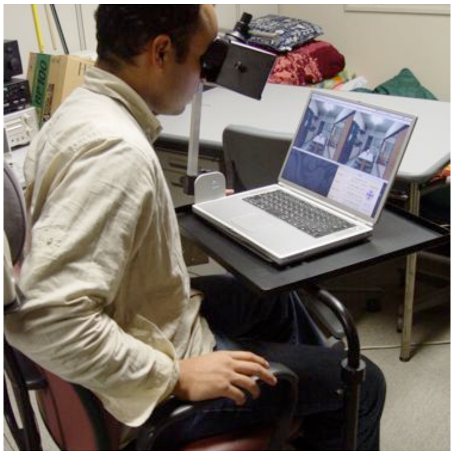


Figure 8: Stereoscopic display of QTVR imagery. A nomadic version, “μVR4U2C,” platformed on the Sharp SH25liS, provides an autostereoscopic equivalent for mobile phones.

3.4 “*Kuru-kuru* Pitcher” Multiplayer Shooting Game

We have developed a multiplayer game that exploits some unique features of our networked rotary motion-platform, loosely resembling a disk/disc access driver, in which “spindled” players race to acquire circularly arrayed dynamically arriving targets. We PLAN to extend the groupware capabilities of the $^{S}{}_{c}\mathrm{ha}_{i}\mathrm{r}^{e}$ beyond the two-person game (Adachi, Cohen, Dumiduwardena, & Kanno, 2004). As shown in Figure 3, an arbitrary number of similarly equipped chairs can be networked, with application-determined distribution (linkage/coupling) of cybernetic torque and arbitrary C/D ratios, fan-out, etc. of the Internet Chair with explicit haptic display.

4 CONCLUSION AND FUTURE RESEARCH

We have developed our rotary motion platform clients to support various azimuthal interfaces, including SQTVR, stereographic Java3D scenes, driving simulators with way-finding, mobile interfaces, and **l**ocation-**b**ased **e**ntertainment. Synchronizing panning graphics and spatial sound with propriocentric sensation enables a “personal practically panoramic” multimodal interface. To use the $^{S}{}_{c}\mathrm{ha}_{i}\mathrm{r}^{e}$ as a conferencing platform, we will deploy microphones for voice input, configure wireless network audio streaming communication, and extend our narrowcasting protocol to encompass the chair's audio capability, as outlined above. We plan to use SIP (Session Initiation Protocol) to establish realtime multimodal conferences (Alam, Cohen, & Ahmed, 2005).

REFERERNCES

[1] Kazuya Adachi, Michael Cohen, Uresh Duminduwardena, and Kayoko Kanno. “*Kuru-kuru* Pitcher”: A Game for the $^{S}{}_{c}\mathrm{ha}_{i}\mathrm{r}^{e}$ Internet Chair. In Matthias Rauterberg, editor, Proc. ICEC: Third Int. Conf. on Entertainment Computing, volume 3166 of Lecture Notes in Computer Science, pages 35–45, Eindhoven, Netherlands, September 2004. Springer. ISBN 3-540-22947-7.

[2] Kazuya Adachi, Ken'ichiro Iwai, Eiji Yamada, and Michael Cohen. Multimodal wayfinding in a driving simulator for the $^{S}{}_{c}\mathrm{ha}_{i}\mathrm{r}^{e}$ Internet Chair, a networked rotary motion platform. Proc. ICEC: Int. Conf. on Entertainment Computing, pages 511–514, Kobe-Sanda, Japan, September 2005. ISBN 3-540-29034-6.

[3] Mohammad Sabbir Alam, Michael Cohen, and Ashir Ahmed. Design of Narrowcasting Implementation in SIP. Proc. HC-2005: Eighth Int. Conf. on Human and Computer, pages 255–260, Aizu-Wakamatsu, Japan, August 2005.

