The Magnifying Glass Approach to Augmented Reality Systems

JUN REKIMOTO

Sony Computer Science Laboratory Inc.

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

http://www.csl.sony.co.jp/person/rekimoto.html

Abstract

This paper proposes a novel way to realize augmented reality (AR) systems. Unlike previous AR systems, which use head-mounted or head-up displays, our approach employs a palmtop-sized video see-through device to present computer augmented view of the real world. This configuration, which we call *the magnifying glass approach*, has several advantages over traditional head-up or head-mounted configurations. A user doesn't have to put on any cumbersome head gear. Like a magnifying glass, the user can easily move the device around and compare real and augmented images. We have built a prototype augmented reality system called NaviCam, based on the proposed magnifying glass approach. NaviCam uses a miniature gyro sensor to determine the orientation of device. It also has a vision-based ID recognition capability to detect a rough position of the device in the real world, and real world objects in front of the device. The experiment conducted by using NaviCam shows the great potential of the video see-through palmtop display. It was significantly faster than the head-up configurations, and its subjective score was also higher.

1 Introduction

Augmented Reality (AR) is a variant of Virtual Reality (VR) that uses see-through Head Mounted Displays (HMDs) to overlay computer generated images on the user's real sight [14, 9, 7, 2, 8, 6, 5, 1]. AR has a number of potential applications including maintenance, training, medical imaging, and entertainment. Normally, AR superimposes computer generated graphics relating to the real world objects (e.g., annotation). Since we are so familiar with the real world view and we can get better sense-of-3D from real world images than computer generated 3D images, see-through displays are also useful to VR.

There are many research issues in realizing see-through displays. One is, of course, how to achieve correct visual registration between real and virtual objects. Other than that, we would also like to raise human-factor issues. Unlike VR, where users can simply deal with computer synthetic worlds, users of AR systems are closely related to real world tasks. We can expect that the future AR systems will be used in various places in our daily lives such as offices, libraries, schools, homes, museums, or even on the road, as well as special environments such as surgery rooms. Thus it is crucial to select a comfortable device setup that is socially acceptable and allows users to perform real world tasks with minimum penalty.

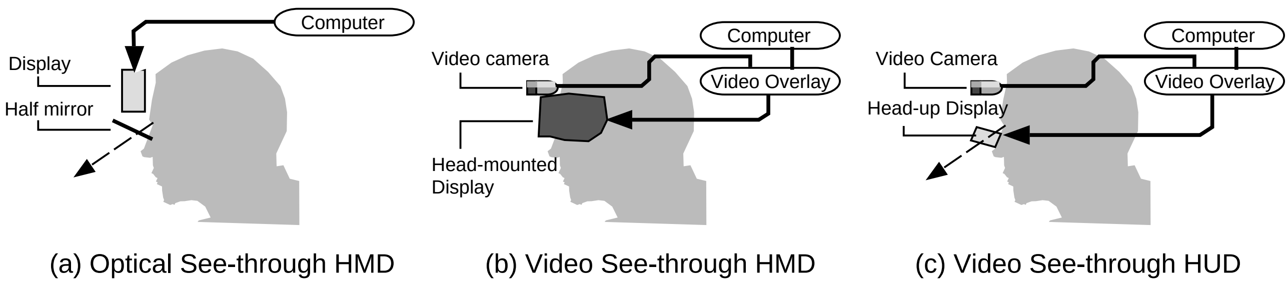


Figure 1: Previous See-through display configurations

This paper discusses several AR configurations from the viewpoint of human-factors, and proposes a new way to realize AR systems using a video see-through palmtop display. The rest of this paper is organized as follows. In the next section, we will briefly review previous AR configurations, and introduce our magnifying glass approach. In Section 3, we will present the prototype AR system called NaviCam, which is based on our proposed configuration. Section 4 explains the experiment to evaluate head-up and hand-held AR systems. Finally, Section 5 discusses several aspects of hand-held AR systems.

2 Display Configurations for Augmented Reality

This section discusses previous AR display configurations and their problems.

2.1 *Optical See-through HMDs*

Originally, AR merges virtual and real world optically, by using a half-silvered mirror. We call this *optical see-through HMD* (Figure 1-a). However, this configuration has several difficulties. Aside from the same problems with opaque HMDs (e.g., time consuming setup, weight, lack of high-resolution), optical see-through requires careful setup and it is very hard to calibrate the correct correlation between virtual and real positions. The system must know the accurate position of human eye-balls, but this position can be easily shifted during operation.

