An optical see-through display for mutual occlusion with a real-time stereovision system

Kiyoshi Kiyokawa\textsuperscript{a,*}, Yoshinori Kurata\textsuperscript{b}, Hiroyuki Ohno\textsuperscript{a}

\textsuperscript{a}Emergency Communications Group, Communications Research Laboratory, 4-2-1 Nukui-Kitamachi, Koganei, Tokyo 184-8795, Japan
\textsuperscript{b}Topcon Co.75-1 Hasumuma-cho, Itabashi-ku, Tokyo 174-8580, Japan

Abstract

Conventional optical see-through displays are not capable of correctly presenting the mutual occlusion of real and virtual environments, since the synthetic objects always appear as translucent ghosts floating in front of the real scene. We have been developing an optical see-through display that is capable of such mutual occlusion. Our display, named ELMO, has both the features that are necessary for mutual occlusion: a selective light-blocking mechanism and a real-time depth sensing mechanism. Firstly, our novel optics cut off incoming light selectively without themselves going out of focus, so virtual objects that should be in front of the real scene are made opaque and sharply occlude the background. This feature also ensures that the virtual objects remain in their intended colors. Secondly, the sensing mechanism, a stereovision system built into ELMO, is used to acquire a depth map of a real scene in real-time, the system can be used in unknown and dynamic environments. As a consequence, ELMO does not require particular environmental settings such as light conditions, and can be used anywhere, e.g., outdoors. We have built prototype displays and confirmed the effectiveness of the approach. In this paper, the basic concept, design of the optics, our prototype displays, and some empirical studies are presented and discussed. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Augmented reality; Mixed reality; Optical see-through display; Mutual occlusion; Real-time depth sensing

1. Introduction

The goal of mixed-reality technology is to produce environments by seamlessly integrating both real and virtual worlds [1]. Realistic occlusion phenomenon is an important part of making mixed-reality environments convincing, as occlusion is well-known to be a strong depth cue. The mutual occlusion of real and virtual objects enhances the user’s impression that virtual objects truly exist in the real world. This is an essential feature for certain mixed-reality applications, such as architectural previewing. Besides, in terms of cognitive psychology, incorrect occlusion confuses users [2]. Fig. 1 shows some examples of mutual occlusion. Fig. 1(A) is a real scene on which a synthetic image is to be superimposed. Fig. 1(B) is the synthetic image with correct occlusion attributes, and Fig. 1(C) is the ideal result of the image overlay. Without occlusion attributes, the merged image looks weird, even though the synthetic image is perfectly registered on the real image (see Fig. 1(D)).

Two items are needed to correctly present mutual occlusion of real and virtual objects, (1) a display that is capable of showing images with mutual occlusion and (2) depth information on the real scene. See-through displays used for mixed reality are of two types: video see-through and optical see-through [3,4]. Conventional optical see-through displays are not capable of correctly presenting mutual occlusion, because the synthetic objects always appear as semitransparent ghosts floating in front of the real scene, as is illustrated in Fig. 1(D). Video see-through displays have thus, up to now, been exclusively used to realize mutual occlusion [5,6].
However, such displays reduce the spatial and temporal resolution of the real scene, have a limited depth-of-field, fixed-focus, and so forth. We have designed an optical see-through display named ELMO that is able to handle the mutual occlusion of real and virtual environments, and have built prototype displays to confirm our approach [7].

It has recently become possible to acquire the depth information of a real scene in real-time by using stereo-paired cameras or laser range finders [5,8,9]. There is a multi-camera real-time stereovision system inside the body of our latest prototype display, ELMO-3. By comparing the acquired depth map with that of the virtual scene (namely, Z-buffer), correct occlusion can be reproduced within unknown or dynamic environments in real-time.

This article is a review of our work thus far on the ELMO optical see-through display. In this article, we describe the design of our display and look at our prototype displays along with a few experimental results. This paper is organized in the following way: in Section 2, the characteristics of the two types of display, video see-through and optical see-through, are explained. Section 3 gives a brief survey of prior work. In Section 4, the basic concepts and characteristics of our optics design are described. Sections 5–7 describe the three generations of prototype displays and some experiments we have carried out with them. Finally, Section 8 gives conclusions and some of the directions our future work will take.