[4] Alamshah Bolhassan, Michael Cohen, and William L. Martens. VR$_4$U$_2$C: A Multiuser Multiperspective Panorama and Turnorama Browser Using QuickTime VR and Java Featuring Multimonitor and Stereographic Display. TVRSJ: Trans. Virtual Reality Society of Japan, 9(1):69–78, March 2004. ISSN 1344-011X.

[5] Michael Cohen. The Internet Chair. IJHCI: Int. J. of Human-Computer Interaction, 15(2):297–311, 2003. ISSN 1044-7318. www.u-aizu.ac.jp/~mcohen/welcome/publications/ic4.ps

[6] Uresh Chanaka Duminduwardena and Michael Cohen. Control System for the $^{S}{}_{c}\mathrm{ha}_{i}\mathrm{r}^{e}$ Internet Chair. Proc. CIT: Int. Conf. on Computer-Information Technology, pages 215–220, Wuhan, China, Sept. 2004.

[7] Nobuo Koizumi, Michael Cohen, and Shigeaki Aoki. Japanese patent #3042731: Sound reproduction system, March 2000.

CupHolder: A Multi-Person Interactive High-Resolution Workstation

Mark Hereld* Michael E. Papka* Justin Binns* Rick L. Stevens*
Argonne National Laboratory, University of Chicago, and the Computation Institute

ABSTRACT

Demand for high resolution visualization, large pixel real estate collaborative workspaces, and interactive computer interfaces continue to drive researchers to develop new physical portals connecting them to their computational tools, to their data, and to their colleagues around the world. In this paper we describe CupHolder, a high performance workstation designed to support interactive collaboration and research activities. It is configured to enable display of high resolution imagery while enabling a comfortable interactive environment for one to several co-located researchers. Moreover it is driven by a high performance commodity cluster that provides substantial local rendering muscle as well as a high performance interface to Grid-based computational tasks. CupHolder is comprised of commodity components. It is primarily novel because it represents an integration of these components into a new form factor that we believe is a useful precursor and testbed for future integrated workspaces.

CR Categories: B.4.2 Input/Output Devices – Image display, H.5.2 User Interfaces – Input devices and strategies, H.5.2 User Interfaces – Interactive styles, H.5.3 Group and Organization Interfaces—Computer-supported cooperative work, I.4.9 Image Processing and Computer Vision – Applications, C.0 General – Hardware/software interfaces

Keywords: tiled displays, human computer interface

1 INTRODUCTION

Commodity display and computation resources play an increasing role in scientific research at all levels. Many approaches to aggregating these resources have been applied to the task of creating high performance facilities in the form of large cluster computing systems and large format tiled displays.

Creating large format high-resolution display systems has a relatively long history which includes much work on projection-based tiled displays [9]. For a summary of issues and pointers to relevant research see Hereld et al. (2000) [3]. However, a path to higher pixel density systems leveraging now ubiquitous LCD flat panel display technology has become increasingly appealing [7]. Displays of this kind with as many as 100 Mpixels (LambdaVision) have been developed within the OptiPuter project [8]. A table format display prototype with a clear plastic protective sheet has also been developed by the same project, the LambdaTable. These often enormous displays are driving development of new technologies for content delivery over very high bandwidth networks [6]. The table format displays invite interactive styles that include gestural input, as well as the inclusion of physical objects to bridge the gap between the real and virtual [5]. Many interesting variations on these themes have been explored. The Magic Table combines projector, camera, machine-vision and physical tokens on a table surface to create a multi-person interactive whiteboard [2]. And the Escritoire is a multi-resolution projection-based personal writing table [1]. None have folded the high performance visualization display capabilities into the personal interactive workspace.

Figure 1: Our StationOne concept drawing from 1998. It was a combination of functional display regions driven by groupings of computers selected for best support of the various functions comprising the workstation: graphics, control, and interface. In addition the frame was conceived as a flexible substrate for customization by the user – placement of user devices, surfaces and hooks for personal effects.