Another problem is the sense of depth. Even when virtual shapes are presented as stereo images, their focal length are fixed. Our vision system is confused by this contradiction especially when virtual images are compared with the real scene. For example, stereo disparity tells a user that a virtual image is behind a real object, while information from the focal length tells them that it is not true.

A third problem is occlusion. With the optical see-through configuration, it is hard to hide the real scene by virtual objects. Instead, every virtual object looks translucent. Occlusion gives the strongest effect on the sense of depth, but optical see-through HMD can not provide this effect correctly.

2.2 *Video See-through HMDs and HUDs*

To overcome these difficulties, another way to realize augmented reality has been investigated recently. This approach, which is called *video see-through*, overlays virtual images onto video images of the real scene, instead of the real scene itself. Bajura et al. first introduced this approach in their virtual environment ultrasound system [2].

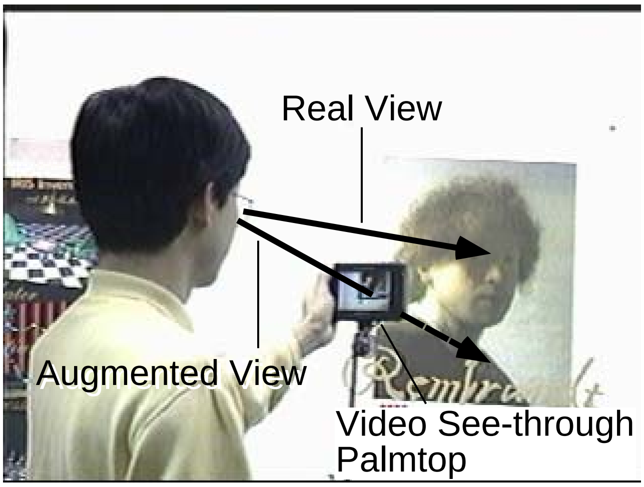


Figure 2: The magnifying glass metaphor

The video see-through approach has several advantages over the optical see-through approach. Once the correct visual registration of the virtual image is established, it is possible to provide correct occlusion between virtual and real objects. With a simple chroma-key video mixer, for example, it is easy to determine which image should be displayed, pixel-by-pixel. Object registration itself is easier than that of optical see-through. Bajura and Neumann suggested the method of obtaining coordinates information directly from incoming video images [3].

More precisely, there are two variation in video see-through AR. Bajura et al.'s system uses an opaque HMD (VPL EyePhone) to present virtual and real images, so the user's sight is completely shielded from the real world view by a HMD. We call this the *video see-through HMD* approach (Figure 1-b). Another configuration is to use a head-up display (HUD) that partially covers a viewer's dominant eye to display images, so the viewer watches a *virtual* floating window in front of his/her face (*video see-through HUD*) (Figure 1-c). PrivateEye and VirtualVision are typical examples of HUDs and can be used as video see-through HUDs. For example, Kuzuoka used the video see-through HUD configuration in his CSCW system called SharedView [10]. In his system, an operator puts on an HUD that displays images from a head-mounted camera merged with an instructor's hand.

However, both head-mounted and head-up video see-through approaches have problems. With the video see-through HMD, the user can only see the real world through the video. It means that the user's visual ability of recognizing real world is limited by the quality (resolution, color separation, and field of angle, etc.) of video images. It becomes a serious problem when an application of AR is a subtle real world task (e.g., surgery). The video see-through HUD has other problems: it introduces a dead angle area in front of the user's head, and it forces only one eye to see virtual images, which might put stress on viewers' eyes.

2.3 *Video See-through Palmtops – The Magnifying Glass Approach*

This paper proposes another configuration to realize augmented reality systems. Instead of wearing a head gear device, our approach uses a palmtop video see-through device. This approach is also called *the magnifying glass metaphor approach* (Figure 2). While a real magnifying glass optically enlarges the real world, a system based on this approach enlarges it in