2. Conventional displays

2.1. Video see-through displays

A video see-through display electronically combines a synthetic image with a real image that is captured by a video camera mounted on the user’s head. The combined image is then presented in front of the user’s eyes. Video see-through displays have the following three advantages.

1 Pixel-based image processing of the real scene, e.g., correction of intensity and tint, blending ratio control, is available.
2 Temporal registration error in each rendered frame can be eliminated by synchronously processing and presenting the real and virtual images.
3 Implicit and explicit visual information from the real scene is available for utilization. For example, the depth information of a real scene can be calculated from multiple images [5], and the relative orientation in and translation of the camera (user’s head) and objects in the environment can be acquired by using computer-vision techniques [6].

The first advantage allows video see-through displays to handle the occlusion problem without difficulty. If the system has the depth information of the real scene, rendering a combined image with correct occlusion is a simple choice, from the two image planes, synthetic and captured, of the source for each pixel. Hence, this has so far been the type of display used in realizing mutual occlusion. However, a video see-through display reduces the rich information content of the real world because of the display’s low spatial and temporal resolution, limited depth-of-field, fixed-focus, and so forth. Another disadvantage is that a user may temporarily lose his/her vision when there is system trouble.

2.2. Optical see-through displays

A typical optical see-through display allows its user to see the real world and a virtual environment simultaneously by applying a mirror that is partially transmissive and partially reflective. Such displays have relatively simple structures and are widely used. Unlike video displays, they preserve the real image as it is. However, conventional optical see-through displays have a significant disadvantage: the synthetic objects always appear as translucent ghosts floating in front of the real scene, because, with a normal beam splitter, there is no way of selectively blocking out the real scene. Each pixel of the synthetic image is affected by the color of the real image at the corresponding point, and never directly shows its intended color. The nature of the approach avoids the problem of the mutual occlusion of the real and virtual environments.
3. Previous work

3.1. Ray paths in an optical see-through display

Researchers have recently been attempting to build new optical see-through displays that tackle the problem of occlusion. Before describing some of this work, let us consider the ray paths of an optical see-through display so that we can better understand and classify them.

To realize mutual occlusion with an optical see-through display, we must selectively cut off or pass the rays of the real scene by using some kind of mask pattern. Fig. 2 shows ray paths of an optical see-through display. (1) Firstly, rays are emitted from the light source, (2) they are then reflected from real objects and (3) pass through the air. (4) Some of them are then blended with synthetic images by an optical combiner, and finally arrive at the viewer’s pupils. Corresponding to these four portions in a path, the following four approaches are conceivable ways of covering a real scene with a synthetic image.

1. Cut rays off between the light source and objects [8].
2. Cut rays off by preparing real counterparts of the virtual objects [10,11].
3. Cut rays off between the objects and the user’s eyes.
4. Decrease the visibility of the real scene by increasing the relative intensity of the synthetic image.

3.2. Approach 1: pattern light source

Noda et al. developed a unique approach using a pattern light source [8]. Fig. 3 shows the system’s configuration. Firstly, a laser rangefinder acquires a depth map of the real objects in the darkroom. Then, those parts of the real objects that should appear are lit by a front projector. Finally, the pixels of the virtual objects that should be in front of the occluded real objects are rendered. Consequently, a correct mutual occlusion phenomenon is observed from the range finder’s viewpoint. With this approach, it is possible to form mask patterns of any shape, in real-time, to correspond to the geometry of the real objects. However, the lighting conditions must be strictly controlled (a darkroom is essential), so its area of potential application is extremely limited.