We have long been experimenting with display system and workstation form factors in search of designs that will naturally support the high performance demands of today's scientific applications (Fig. 1). In addition to high resolution, vast numbers of pixels provide a canvas for support of interactive collaborative environments enabling the side-by-side presentation of complex simulation results and video streams of distant collaborators. A progression of devices and technologies map the arc of our work to date: ActiveMural, μMural, and the AccessGrid. With the development of Grid infrastructure in mind, we have become keenly interested in developing a prototype interface to potentially huge computational and data resources. In particular, we seek an instance of such an interface with a form factor that is suitable for an office.

In this paper we describe a new configuration of commodity components which targets high performance display and computing while supporting collaborative work styles and enabling interactive input modalities based on computer vision techniques. In the following section we will describe the design of the CupHolder. In section 3 we will discuss applications that

---

*e-mail: {hereld, papka, binns, stevens}@mcs.anl.gov

have debuted on it. And in section 4 we will add observations and comments about the future of this design point.

2 CUPHOLDER DESIGN

In this section we will discuss the various aspects of the CupHolder design (Fig. 2), what motivated our choices, and how these were ultimately implemented.

Visualization Field. The main surface for display of demanding visualizations and multimedia content is provided by the 3 x 3 array of Dell UltraSharp 2001FP flat panel displays (20.1 inch, 1600 x 1200 pixels). The aggregate 4800 x 3600 pixel surface is bright and compelling. Each display panel is driven by a PC (3.6 GHz P4, 1MB cache, 2 GB ram, 750 GB disk, on-board copper gig-e network) with an nVidia GeForce 6600GT 128MB graphics card in the PCI-E x16 slot. We selected this box for its form-factor – the Shuttle XP is about 8 inches wide, 8 inches tall, and 12 inches deep – so that we could house them within the envelope of the CupHolder frame.

Console and Table. To this field of pixels we added two multi-purpose functional units. The console field is provided by 3 of the same flat panel displays at a 45 degree angle adjoining the bottom of the visualization field. The table field is provided by another 3 LCD panels lying horizontally and adjoined to the bottom of the console field (Fig.2). This substantial dollop of pixels is driven by a single Dell XPS 600 workstation (3.2 GHz Pentium D) with two dual-headed 256MB nVidia 7800 GTX adaptors in the two 16x PCI-E slots of the box and a single nVidia GeForce FX 5200 Ultra in a PCI slot. It has 4GB of ram and 1 TB of disk. These two ranks of displays are covered by a cold formed transparent polycarbonate cowling that protects the six displays from spilled liquids and damage from objects place on it (keyboards, mice, coffee cups).

Human Factors. In designing the CupHolder we wanted a system that would fit in an office, serve one to several people, and encourage a comfortable work style. To this end, we decided on that it should be used principally while standing, mainly to support several people. The integrated horizontal surface provides a convenient resting place for personal effects, a keyboard, and future tangible user interface devices.

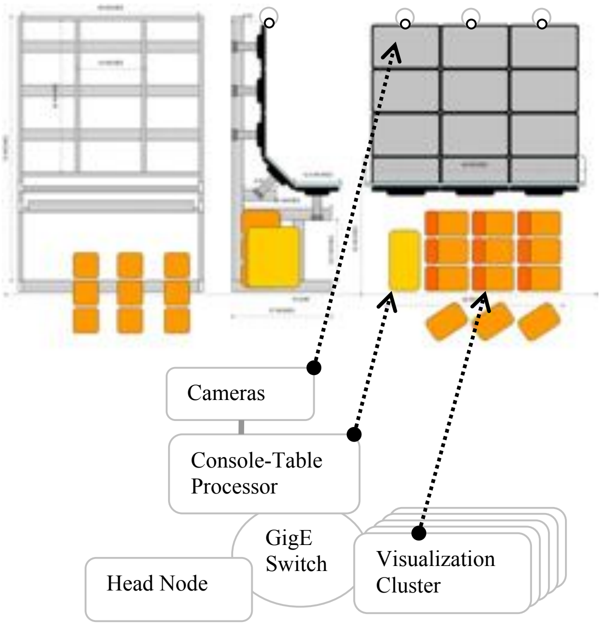


Figure 2: Diagram of the CupHolder. The 3 x 3 array of LCD flat panel displays are driven by a cluster of PCs. The console bank of 3 displays and the table bank of 3 displays are driven as a single desktop by a 6-headed PC. Three USB cameras provide overlapping coverage of the console, table, and surrounds from above.