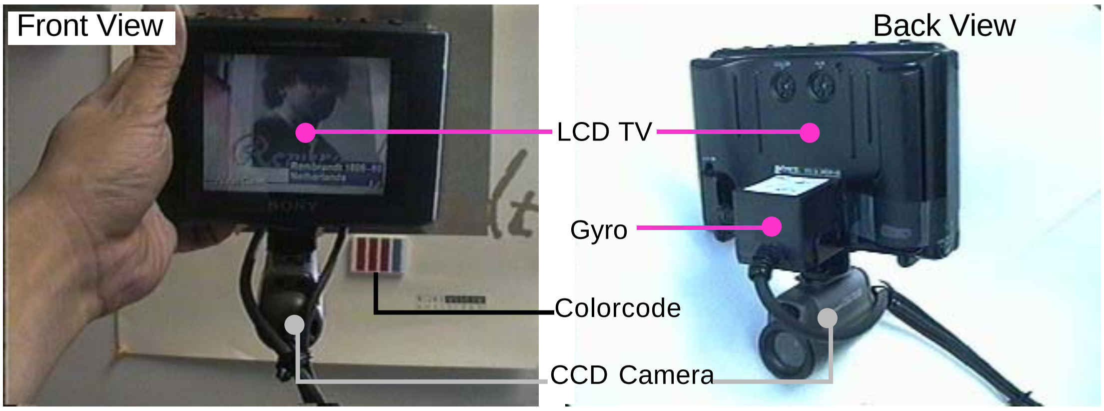


Figure 3: The outer view of NaviCam

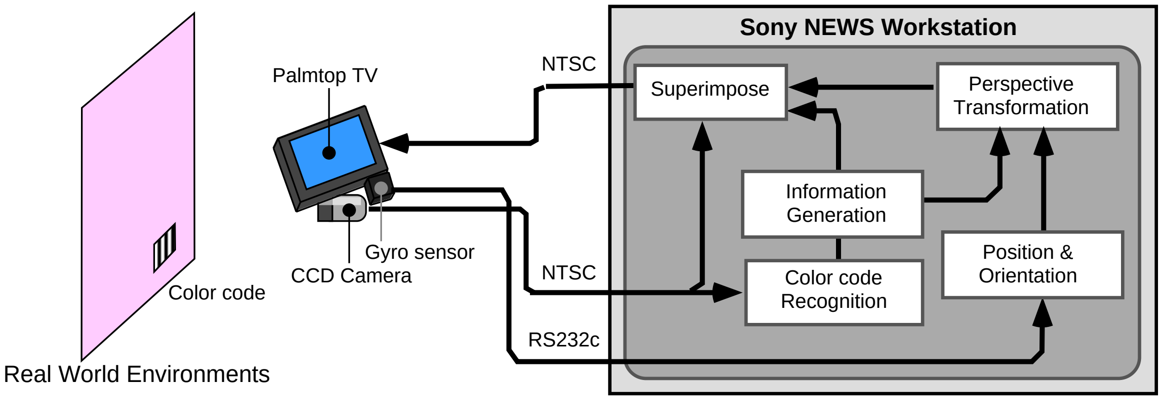


Figure 4: The system architecture of NaviCam

terms of *information*.

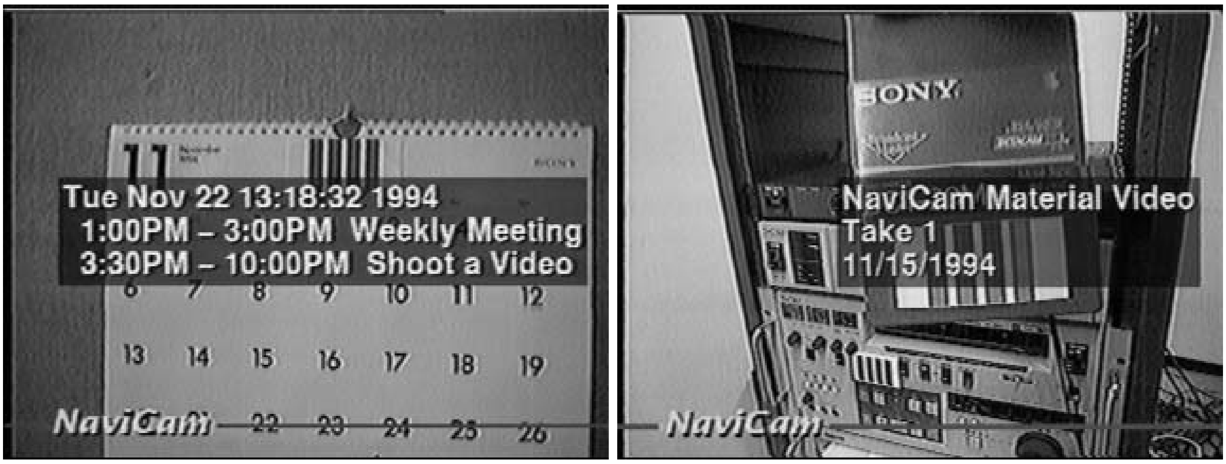
The magnifying glass approach has several advantages over previous AR configurations. First, it inherits advantages from video see-through systems. Secondly, it does not require any time to set up, because there is no head gear to wear. Unlike head-mounted or head-up displays, hand-held devices are perfectly socially acceptable and it is quite easy to quickly switch between the computer world and the real world. Finally, just as with a real magnifying glasses, it is easy to move the device around in the environment, to move it toward an object, and to compare the real image and the information-enhanced image.

