TRANSVISION: A HAND-HELD AUGMENTED REALITY SYSTEM FOR COLLABORATIVE DESIGN

Jun Rekimoto

Sony Computer Science Laboratory Inc.
Takanawa Muse Building,
3-14-13 Higashi-Gotanda, Shinagawa-ku,
Tokyo 141, Japan
Phone: +81-3-5448-4380, Fax: +81-3-5448-4273
E-mail: rekimoto@csl.sony.co.jp

Abstract

This paper presents the shared augmented reality system called TransVision. TransVision augments real table-top with the computer graphics objects. Two or more participants hold a palmtop size see-through display and look at the world through the display. They can share the same virtual environment in the real world environment. Since users are not isolated from the real world, natural mutual communications such as body gestures can effectively be used during collaboration. This paper describes the architecture of the TransVision system and reports some early experiences.

1 Introduction

In traditional design studios, when a group of engineers is designing a new automobile, they would build a clay model of the automobile and examine it. During this process, discussions among engineers are often supported by gazing (e.g., looking at some part of the model and saying something) and hand gestures (e.g., one engineer pointing at a part of the model and saying "the design of this part has to be modified."). This rather old-fashioned approach is being replaced by computer aided design tools. However, the problem of current CAD-based design process is its lack of such intuitiveness.

The TransVision system described in this paper is an attempt to use augmented reality (AR) technology for collaborative designing. The system uses the palmtop video-see-through display instead of bulky head-mounted displays. The user can see a computer-generated 3D model superimposed on the real world view. The position and orientation of the display are tracked by the system such that the computer-generated model appears to occupy real space. Two or more participants can share the same computer model as if it were real. This situation seems similar to shared virtual environments, but there are fundamental differences between shared virtual reality and shared augmented reality approaches. By taking the augmented reality approach, two or more users can see each other without any difficulty. Since the coordinate system of each participant is identical, actions such as pointing gesture are meaningful to all participants. With virtual reality, we have to re-generate and render the human body as well as computer models.