Interactive Interface Testbed. The volume above the table and the space in front of the visualization field are targets for gestural input (Fig. 3). Cameras (two or three) are stationed above the CupHolder and casually positioned to take in the target volume. Overlap in the field-of-view of these provides for modest positional resolution in the vertical direction.

Figure 3: The CupHolder. The angled console and horizontal table surfaces both provide substantial pixel real estate for application control interfaces that may involve gesture recognition.

3 APPLICATION EXPERIMENTS

We have engaged the CupHolder in a number of tests to see how it performed and how it felt for a number of use cases. All of these application experiments were carried out informally at SuperComputing 2005, on its maiden voyage.

Monolithic Visualization. In this use case a single application drives the cluster, using Chromium [4] or homegrown solutions, to produce a visualization that covers the 3 x 3 array of LCDs as if they were a single panel. In this environment the user is usually interested in the large number of pixels to provide a canvas on which a large number of data points can be mapped. The mullions provide a windowed view into the visualization that users feel comfortable with, much like looking out a normal window. For many situations care needs to be taken that the mullions appear as obstacles in the view path and not simply as discontinuities to enhance the sense of realism for the user.

Distributed Simulation Monitor. Here, the application drives the panels separately, deriving rendered results from processes running on Grid computing resources around the country (Fig. 4). The central tile shows overall progress of the simulation with a composite view of the entire visualization. This mode of operation can be combined with the approached described above merging multiple panels to form a larger high-resolution view. This mode of use provides an effective way to monitor a large number of different events, be they the results of a simulation, the results of a simulation from multiple views, or using different visualization methods for the same data. It can be used to provide a method of presenting diagnostics of how the simulation is performing or other information about the environment in which the simulation is being performed.

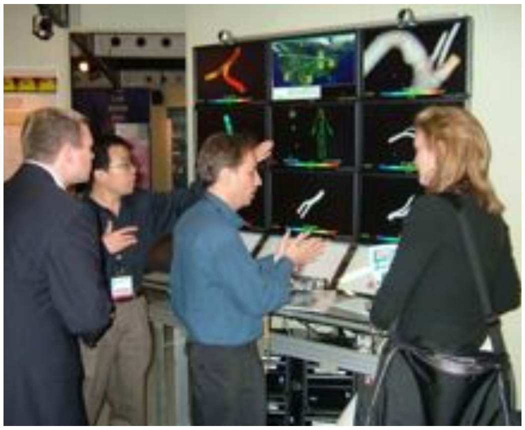


Figure 4: The CupHolder as a demonstration vehicle at SuperComputing. This application shows CupHolder in its capacity as high performance visualization front end to a large distributed simulation run on Grid computers at several locations around the country and one in the United Kingdom.

Cartographic Navigation. Here we begin to explore the look and feel of interacting with flat 2-D representations using hand gestures to select viewing modes and to navigate by pan and zoom (Fig. 3). For these experiments we developed simple techniques for interpreting camera images and for passing events to the pre-existing application program.

The image processing prototyping was done in Matlab. A calibration step required the user to place his hand into a zone within the camera field of view where the mean and standard deviation of hue and saturation were computed and used for future identification of skin tone in the field. The left and right images were then processed to provide 2D position of the left and right hand in camera coordinates. The left hand image was tested for proximity to virtual button positions. The right hand image was tested for significant changes of position and interpreted as "mouse" input to be used for pan and zoom operations – hand orientation was interpreted as mouse up or mouse down state.

A python interface to the Matlab code was provided using PyMat. Detection of state changes from the button interface on the left hand and position changes from the mouse interface on the right hand were passed to the application program via its xmlrpc interface.