In spite of many advantages, the major drawback of the hand-held configuration is that it does not allow hands-free operations. We will come back to this issue later in the discussions section.

3 NaviCam: The Palmtop Augmented Reality System

3.1 *System Architecture*

NaviCam is a prototype augmented reality system being developed at Sony Computer Science Laboratory, which is based on the magnifying glass metaphor. Figure 3 shows the outer view of the device. NaviCam consists of an LCD-TV screen, a miniature CCD-camera. and a gyro-sensor.



(a) Augmenting a paper calender (b) Annotation of a movable object

Figure 5: Snapshots from the NaviCam screen

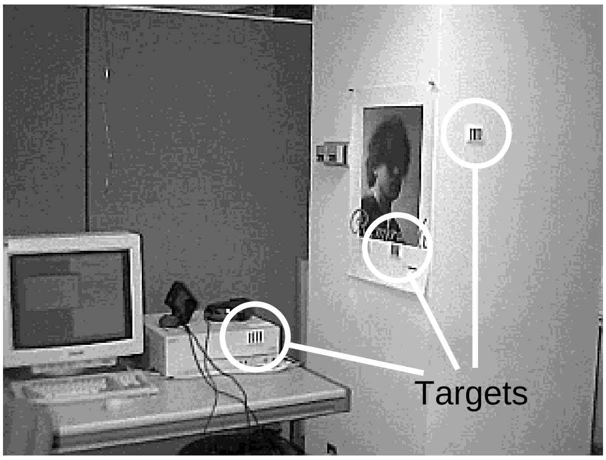
The information flow of the system is shown in Figure4. The real world view is captured by the CCD camera. Captured images are sent to the Unix workstation (Sony NWS-5000X with video capture and encoder boards), then merged with computer-generated information (either text or graphics), and sent back to the LCD screen. Thus, NaviCam presents the view at which the user is looking as if it is a transparent board, with computer augmented information. Video superimposition is performed by software and the system can process at a rate of 10 frames per second.

The gyro (JAE MAX3) used with this system is a solid state inertia-based position tracker consisting of three acceleration sensors and three orthogonal angular rate sensors. It is a 6DOF tracker so that it can report x–y–z positions as well as orientations (yow, pitch, and roll). Unlike other magnetic or acoustic position sensors, this device can detect position and orientation without other transmitter or receivers. Currently, NaviCam does not use position information reported from the gyro, because of its unacceptable inaccuracy. This is partially due to the limited precision of the gyro used. Another reason is that to obtain a position, it is necessary to transform a received acceleration vector in the body coordinates into its representation in the local coordinates by using angular information from the sensor. Since angular information may also have an error, it affects on the accuracy of positional information. This error can easily be accumulated during the double integration process of calculating a position from the acceleration.

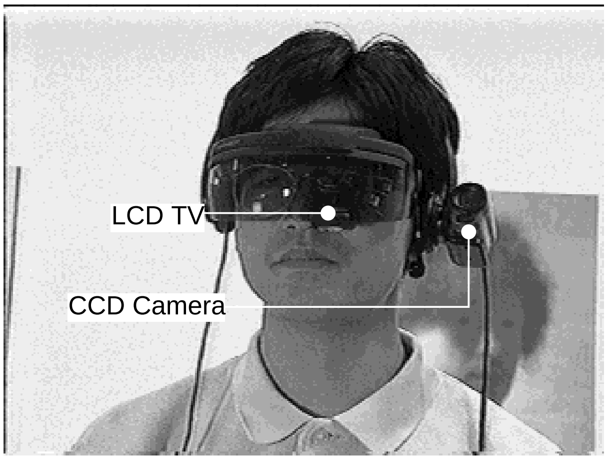
3.2 *ID Recognition*

Another unique feature of NaviCam is its ID recognition capability. By analyzing video images from the CCD camera, it can recognize a color-code, which is a kind of barcode, in front of the camera. This capability allows the system to detect a real world object in front of the system and to present information regarding to that object.