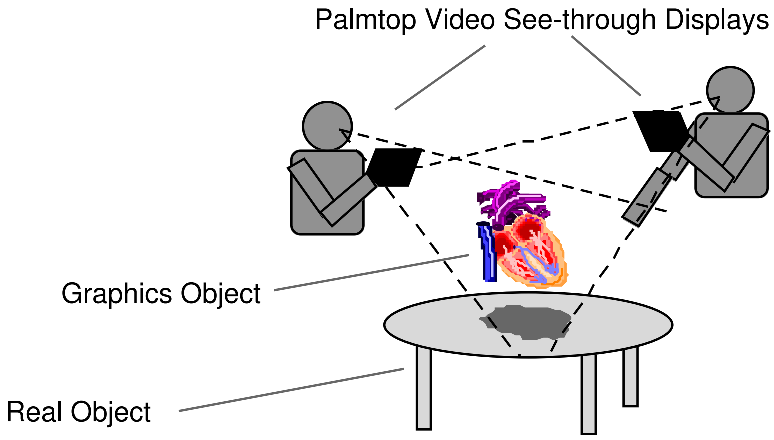


Figure 1: The concept of the shared augmented space

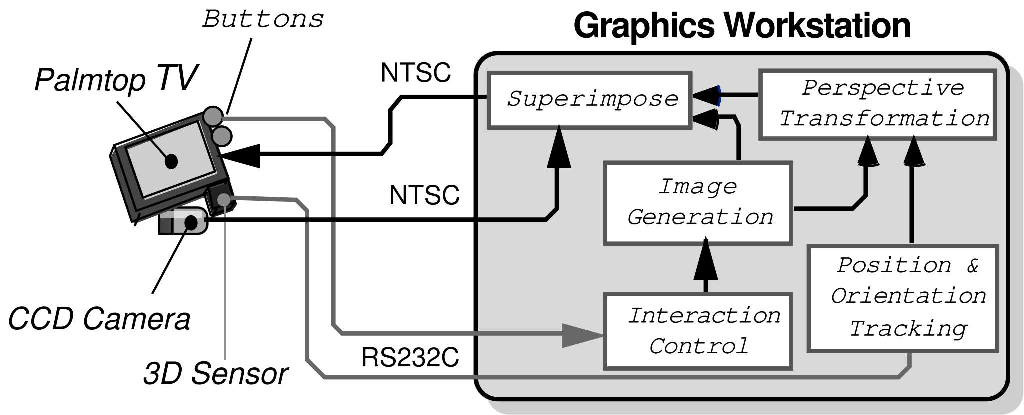


Figure 2: The system architecture

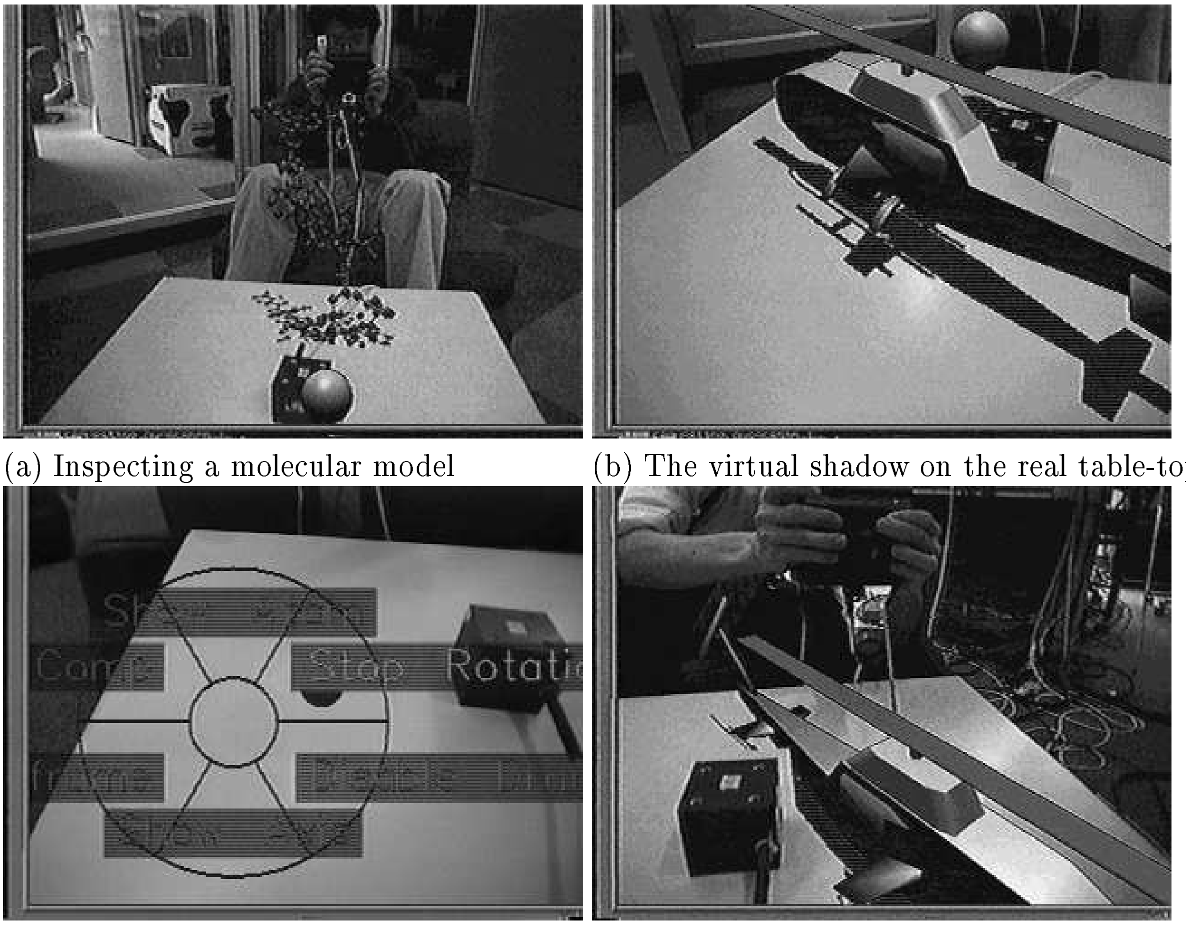
2 Previous Work

There are a number of researches on shared virtual reality systems [1, 8, 10, 11], but all of them were trying to make a virtual environment which is isolated from the real world.

The palmtop configuration is becoming popular in building VR and AR applications. The idea of using a palmtop display as a VR viewer was originally introduced by Fitzmaurice [2]. Rekimoto's NaviCam is the first AR system that uses the video see-through palmtop configuration [6]. Noma et al. integrated force-feedback with their palmtop VR system [4].