4 DISCUSSION

Although we have only begun to apply the prototype CupHolder to the various tasks for which it was designed, it is clear that the combination of visualization real estate, integrated (spill-proof) display table and control console, cameras, and its standing configuration provide a compelling interface.

The large display surface available for parallel rendering is very appealing (bright, crisp, dense in pixels) and useful for the range of applications tested on it. Pan and zoom on huge images filling the display was compelling. Applications which provided distinct components fit naturally into the segments of the tiled display.

The console and table sectors of the workstation provide a compelling environment for program development as well. The large available real estate in the table and console display fields is useful for coding, monitoring, and googling while testing applications on the visualization field. Of all the display systems we have designed, this one elicits that greatest "*I want one!*" reaction from people who see it.

With the limited number of "flying hours" that we have accumulated on this system we can make a few tentative observations about its ergonomics.

- Standing in front of the display and discussing visualizations in groups of two to five is quite effective and comfortable.
- Facing the display and working with keyboard and mouse on the table surface is comfortable as well, but depends on the height of the table which is fixed in the current design.
- Facing the table and controlling applications with hand gestures can be taxing on the back. This probably derives mostly from the slow frame rate of the current input processing which in turn requires the user to move slowly with arms cantilevered out over the control surface.
- Working at the console while seated in a "standing chair" is likely to provide added comfort for protracted sessions.

In this paper we have described a novel integration of commodity display, PC, and camera components into a functional unit that combines high performance visualization and interactive capabilities in a multi-purpose workstation. We described several use cases. This prototype device will allow us to experiment with and develop interactive input methods, distance collaboration, and remote visualization.

As a prototype, CupHolder will allow us to evaluate this breed of high performance workstation and continue to design and develop improved systems. Some of the areas that will need attention are summarized in this list of notes.

- The single most damning aspect of this approach to high resolution large format displays is the regrettable width of the border zone on each LCD panel. Some of this frame width could no doubt be eliminated at the packaging level. But there are likely more fundamental issues that would require creative re-engineering of the panels to allow for minimal abutting losses on all four of the sides of each panel.
- Another interesting possibility is to marry this display system with a pixel-based network transport, such as Sage [6]. Such an approach would reduce the expense of the local computing fabric, and would accrue benefit from the straightforward homogeneous model for content display – pixel shipping, that is.
- Because of its concave shape the CupHolder form factor provides a natural volume for capture and interpretation of gestural input, an area that needs further development in this context.
- The flat surface of the table field, backlit with a dense array of pixels provides an interesting arena for mixed mode user interface combining aspects of tangible UI wherein objects placed on the surface and provide means for interacting with displayed information on the table.
- We intend to integrate sound more tightly with the design. This enhancement will support distance collaboration applications.
- The present physical design would benefit from the addition of more unprogrammed space in the form of hooks, racks, and surfaces for personal items and customization.

5 ACKNOWLEDGMENTS

The authors acknowledge help from the other members of the Futures Laboratory, particularly Ti Leggett, in preparing for and executing the CupHolder deployment to SuperComputing 2005. This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract W-31-109-ENG-38 with additional support provided by the Pantheon Testbed project.

REFERENCES

[1] Mark Ashdown and Peter Robinson, “The Escritoire: A personal projected display”, Proceedings of the 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG 2003), Pilsen, Czech Republic, 3-7 February 2003, pages 33-40.

[2] Bérard, F., “The Magic Table: Computer-Vision Based Augmentation of a Whiteboard for Creative Meetings,” in IEEE workshop on Projector-Camera Systems (PROCAMS 2003), Nice, France.

[3] Mark Hereld, Ivan R. Judson, Rick L. Stevens, "Tutorial: Introduction to Building Projection-based Tiled Display Systems", *IEEE Computer Graphics and Applications* 20(4): 22-28 (2000).

[4] G. Humphreys, M. Houston, Y.-R. Ng, R. Frank, S. Ahern, P. Kirchner, and J. T. Klosowski, "Chromium: A Stream Processing Framework for Interactive Graphics on Clusters," in *Proceedings of ACM SIGGRAPH 2002*, San Antonio, pp 693-702.