Figure 5 (a) is an example usage of ID-awareness. There is a *real* paper calender with a color-code ID tag. NaviCam detects its ID, and recognizes that the user is looking at a paper calender. Then the system retrieves schedule information from the database and superimposes it on the screen.



(a) The experimental Setting



(b) The head-up configuration

Figure 6: The Experiment

ID-awareness has some advantages over the traditional position-aware approach used by previous augmented reality systems. By using ID detection, the system can augment information about objects that might move. For example, the system can display information about a magnetic tape (Figure 5 (b)). Only using location information, this capability is very difficult to achieve. Feiner et al. tried to augment information on a movable object (e.g., a tray of a laser printer) by attaching another position sensor to that object [8]. However, this approach is unrealistic because it is almost impossible to attach position sensors to all objects in the real world.

4 Evaluation

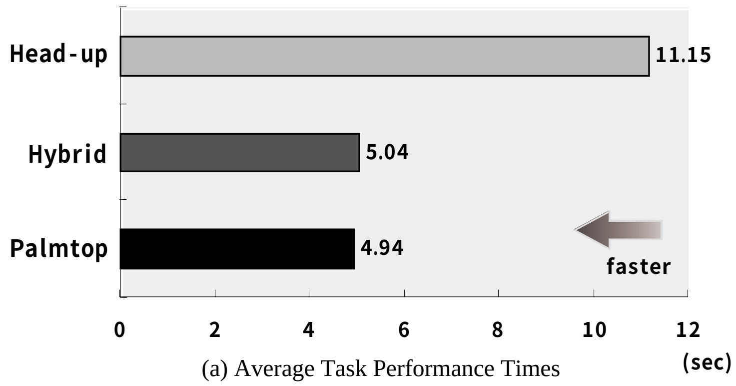
Using NaviCam as described in the previous section, we conducted an experiment to see the effect of head-mounted and palmtop configurations on the usability of augmented reality.

4.1 *Design*

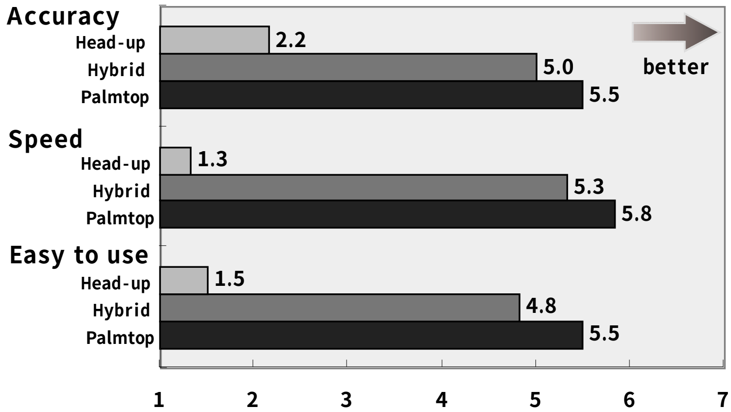
In this experiment, we placed three targets in an office environment and a subject *visits* these targets. The targets were placed on the wall, on a picture, and on a workstation (Figure 6 (a)). "*Visit*" means that subjects should locate a target about the center of the screen (either a head-up display or a palmtop display). To ensure this, we used a color-code tags as targets. When the system recognizes a color-code, it generates a beep sound and displays a message on the screen to inform the subject that he/she can proceed to the next target. Three targets were placed in different positions and different orientations. Thus the subjects must move the device both vertically and horizontally, and turn it to visit all three targets.

Subjects perform this task with the following three different conditions. The same software system was used in all conditions.

palmtop Subjects perform the task with a NaviCam palmtop system, as described in the previous section.



(a) Average Task Performance Times



(b) Subjective Rating Scores

Figure 7: Experimental results

head-up Subjects perform the task with a head-up display (VirtualVision Sport) and a head-up CCD camera (same as NaviCam 's camera), as shows in Figure 6 (b).

hybrid Subjects perform the task with a head-up display and a CCD camera in their hand.