3.3. Approach 2: real counterpart object

Kameyama developed a simple CAD system that allows users to perceive mutual occlusion, by appropriately controlling the intensity of a real environment [10]. In this system, the user manipulates a black input device on which a synthetic image is superimposed through a half-silvered mirror. The user’s hands are so brightly lit that his or her fingers in front of the device are clearly visible through the superimposed image. As a result, the user’s fingers are perceived as covering the synthetic image. Inami et al. developed a head mounted projector [11] that uses real objects covered with retro-reflective material as screens onto which a synthetic scene is projected. Fig. 4 shows the configuration of the system. Stereo-paired synthetic images are cast on the real scene from projectors placed in positions optically equivalent to those of the user’s eyes. Most of the rays that hit the screen are then reflected and go back to each user’s eye, while most of the rays that hit other real objects are randomly diffused so that negligibly few go back to the user’s eyes. Therefore, the user sees only virtual objects on the screen. Consequently, any other real objects in front of the screen occlude the virtual objects and the screen itself occludes any real objects behind it. Mutual occlusion is thus realized. Although these approaches are useful and relatively simple, it is
inconvenient to prepare a real counterpart for each virtual object, and lighting conditions must be carefully controlled.

3.4. Approaches 3 and 4

Approach 4 does not require any special environmental factors so it can be used anywhere, including outdoors. As a matter of fact, this approach has most commonly been used for the head-up displays of aircrafts. However, this approach strictly restricts the range of available colors to bright colors alone. This approach is thus exclusively used to display supplemental information rather than realistic virtual objects.

On the other hand, approach 3, when the light blocking mechanism is embedded in the HMD, is promising in terms of providing a more generally applicable approach. In this case, the display can be used anywhere and is capable of presenting any color, including black. Such an approach, however, is not easy to realize optically, due to the out-of-focus problem. That is, even if an addressable optical filter is placed on the half-mirror, the formed mask pattern will itself not be in focus when the user is focusing on outside scenery. As a result, the edges between synthesized and real images look unnatural. In the next section, we will describe how we have actually tackled this problem.

4. Optics design

We employed approach 3 as discussed in the previous section. Fig. 5 shows the basic idea of this optics. The heart of the design is very simple. That is, a liquid crystal display (LCD) panel is placed between a conventional optical see-through display and the outside scenery in order to selectively block the entry of arbitrary rays from the outside. The problem is that the LCD panel is so close to the user’s eyes that the pattern on the LCD panel will be out of focus when the user is viewing external objects.

To solve the out-of-focus problem, we take two convex lenses, each with the same focal length \( f \), and place one in front and one behind the LCD panel to structure a kind of telescope with a magnification of one. We then add an erecting prism to erect the inverted outside scenery. Fig. 6 shows the basic design of the optics. With this optical system, a viewer can simultaneously focus on both the outside scenery and a pattern on the LCD panel. By opening and shutting the pixels on the LCD panel at positions where real objects should appear and disappear, respectively, it is possible to optically present a true mutual occlusion of real and virtual worlds. Sony proposed the same optical idea in 1992 [12], prior to our independent invention of it, although there is no report of their having carried out a feasibility study. Our goal is to develop actual displays for practical use, and this requires much work and a number of artifices.

We named this display ELMO, which is an abbreviation for an enhanced optical see-through display using an LCD for mutual occlusion. One of the major advantages of ELMO is that it neither affects the real environment nor requires any additional environmental settings. It can thus be used anywhere in any situation, including outdoor applications. ELMO is able to surely block any light coming from outside scenery in any situation, so the virtual images also retain their intended colors, and ELMO’s color fidelity is thus greatly superior to conventional optical see-through displays. On the other hand, if we make this optical system in a straightforward way, it adds a certain amount of offset to the user’s viewpoint, causing inconsistencies between proprioception and visual perception. In Section 6, we will show you how we eased this problem with ELMO-2.