[5] Ratti, C., Wang, Y., Ishii, H., Piper, B., Frenchman, D., “Tangible User Interfaces (TUIs): A Novel Paradigm for GIS”, in *Transactions in GIS*, 2004, 8(4): pp. 407-421.

[6] Renambot, L., Rao, A., Singh, R., Jeong, B., Krishnaprasad, Naveen, Vishwanath, V., Chandrasekhar, V., Schwarz,N., Spale, A., Zhang, C., Goldman, G., Leigh, J., Johnson, A., “SAGE: the Scalable Adaptive Graphics Environment”, in *Proceedings of WACE 2004*, Nice, France.

[7] Timothy A. Sandstrop, Chris Henze, and Creon Levit, “The hyperwall”, in *IEEE Proceedings of the International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV’03)*, ed. Jonathan C. Roberts, p.124.

[8] Larry L. Smarr, Andrew A. Chien, Tom DeFanti, Jason Leigh, Philip M. Papadopoulos, "The OptIPuter," *Communications of the ACM*, Volume 46, Issue 11, November 2003, pp. 58-67.

[9] R. Surati. “Scalable Self-Calibrating Display Technology for Seamless Large-Scale Displays”, PhD thesis, Massachusetts Institute of Technology, 1999.

A Light-weight Annotation System Using a Miniature Laser Projector

Ryugo Kijima *, Toshiyuki Goto **

* Faculty of Engineering, Gifu University

** Guraduate School Engineering, Gifu University

ABSTRACT

A small, hand-held laser projector is developed using a MEMS scanning mirror. The vertical scan is done by the vibrating mirror and the horizontal scan is done by the user's action to swing this projector. The prototype is equipped with the photo sensor and barcode reader circuit module to identify the real object with the retroreflective barcode. When the user swing this projector, projector begin to search the object with barcode ID, and when the projector detect such object, the annotation associated with the object's ID is displayed.

Keywords: Projection Display, Laser Projector, Mixed Reality, Objective Display

1 1. INTRODUCTION

The video projector is now widely recognized as a device for the augmentation of the real object. There have been many examples in fixed configuration [1][2], in hand-held configuration [3] and in wearable configuration [4]. Generally in mixed reality, the virtual object i.e. projected object is connected with the real object in terms of the location, of the shape, or of the semantics. To have system recognize what is the target object to be augumented, the communication from the projector to the object and/or that from the object to the projector is necessary. For this aim, there would be many potential combination among, for example, camera and color bar [5], barcode, shape recognition using the coded lighting, modulation on the retroreflective materials, RFID-tag, etc.

A remarkable feature in this area is that the function of image projection device is multiplexed. One typical functionality is to determine the relative location of the projector to the object. One example use the camera attached to the object that watches the projector's iris [6]. Others use the photo detector attached to the object and the time-based index of the projection such as the simple scanning, space coded rays [3], etc.

This article is not for the trial to fill another vacant cell in the combination matrix nor to add the new row or column to the matrix. The authors simply intend to introduce the new prototype of the small laser projector that recognize the traditional barcode on the object and project the suitable annotation to the object.

2 REACTIVE PROJECTOR

The authors call a type of system as the reactive projector. The projector gives some information to the real object using the light and the object send back another information to the projector. Based on this communication, the content of the projection is changed to fit to the object.

As mentioned above, there are so many means of the communication and the potential usage of their combination. One direction of the research would be a trial to find the new means, combination, and application. Then a point of evaluation would be if it is or will be technically or economically feasible or not.

Another direction is the trial to realize the feasible device for this area.

In this paper, the authors pick up a simple application that is to give the annotation on the object. When the user swing a small hand held projector, it detects the retroreflective barcode pattern on the object and project the annotation image on or close to the object. In this simple system, the key issue is how small the projector could be. In the next section, the details of the system will be described.