3 TransVision: The Shared Augmented Reality System

Figure 2 shows the system architecture of the TransVision system. The system comprises of two sub parts, the display part and the graphics subsystem (SGI Indigo2). These two parts are connected by two video cables and a serial line for position tracking data. The display



(a) Inspecting a molecular model (b) The virtual shadow on the real table-top

(c) The pie menu is in use (d) The virtual beam is emitted from the camera

Figure 3: Screenshots of the TransVision system

part consists of an LCD TV, a CCD camera, a 3D (Polhemus) sensor, and two buttons for operations. The software in the graphics subsystem is developed by using the MR toolkit [3].

3.1 Graphics Overlay

The graphics subsystem renders a scene according to the current position and orientation of the video camera. The position and orientation of the camera is tracked by the attached Polhemus sensor. Graphics overlay itself is achieved by using hardware video blending (i.e., chroma-keying).

To correctly overlay graphics images on the video image, the system needs to know camera intrinsic parameters as well as position and orientation. Tsai's camera calibration algorithm [9] was used to measure the focal length and the piercing point of the camera coordinate plane. These parameters are used as premeasured constants to define the perspective projection of graphics imaging.

The system has a simple model of the real world (where the table top lays, etc.). These real objects are rendered as invisible virtual objects, hence occlusion between virtual and (measured)

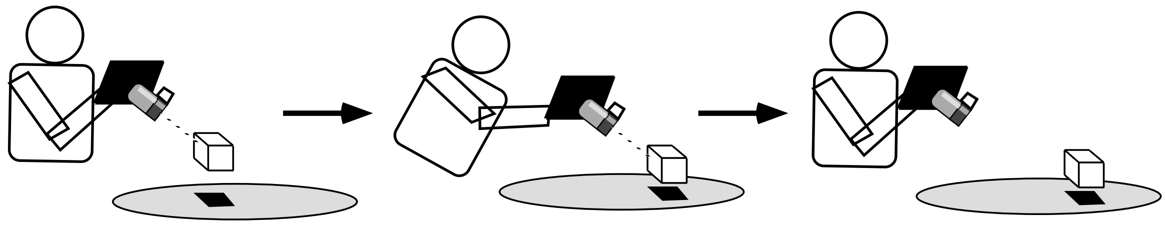


Figure 4: Virtual Object Manipulation

real objects is correctly visualized. Using the world information, the system also displays shadows of virtual objects on the real table top (Figure 3-(b)). These *virtual shadows* act as an effective visual aid to understand the relative position and the size of virtual objects.

3.2 Interaction Techniques

The display has two buttons for input. In current implementation, the upper button is used for object selection and manipulation, and the lower button is used for menu selection (Figure 3-(c)). The system uses a variation of pie menu with tilting operation that allows a user to select an menu item by tilting the display [5].

For selecting an object, a virtual beam is emitted from the camera along the orientation of the looking-at vector, and an object which is most close to this ray will be selected. From the user's point of view, an object at the center of the screen is selected.

Object manipulation also uses this virtual beam interaction model. Once the object is grabbed by pressing the button, the position of the object is fixed to the camera coordinate system. That is, the user can move the object by moving or rotating the display, as if the object were skewered by the virtual beam. We found that the user moves his/her entire body during this operation as they moved a real object on a table (Figure 4). Though further evaluation is needed to be conclusive, but we think that this similarity increases the naturalness of virtual object manipulation.

3.3 Data Sharing

In the TransVision system, two or more participates can share the same 3D models simultaneously, and any modifications to the model are also shared among participants. Like other shared VR systems, each user has its own 3D model database and the contents of the databases are kept coherent by propagating modifications from one user to other participants. The system uses Shaw's peers package [7] for the underlying inter-process communications method.

The current implementation does not allow simultaneous modification of the object database; only one participant has the ownership and can manipulate the object. The ownership is passed from one participant to another when the user pressing the button to select an object. Thus there is no explicit *get-ownership* command which might bother users.

To show who has the ownership, a virtual marker is overlaid on the camera of a participant who has the access right. In addition, a virtual beam is also overlaid on the other users' screen during manipulation (Figure 3-(d)).

4 Observations

This system was demonstrated at the Sony CSL open house and more than 100 people used the system. Here is the summary of observations that we found during this trial.