In the following three sections, the three prototype displays are explained in the order in which we built them. The bench-top prototype, ELMO-1, was built first to confirm the validity of the basic concept. ELMO-2 was built next, and featured a stereoscopic capability and compact optics. ELMO-3 is the latest prototype and features the inclusion of a real-time rangefinder and color displays inside its body. ELMO-3 is the first optical see-through HMD in the world to be able to handle mutual occlusion appropriately without imposing special and extraneous hardware or environmental settings.
5. ELMO-1: the first prototype display

5.1. Overview

We implemented the basic design of the optics in building the first prototype display, ELMO-1, to confirm the feasibility of the approach. Fig. 7 shows the appearance of ELMO-1. We developed new lenses and a prism, and employed a commercially available 10.4-inch LCD panel (Integral Electronic Japan, IDB-9344W). The specifications of the optics of ELMO-1 are summarized in Table 1. The dot pitch of the panel is 0.27mm and the effective area of the panel is 31mm × 31mm. This setup is thus able to display a mask pattern of about 115mm × 115 pixels. The degrees of transparency of open/closed pixels of the panel are varied by changing the resistance of the panel’s circuit (see Fig. 8), and we adjust them to 18% and 2%, respectively, to maximize the contrast.

5.2. Experiment

Fig. 9 shows ELMO-1 with an optical see-through HMD (Shimadzu STV-E). A PC (SGI VWS540) was used to render color images for sending to the HMD via the PC’s video output. A digital video camera (Sony DCR-TRV900) is used to capture the image seen through the display. Fig. 10 illustrates the configuration of the experimental setup. All virtual objects and the real white box, which is only one real object here for occlusion test, were modeled in advance (Fig. 10(A)) of the experiment. The 3D magnetic tracker (Polhemus Fastrak) transmits the position and orientation of the real box (Fig. 10(B)) to the PC (Fig. 10(C)). The PC then renders a virtual scene that is composed of virtual objects including two counterparts of the real box, a mask pattern (Fig. 10(D)) and a color image (Fig. 10(E)). Note that the counterpart of the real object is rendered transparent by using OpenGL’s alpha-blending feature [4]. By doing this, the Z-buffer is appropriately modified according to the real object’s geometry while the corresponding color pixels are kept transparent (i.e. black). Combining these rendering techniques, a mask pattern can easily be rendered by disabling virtual light sources. As a result, correct mutual occlusion is presented by this setup as illustrated in Fig. 10(G), while conventional displays are only able to present a ghost image, as shown in Fig. 10(F).

Fig. 11 shows four images captured by this setup. Conventional optical see-through displays present an

---

Table 1
Specifications of ELMO-1

<table>
<thead>
<tr>
<th>Eyepiece/Objective lenses</th>
<th>Focal length</th>
<th>70 (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effective aperture</td>
<td>37 (mm)</td>
<td></td>
</tr>
<tr>
<td>Center thickness</td>
<td>15.5 (mm)</td>
<td></td>
</tr>
<tr>
<td>Weight (for each)</td>
<td>56 (g)</td>
<td></td>
</tr>
<tr>
<td>Erecting prism</td>
<td>Window shape</td>
<td>35 × 35 (mm)</td>
</tr>
<tr>
<td>Refractive index</td>
<td>1.755</td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td>173 (g)</td>
<td></td>
</tr>
<tr>
<td>Effective field of view</td>
<td>&gt; 25 (degree)</td>
<td></td>
</tr>
<tr>
<td>Exit pupil aperture</td>
<td>&gt; 5 (mm)</td>
<td></td>
</tr>
<tr>
<td>Eye relief</td>
<td>&gt; 60 (mm)</td>
<td></td>
</tr>
<tr>
<td>Viewpoint offsets</td>
<td>Horizontal</td>
<td>291.9 (mm)</td>
</tr>
<tr>
<td></td>
<td>Vertical</td>
<td>35.7 (mm)</td>
</tr>
<tr>
<td>Transparency/reflectance</td>
<td>(lambda = 550 (nm))</td>
<td></td>
</tr>
<tr>
<td>Eyepiece &amp; objective lens</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>Prism</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.88</td>
<td></td>
</tr>
</tbody>
</table>
overlaid image like the one shown in Fig. 11(A). In this case, the occlusion relationship is not presented so that you are unable to tell which portion of the virtual image is in front of the real objects. With depth information on the real object, the virtual objects can be covered by the real one (Fig. 11(B)). However, they remain semitransparent ghosts and the real object behind the virtual image remains visible. On the other hand, ELMO-1 is able to block out the real scene selectively as is shown in Fig. 11(C). Finally, by combining the mask pattern with colorful virtual images, which are presented by a normal optical see-through display, it is possible to present a mixed world with correct occlusion and high color fidelity as shown in Fig. 11(D).