6 subjects, all were computer engineers, took this experiment. Each subject performed 10 trials (visiting three targets) with each condition, 30 trials in total. The order of use of the three conditions was conter-balanced across the subjects. Prior to each task, some training trials were given. The task completion times were logged by the computer. After the experiment, each subject rated three conditions on a scale from 1 (worst) to 7 (best) in terms of accuracy, speed, and easy to use.

4.2 *Results*

The results of the experiment are shown in Figure 7. The ANOVA test was carried out on those results. Here are summaries of the analysis. Regarding task performing times, both the palmtop and the hybrid were significantly faster than the head-up condition ($F(2,177) = 54.31, p \ll 0.001$ and $F(2,177) = 52.46, p < 0.001$, respectively), while there was no significant difference between the palmtop and the hybrid. Regarding to subjective ratings, the subjects felt that the head-up was harder than the palmtop ($F(2,15) = 15.1, p < 0.001$) and the hybrid ($F(2,15) = 10.5, p < 0.01$). The subjects also felt the head-up was slower than the palmtop

$(F(2,15) = 17.9, p < 0.001)$ and the hybrid$(F(2,15) = 14.1, p < 0.001)$, and the head-up was less accurate than the palmtop $(F(2,15) = 7.73, p < 0.01)$ and the hybrid $(F(2,15) = 5.59, p < 0.05)$. There was no significant difference between the hybrid and the palmtop.

As confirmed by the experiment, the palmtop was significantly faster than the head-up, and all subjects preferred it to the head-up configuration. Other observations we found were as follows. The subjects with the head-up and the hybrid often confused the virtual and the real images. Some subjects even preferred to close one eye so that they could concentrate on the virtual view. There was no such confusion with the palmtop condition. With the hybrid condition, which was as fast as the palmtop, some subjects simply ignored virtual images on the head-up display and only relied on the sound feedback from the system. Thus, the result might be different if the task was designed to require information from the virtual view, without the help of sound feedback.

Paucsh et al. also reported that a HMD with a head-coupled tracker was able to improve performance in their experimental tasks, as compared with a combination of a stable display and a hand-held viewpoint control device [11]. Their result looks to be contradicting with ours, although the tasks performed were different.

To explain these different results, we would like to argue that the difference between virtual and augmented reality is very significant. In VR, a user recognizes the location and orientation *only* from VR devices. Thus, when the user's head movement is decoupled from virtual viewpoint movement, the user easily gets confused. In AR, on the other hand, a user can use their own spatial ability, because the virtual world is always related to the real world. As observed throughout the experiment, even though the palmtop device was not coupled with a subject's head movement, the subject was able to use their *real world view* to perform the task.

The other observation to be noted is the weight. The weight of the current prototype palmtop device is about $430g$ and some subjects complained of holding it in the air for a long time.

5 Discussions

Table 1 summarizes the features of AR configurations. As indicated in the table, it is clear that the major disadvantage of the palmtop configuration is its lack of hands-free capability. The palmtop approach is thus not suitable for some applications requiring two handed operation (e.g. surgery). On the other hand, putting on head-up gear is, of course, rather cumbersome and under some circumstances might be socially unacceptable. This situation will not change until head-up gear becomes as small and light as bifocal spectacles are today.

For the ID detection purpose used in NaviCam, the head-up configuration is also somewhat impractical because it forces the user to place their head very close to the object. Since hand mobility is much quicker and easier than head mobility, the palmtop configuration appears more suitable for browsing through a real world environment.

Another potential advantage of the palmtop configuration is that it still allows traditional interaction techniques through its screen. For example, you could to annotate the real world with letters or graphics directly on the palmtop screen with your finger or a pen. You could also operate it by touching a menu on the screen. This is quite plausible because most existing palmtop computers have a touch-sensitive, pen-aware LCD screen. On the other hand, a head-up configuration would require other interaction techniques with which users would be

Table 1: A summary of comparison among augmented reality configurations

see-through type display type	optical head-mounted [9, 8, 14, 1]	video head-mounted [2, 12]	video head-up [10, 13]	**video** **palmtop**
setup	very difficult	difficult	easy	**very easy**
real world view	fair	bad	fair	**good**
dead angle	no	no	yes	**no**
movement	slow	slow	slow	**fast**
social acceptance	no	no	may be	**yes**
on screen interaction	no	no	no	**yes**
hands free	yes	yes	yes	**no**

unfamiliar. For example, CMU's ViewMan system uses a special dial-like input device [4], but its ability looks quite limited as compared with a pen or a touch panel.