3 MEMS LASER PROJECTOR AND SYSTEM CONFIGURATION

In short, our prototype is the mixture of the laser barcode reader and the laser projection display. Figure 1 shows the primary component of the current prototype. A laser diode produces a ray whose brightness is modulated by the image generation circuit. The ray is reflected by a MEMS (Micro Electro Mechanical System) mirror that swings the ray in one degree of freedom controlled by a driver circuit. The coil is printed around the central mirror and this generates the driving force between the fixed permanent magnet, according to the drive current.

The user is asked to sweep this projector to observe the image. At first, the projector generates the laser ray at a constant brightness to find the target object. If the laser ray reach to the retroreflective material on the object, strong reflection goes back to the projector. Since the reflected ray is modulated by the barcode pattern printed on the retroreflective material, the photo sensor can detect the change of the brightness and recognize the ID. Then the projector displays the annotation image on or close to the object associated with the ID.

Please note the scan in one direction is by the vibrating mirror and that in another direction is done by the user's hand. The swing of the hand is the action to find the target object as well as to display the image. This enable the MEMS mirror to be simpler and smaller, though the mirror itself in the current prototype originally has a capability to be driven in two DOF.

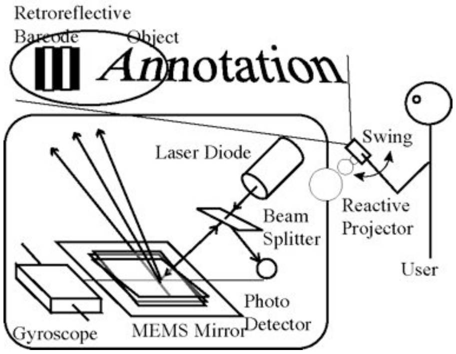


Figure 1. Reactive Projector System Overview

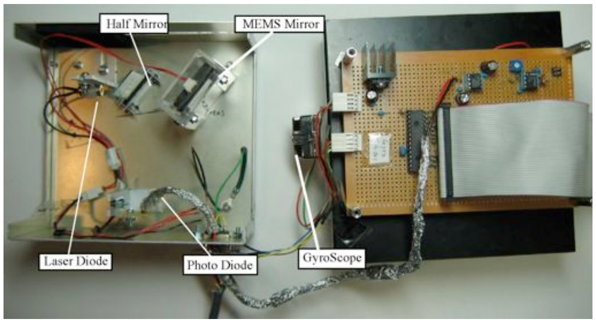
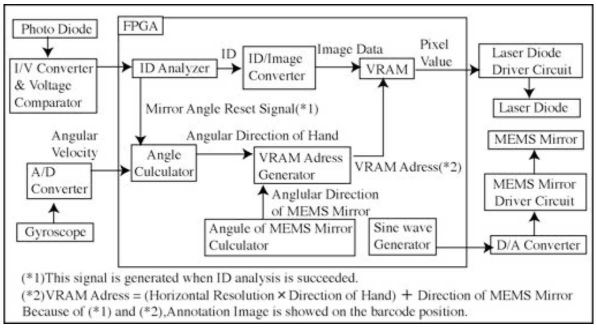


Figure 2. Appearance of the Prototype

The scanner driver, barcode reader, and the image generator are realized in a FPGA (Field Programmable Gate Array). The scanner driver generates the sine wave to drive the MEMS mirror. The image generator sends the appropriate pixel in the VRAM to the laser diode, referring to the phase of the scanner driver output to determine the row of the image, and to the gyroscope sensor to determine the column. The barcode reader module continuously reads the output of the photo sensor to find the start bit of the barcode pattern. After detecting this, it calculates the relative size of the barcode based on the time width of the start bit.



(*1)This signal is generated when ID analysis is succeeded.
(*2)VRAM Adress = (Horizontal Resolution × Direction of Hand) + Direction of MEMS Mirror
Because of (*1) and (*2),Annotation Image is showed on the barcode position.