6. ELMO-2: the second prototype display

6.1. Overview

Although the feasibility of our approach was confirmed by ELMO-1, some of its properties were not fully satisfactory. The optics needed to be made stereoscopic and compact, the length of the viewpoint offset needed to be shortened, and the field-of-view needed to be enlarged. We had to improve the transparency, contrast, response time, and resolution of the LCD panel.

To achieve some of these goals, we designed the new compact folded system of optics shown in Fig. 12. We independently contrived the system after considering several geometrically similar layouts of prisms and mirrors. The system has four rectangular prisms and two mirrors for each eye, thereby reducing its size and the viewpoint offset. From the viewer’s eyeball to the opposite side of the display, we arranged an eyepiece, prisms 1 and 2, the LCD panel, prisms 3 and 4, an objective lens, and mirrors 1 and 2. A big advantage of this system is that it eliminates horizontal and vertical viewpoint offsets. Although similar layouts were presented in Sony’s patent, the fields of view they provided are impractically narrow because the mirrors used in folding the optical paths obstruct the view. Another unique advantage of our system of optics is its...
separation into front and rear parts so that two LCD panels of any size or a single shared panel can be used. High-refractive-index prisms are used to gain enough optical path length to ensure a wide field of view. Taking weight, refractive index, cost and so on, into account, we chose an LaSF-40-equivalent material, which has a specific gravity of 4.43, for prisms 1 and 2, and a BK-7-equivalent material, which has a specific gravity of 2.52, for prisms 3 and 4. Mirrors are used for those components that are farther than the objective lenses so that the system is kept light. This layout brings the center of gravity closer to the head and decreases the moment of rotation.

Based on these ideas, we built our second prototype display, ELMO-2. Three views of ELMO-2 are shown in Fig. 13(A)–(C). The specifications of ELMO-2’s optics

Fig. 11. Captured images of the view presented by ELMO-1.

Fig. 12. Compact folded optics.
are summarized in Table 2. The outer frame is made to house a 10.4-inch LCD panel. The optical parts weigh about 1 kg and ELMO-2’s total weight with the outer frame is about 2 kg. The eyepieces and LCD panel are movable along the optical axis, and the inter-pupil distance is also adjustable. We also made a supporting and linking unit for ELMO-2, since it is still heavy for a head-mounted device. The appearance of ELMO-2 with the unit is shown in Fig. 14.

### Table 2: Specifications of ELMO-2

<table>
<thead>
<tr>
<th>Eyepiece/Objective lenses</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Focal length</td>
<td>75 (mm)</td>
</tr>
<tr>
<td>Effective aperture</td>
<td>38 (mm)</td>
</tr>
<tr>
<td>Center thickness</td>
<td>15.4 (mm)</td>
</tr>
<tr>
<td>Weight (for each)</td>
<td>56 (g)</td>
</tr>
<tr>
<td>Erecting prism</td>
<td></td>
</tr>
<tr>
<td>Window shape</td>
<td>40 × 40 (mm)</td>
</tr>
<tr>
<td>Refractive index</td>
<td>1.516 ~ 1.834</td>
</tr>
<tr>
<td>Weight</td>
<td>161 ~ 284 (g)</td>
</tr>
<tr>
<td>Effective field of view</td>
<td>&gt; 25 (degree)</td>
</tr>
<tr>
<td>Exit pupil aperture</td>
<td>&gt; 5 (mm)</td>
</tr>
<tr>
<td>Eye relief</td>
<td>&gt; 60 (mm)</td>
</tr>
<tr>
<td>Viewpoint offsets</td>
<td></td>
</tr>
<tr>
<td>Horizontal</td>
<td>216.9 (mm)</td>
</tr>
<tr>
<td>Vertical</td>
<td>0 (mm)</td>
</tr>
<tr>
<td>Transparency/Reflectance (lambda = 550 (nm))</td>
<td></td>
</tr>
<tr>
<td>Eyepiece &amp; Objective lens</td>
<td>0.98</td>
</tr>
<tr>
<td>Prism</td>
<td>0.96 ~ 0.98</td>
</tr>
<tr>
<td>Mirror</td>
<td>0.92</td>
</tr>
<tr>
<td>Total</td>
<td>0.75</td>
</tr>
</tbody>
</table>