4.1 Graphics Overlay

Overlaying graphics on the real world scene was effective in understanding the location and the size of virtual objects. For example, it was easy to guess the size of a virtual object by comparing it with the real object such as the table. The user even could tell that "This (virtual) ball is about 10cm in diameter." We believe that without overlay, the user would not be able to judge that. The virtual shadow is also a quite effective method to infer the location of the object.

4.2 Mutual Awareness

One of the main goal of this project is to investigate how mutual awareness is achieved by using shared augmented reality. While using the system, what some one was doing was apparent from their body movements. For example, when one user wants to grab an object on the table, he/she first turns the body before making a selection. This body movement is visible to other participants and it implicitly transmits the user's intention.

Hand pointing was also effective. The user was able to say "please move this object" by pointing it with their finger. However, when the occlusion relationship between the hand and the virtual object was incorrect, users were confused. To avoid this problem, some users pointed their finger at the (virtual) shadow on the table.

Initially, we also expected that the user's eye direction (gaze) is also an effective communicative method for collaboration. However, since the user was mostly looking at the surface of the display during manipulation, there were less mutual gaze awareness. We discovered that the users gradually regarded the camera direction as gaze. The user became able to guess the other user's focusing area from the camera direction of that user. The virtual beam (Figure 3) also enforced this inclination.

5 Conclusion and Future Plans

We built a shared augmented reality system using a video see-through palmtop display. Early and informal observations suggested that overlaying virtual images on the real world scene helped the user to understand the location and the size of virtual objects. During collaboration, natural mutual awareness (e.g., body movement) was extensively used as well as awareness caused by artifacts (e.g., the virtual beam).

The current system only allows creation and manipulation of simple graphics objects. Future plans include development of more complete 3D interaction and application of this technology to the domains of 3D visualization and shared virtual games. We are also planning to combine vision-based position tracking to achieve more accurate alignment of virtual objects on the real objects.

Acknowledgment

The author is grateful to Yuji Ayatsuka, Trevor Stricker and Toshiyuki Masui for helpful discussions. The author would also like to thank Mario Tokoro and Toshi Doi for their continuing support.

References

[1] Lennart E. Fahlén, Charles Grant Brown, Olv Ståhl, and Christer Carlsson. A space based model for user interaction in shared synthetic environments. In *INTERCHI'93*, pp. 43–48, 1993.

[2] George W. Fitzmaurice, Shumin Zhai, and Mark H. Chignell. Virtual Reality for palmtop computers. *ACM Transactions on Information Systems*, Vol. 11, No. 3, pp. 197–218, July 1993.

[3] Mark Green and Dani Beaubien. *Minimal Reality Toolkit Version 1.2 Programmer's Manual.* Department of Computing Science, University of Alberta, Edmonton, Alberta, 1992.

[4] Haruo Noma, Tsutomu Miyasato, and Fumio Kishino. A palmtop display for dextrous manipulation with haptic sensation. In *CHI'96 proceedings*, pp. 126–133, 1996.

[5] Jun Rekimoto. Tilting operations for small screen interfaces. In *Proceedings of UIST'96, ACM Symposium on User Interface Software and Technology*, 1996. to appear.

[6] Jun Rekimoto and Katashi Nagao. The world through the computer: Computer augmented interaction with real world environments. In *Proceedings of UIST'95, ACM Symposium on User Interface Software and Technology*, pp. 29–36, November 1995.

[7] Chris Shaw. The MR Toolkit peers package and experiment. In *Proceedings of the Fifth Annual Western Computer Graphics Symposium*, 1993.

[8] Haruo Takemura and Fumio Kishino. Cooperative work environment using virtual workspace. In *ACM 1992 Conference on Computer-Supported Cooperative Work*, pp. 226–232, Tronto, Canada, November 1992.

[9] Roger Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 364–374, 1986.

[10] Michael J. Zyda, Rich Gossweiler, John Morrison, Sandeep Sighal, and Michael Macedonia. Networked virtual environments. In *A panel at Virtual Reality Annual International Symposium (VRAIS) '95*, pp. 230–231, 1995.

[11] M.J. Zyda, D.R. Pratt, J.G. Monahan, and K.P. Wilson. NPSNET: constructing a 3D virtual world. In *Computer Graphics*, 1992.