Figure 3. Block Diagram of Controller

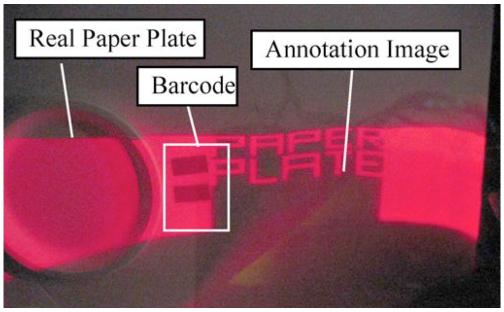


Figure 4. Real Object with Projected Annotation

4 DISCUSSION

4.1 Speed of Manual Scan and Human Vision System

Though our prototype works fine, there are several points in design remaining to be considered.

The first one is the adequate speed of the manual scan. When the user sweeps the ray slowly, the width of the image observed at once is narrow. To watch the larger area of the image, the user needs to swing the projector quickly. However the image resolution is reduced according to the speed of manual scan. Since the resolution depends on the frequency of the mechanical scan, we need the faster scanning mirror to realize the enough resolution.

The second point is concerning the fact that this display shows only the one-shot image to the user. One problem is that the user can read only several characters while they can see the whole image at once. To achieve the optimal design, we need to consider at least three issues, the size of the high resolution area in the human retina, tracking system by eyeball movement, and the relation between the contents and the human capability of the information processing. For example, in case of reading characters, the image is processed as a series of chunks and a chunk is processed at once. Even if the central area of the user's retina observed many words with enough resolution, the user's brain can process only a limited number of chunks. This issues concern not only to the physical aspect but also to the semantic side of the contents.

In our short experience to use the prototype, we noticed that the view of the projected image when the user was gazing at the fixed object is different from when the user was trying to follow the manual scanning. With very slow manual scan, the user's eye seems to pursuit the image or the scan line. At the medium speed, the eye seems to try to pursuit the line smoothly and sometime repeat the saccade around the scan line to follow. At the high speed, the eye cannot follow the scan and watch the fixed object. We need to explore which mode of the eye movement is suitable as well as how to have the user to naturally use in this mode

4.2 Future Work

The current prototype is just the assembly of the each parts and further larger than we finally expects. If the laser and photo diode, MEMS mirror will be made in one component, it would be possible to miniaturize to be in the cellular phone.

REFERERNCE

[1] Ishii, H., Underkoffler, J., Chak, D., Piper, B., Ben-Joseph, E., Yeung, L., Kanji, Z. Augmented Urban Planning Workbench: Overlaying Drawings, Physical Models and Digital Simulation, in Procs of Intl Symp on Mixed and Augmented Reality (ISMAR '02),

[2] Kondo, D., Goto, T., Kouno, M., Kijima, R., Takahashi, Y.,"A Virtual Anatomical Torso for Medical Education using Free From Image Projection", Procs of the 10th Intl Conf on VSMM 2004, pp678-685

[3] Raskar, R.; Beardsley, P.; van Baar, J.; Wang, Y.; Dietz, P.H.; Lee, J.; Leigh, D.; Willwacher, T., "RFIG Lamps: Interacting with a Self-Describing World via Photosensing Wireless Tags and Projectors", ACM Trans on Graphics (TOG) SIGGRAPH, 23-3, pp. 406-415, 2004 ACM.

[4] Kijima, R., Haza, K., Tada, Y., Ojika, T., "Distributed Display Approach using PHMD with Infrared Camera ", Intl Conf on Virtual Reality 2002 (VR2002), pp. 33-40, IEEE, 2002.

[5] J. Rekimoto, NaviCam: A Palmtop Device Approach to Augmented Reality, Fundamentals of Wearable Computers and Augmented Reality, pp.353-379, 2000.

Street Address and Electronic Mail Address
Unit 5, Virtual System Laboratory, Gifu University
1-1 Yanagido, Gifu, Gifu, 501-1193 JAPAN
1-2 {kijima, k03goto}@vsl.gifu-u.ac.jp

EDT 2006
IEEE VR 2006 Workshop
Sunday, March 26, 2006