#### 6.2. Experiment

Fig. 15 shows ELMO-2 with an optical see-through HMD (Olympus Mediamask), from which the polarizing filters and LCD shutters have been removed. Color virtual images are rendered by a PC and sent to the HMD via the PC’s video output as field-sequential stereo images. Stereoscopic mask patterns are rendered on the left- and right-hand visible areas of the LCD panel (about 140 × 140 pixels each) via the PC’s RGB output. Fig. 16 shows a few examples of composed stereo images as captured by a pair of small cameras (Toshiba IK-CU43). In this case, a cross-shaped virtual object is superimposed on the real white box. Through this experiment, we confirmed that ELMO-2 is capable of presenting stereoscopic images with correct mutual occlusion. We have also confirmed that this scene is correctly perceived stereoscopically by naked-eye observation.

### 7. ELMO-3: the third prototype display

#### 7.1. Overview

As we mentioned in Section 1, the depth information of the real scene and an appropriate display are both required to realize mutual occlusion. In our previous work, we had assumed that the real environment was static except for the presence of a few movable objects with 3D trackers, so that the system was able to model the geometry of the real scene. However, in order to use ELMO in unknown or dynamic environments, the
system needs to acquire a dense depth map of the real scene in real-time. A number of real-time depth acquisition mechanisms have been proposed to date [5,8,9]. Of these, only the passive approaches were considered suitable for ELMO, as we wished to keep the advantage of ELMO not affecting the real environment. Considering the resolution of the depth map and processing speeds, we decided to incorporate the FZ930 real-time stereovision system developed by Komatsu Corporation in ELMO-3.

In its original form, the FZ930 has nine cameras and employs a multi-baseline stereo-matching algorithm to improve its sensing accuracy. Fig. 17 shows how the number of cameras affects the accuracy of sensing by the FZ930. As the figure shows, matching errors are drastically reduced when five or more cameras are used. Considering the size of the system, the layout and number of cameras, we decided to use five cameras, each with a focal length of 6 mm and a baseline of 79 mm. The cameras are arranged in an X-shape and the center camera is placed at the center of the two eyes. The appearance of ELMO-3 is shown in Fig. 18. Fig. 19 shows the relation between disparity on image planes and detected distance. The maximum disparity (29 pixels) corresponds to the distance of 831 mm. Fig. 20 shows the relation between the resolutions of distance and depth. Depth resolution varies in proportion to the square of the distance. At the detectable minimum distance (831 mm), depth resolution is 5.5 mm. Horizontal and vertical fields of view of each camera are
about 40° and 30°, respectively, so they are sufficiently larger than those of ELMO’s optics.

Another new feature of ELMO-3 is that there is no longer a need to use a normal see-through HMD to produce color CG, as beam splitters and color LCD panels now replace prisms No. 2 shown in Fig. 12. The detachable 2.5-inch LCD monitors of hand-held video cameras (Sharp, VL-PD7) are used as the color displays. The reflectance and transmittance of the beam splitters are set to about 70% and 25%, respectively; the overall transmittance of the folded optics system without the LCD panel is 50%. This makes real scenes seem brighter than those in the previous prototype displays.