Returning to the magnifying glass analogy, we can identify uses for *real* head-up magnifying glasses for some special purposes (e.g. watch repair). The head-up configuration therefore has advantages in some areas, however, hand-held magnifying lenses are still dominant and most prefer them in their daily lives. Thus we can conclude that there should be many applications where palmtop AR systems are more usable and practical than head-up AR systems.

6 Summary

In this paper, we have proposed a video see-through palmtop approach (also called the magnifying glass approach) to realize augmented reality systems. This configuration is radically different from traditional VR and AR systems, but has great potential to enhance human's ability in interacting with the real world environment. We are currently working on improvement of a gyro device driver to combine ID-awareness and location-awareness capabilities.

ACKNOWLEDGMENTS

We would like to thank Katashi Nagao, Satoshi Matsuoka, and Jonathan Grudin for their helpful discussions. Special thanks also go to the subjects who volunteered for the experiment, Rodger Lea for his valuable comments on the draft, and Hiroshi Yoshimatsu for information on gyro sensors. Finally, we would like to thank Mario Tokoro and Toshi Doi for their continuing support.

References

[1] Daniel G. Aliaga. Virtual and real object collisions in a merged environment. In *Proceedings of the VRST'94 Conference*, pp. 287–298, 1994.

[2] Michael Bajura, Henry Fuchs, and Ryutarou Ohbuchi. Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. *Computer Graphics*, Vol. 26, No. 2, pp. 203–210, 1992.

[3] Michael Bajura and Ulrich Neumann. Dynamic registration correction in augmented-reality systems. In *Virtual Reality Annual International Symposium (VRAIS) '95*, pp. 189–196, 1995.

[4] Len Bass, Dan Siewiorek, Asim Smailagic, and John Stivoric. On site wearable computer system. In *CHI'95 Conference Companion*, pp. 83–84, 1995.

[5] David Drascic, Julius J. Grodski, Paul Milgram, Ken Ruffo, Peter Wong, and Shumin Zhai. ARGOS: A display system for augmenteing reality. In *INTERCHI'93*, pp. 521, 1993.

[6] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. Windows on the world: 2D windows for 3D augmented reality. In *Proceedings of UIST'93, ACM Symposium on User Interface Software and Technology*, pp. 145–155, November 1993.

[7] Steven Feiner, Blair MacIntyre, and Doree Seligmann. Annotating the real world with knowledge-based graphics on a see-through head-mounted display. In *Proceedings of Graphics Interface '92*, pp. 78–85, May 1992.

[8] Steven Feiner, Blair MacIntyre, and Doree Seligmann. Knowledge-based augmented reality. *Communication of the ACM*, Vol. 36, No. 7, pp. 52–62, August 1993.

[9] Steven Feiner and A. Shamash. Hybrid user interfaces: Breeding virtually bigger interfaces for physicall smaller computers. In *Proceedings of UIST'91, ACM Symposium on User Interface Software and Technology*, pp. 9–17, November 1991.

[10] Hideaki Kuzuoka. Spatial workspace collaboration: A SharedView video support system for remote collaboration capability. In *Proceedings of CHI'92*, pp. 533–540, 1992.

[11] Randy Pausch, M. Anne Shackelford, and Dennis Proffitt. A user study comparing head-mounted and stationary displays. In *IEEE Symposium on Research Frontiers in Vritual Reality*, pp. 41–45, 1993.

[12] Jannick P. Rolland, Frank A. Biocca, Todd Barlow, and Anantha Kancherla. Quantification of adaptation to virtual-eye location in see-thru head-mounted displays. In *Virtual Reality Annual International Symposium (VRAIS) '95*, pp. 56–66, 1995.

[13] Jane Siegel, Robert E. Kraut, Bonnie E. John, and Kathleen M. Carley. An empirical study of collaborative wearable computer systems. In *CHI'95 proceedings*, pp. 312–313, 1995.

[14] Ivan Sutherland. A head-mounted three dimensional display. In *Proceedings of FJCC 1968*, pp. 757–764, 1968.