7.2. Experiment

Fig. 21 shows the environment used for the experiments with ELMO-3. As was earlier indicated, ELMO-3 itself requires no particular environmental factors to realize mutual occlusion. However, due to the nature of stereovision algorithms, the accuracy of an acquired depth map depends on the presence of patterns in the real scene. For example, if the background includes no pattern and is monotone, depth maps become unreliable. In this case, we located a partition board with random square patterns to ensure reliable depth maps. Since ELMO-3 has three displays, one LCD panel for the stereo mask patterns and two LCD modules for the color images, the system has to output three images at once. By taking account of the fact that only two 140 × 140 pixel areas are required for the stereo mask pattern, and that the resolutions of the color LCD modules are about a quarter of that of a VGA display (320 × 240 pixels), the system renders the three images onto a single VGA screen, as is illustrated in Fig. 22. The left and right color images are first rendered on two quarters of the VGA screen. Each of these is then scan-converted to NTSC and sent to the corresponding LCD modules. Captured depth maps are rendered three dimensionally and transparently to overwrite Z-buffer, from the positions of the left and right eyes to match the real scene from the respective viewpoints.

Fig. 23 shows four images captured through this setup. Fig. 23(A) is a typical ‘ghost’ image without occlusion. Fig. 23(B) shows opaque virtual images without depth information. In this case, all virtual objects occlude the real scene. Fig. 23(C) shows the scene with semi-transparent virtual objects with depth.
information of the real scene. Fig. 23(D) shows opaque virtual objects with correct occlusion attributes. We confirmed that ELMO-3 shows mixed-reality environments with correct mutual occlusion in real-time. Since the original depth map is acquired from the center of the two eyes, the converted depth maps from the left- and right-eye viewpoints are not perfect, particularly around vertical edges. However, this has little effect that a good overall impression is maintained.

8. Conclusions and directions for future work

This paper is a review of our research on the ELMO optical see-through display. In this paper, we first pointed out the well-known occlusion problem of conventional optical see-through displays, and then described the design of optics for optical see-through displays that are able to present mutual occlusion of objects in real and virtual environments. Our display,
ELMO, can be used anywhere including outdoors, and it is able to display any color, including black, with high fidelity. This is not possible with conventional optical see-through displays. We also presented our prototype displays and gave some experimental results. Through the empirical studies, we confirmed that our display design solved the occlusion problem. Such a display will be especially useful when users want to avoid any degradation in the quality with which the real scene is presented. Medical and outdoor applications will suit this display.

Although the basic concept of our optics is very simple, there is no report of anyone previously trying to build it. To realize the basic ideas, considerable effort is required, e.g., selection of the prisms and their placement in terms of obtaining a wide-enough field of view within a reasonably compact body. As far as we know, we are the first group to have actually built and reported practical displays based on the principles we have described, and to have shown their practicability in the context of mixed reality.

One remaining problem is that the attenuation of light by the LCD panel is still noticeable. However, the overall transparency can be doubled or further increased by replacing the LCD panel with a panel of high transparency and wide aperture. Scotopic adaptation may also function well. As for response times and resolution, we will be able to overcome these problems.
in the near future, for LCD panels that have response times of less than 2 ms and resolution of over 200 dpi have recently become commercially available. The resolution and system latency of the stereovision system should improve in proportion to the qualities of the LCD modules. The size of the display, which is still cumbersome, can be greatly reduced by using smaller LCD panels (the main reason for ELMO’s current bulkiness is the 10.4-inch panel). It will also be possible to miniaturize the optical parts themselves at the expense of the exit pupil apertures (it gets harder to adjust one’s viewpoints). Such miniaturized displays would be effective for use in wide areas and outdoors, which will require some good tracking techniques [13,14]. To make better use of the design of our optics, it would be useful to employ an accommodative compensation mechanism [15]. We would like to pursue all of these issues. Another possible solution for mutual occlusion would be to use DMDs (Digital Micromirror Devices). Using DMDs should eliminate most of the remaining problems, though the optical design would be extremely difficult due to the small tilt angles of the mirrors.

Acknowledgements

We thank the Komatsu Corporation for technical support.
References


