User Experience with Alignment of Real and Virtual Objects in a Stereoscopic Augmented Reality Interface

Ming Hou

Ergonomics in Teleoperation and Control (ETC) Laboratory
Department of Mechanical and Industrial Engineering
University of Toronto, ON, Canada M5S 3G8
ming@mie.utoronto.ca

Abstract

This paper reports two virtual pointer alignment experiments carried out using a stereoscopic augmented reality interface. The purpose was to evaluate users' sensitivity to surface texture, target position at designated probe points on a cylinder real object surface, virtual pointer form and binocular disparity. The results confirmed the main findings from a previous study: that both surface texture and target position have significant influences. Subjective evaluation of virtual pointer form revealed that a three dimensional pointer is preferred over one and two-dimensional pointers. The contributions of size cue and resolution to binocular disparity effects are also discussed in relation to interface design issues for augmented reality applications.

1 Introduction

Mixed Reality, the concept of seamlessly blending real and virtual images within a single display medium [15], is rapidly becoming feasible through advances in computer, communications, and human-computer interface technologies [2]. As one of the typical examples of this, Augmented Reality (AR) blends virtual computer-generated information (graphics or text) with images of unmodelled real world objects.

The information conveyed by the virtual objects in AR is intended to help the human user have a better understanding about the real world, and thus perform associated real-world tasks better. Applications of AR are thus very broad, encompassing medicine, health care, telerobotics, maintenance and repair, education, entertainment, urban planning, and military aircraft navigation and targeting, etc. For example, 'knowledge enhancement' using AR has been investigated for aiding the user's perception of interactions with real world systems [8]. One clear example in health care is the ability to visualise and localise an aneurysm or a tumour relative to the patient's surrounding anatomy, and then measure its dimensions for better surgical outcomes [10]. The enhanced 'virtual hand' has been used to facilitate the investigation of goal-directed human hand movement, including studies related to computer-aided surface design, computer animation of the human body, gestural input, and "smart" interfaces that recognise a user's intent by analysing hand and body movement as part of the user interface of telerobotic systems used in some advanced manufacturing applications [12] [16]. Another instance in telerobotics is the concept of communicating to a robot where "there" is, for execution of a "put that there" kind of instruction [9]. For educational purposes, the MagicBook enables children to pursue the fantasy of being inside the books they read, seeing the characters eye to eye and becoming part of the story, by exploring the transition between physical reality, augmented reality, and immersive Virtual Reality (VR) in a collaborative setting [3].

One of the great potentials of stereoscopic video based AR is to make remote 3D measurements at low cost [10]. However, due to the nature of such interfaces, where combining both virtual and real objects into a single medium is frequently accompanied by perceptual ambiguities about the exact location of the real objects, accurate measurements often remain difficult to accomplish [5]. In the current paper we first explain the nature of these so-called "surface effects" in aligning real and virtual objects in a stereoscopic AR interface. We then report on two psychophysical experiments on users' sensitivity to surface features and virtual object features. Finally, some suggestions on AR interface design will be presented, based on the experimental results.

1.1 Nature of Augmented Reality

In order to solve the ambiguity problem in AR environments, we must first understand the nature of Augmented Reality. In the past few years, Virtual Environments (or Virtual Reality) has attracted a great deal of attention, the basic aim typically being to immerse a user to some extent within a completely computer-generated "virtual world". In contrast, with Augmented Reality the underlying image is composed of real world objects, with computer graphics superimposed. In other words, the real world is supplemented, rather than replaced. In the ideal case, it should seem to the user that the real and virtual objects coexist.

If we view AR within the context of a *Reality - Virtuality (RV) continuum*, it is straightforward to define it as a subset of the class of Mixed Reality (MR) displays [14]. As illustrated in Figure 1, the RV continuum is presented as a framework for describing the spectrum of cases that define whether the primary world being experienced by an observer is *real* or *virtual.* On the left end of the continuum is the purely real world, which can be visually displayed by scanning, transmitting and reproducing image data, as is the case with ordinary video displays – without the need for the display software to "know" anything about the objects in the real world. Another way to present real world objects is by viewing real world scene either directly or via some optical medium. On the other hand, purely virtual images, on the right end of the RV continuum, can be produced only if the computer display system generating the images has a *quantitative model* of the objects being portrayed.

As an excellent example, MagicBook enables people to read a book in the real world while experiencing virtual images that appear attached to the real book pages [3]. Since readers can fly into the virtual images and experience the story immersively, this AR book allows people to experience the full expanse of the RV continuum. In principle, Augmented Reality (AR) enables one to make virtual images appear before the viewer in a fairly well specified location in real space. These images can display task related information, or can serve as interactive tools for measuring or controlling the environment. In contrast, Augmented Virtuality (AV) displays are those in which a primarily virtual environment is enhanced, or augmented, through some addition of real world images or sensations. These additions can take the form of directly viewed objects, where viewers might see their own bodies instead of computer-generated simulations, as is typical with surround type virtual environments, where one might reach into the scene to grasp an object with one's own hand.

In addition to the techniques of combining the virtual with the real in a display, one can also look at the RV continuum in terms of how much we actually know about the objects and the world in which they are displayed. For example, purely real world displays can be looked upon as unmodelled scenes about which the displaying hardware/software knows nothing.

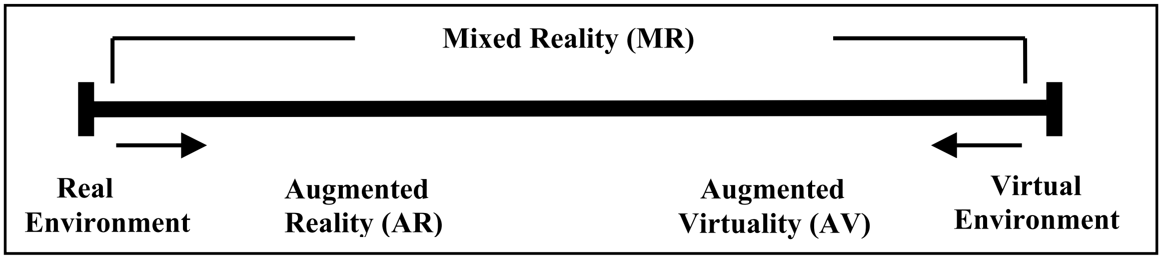


Figure 1. Reality—Virtuality (RV) Continuum

Completely virtual images, on the other hand, derive from modelled worlds, about which everything must be known, in order to display the images. Mixed Reality (MR), which is between these two extremes, encompasses partially modelled worlds, about only partial knowledge is available. As a subset of MR, Augmented Reality (AR) combines unmodelled real world images with computer-generated (modelled) virtual objects.

1.2 Research Motivation

Because the superimposed virtual objects are generated by computer, it is reasonable to assume that one should be able to generate a virtual object to appear at a known location within the real world, *in spite of the fact that no knowledge is available about what is actually present at that location in the real world*. For example, when using a stereoscopic video camera to view a real world scene, it is straightforward to present a graphical object of arbitrary size at an arbitrary position {x,y,z} relative to the optical centre of the camera system (assuming that the camera calibration parameters are known), without having any knowledge of whether or not any real world object also exists at that location. (Rotation or translation of the camera in this case would cause the world to appear to move relative to the monitor, while the virtual object remains fixed relative to it.)

By taking advantage of the knowledge of computer-generated virtual object locations, AR offers the potential for using stereoscopic displays to assist human users to acquire more information about real environments in 3D. This is especially useful for inaccessible and/or potentially hostile working places, where knowing what is present and where everything is located in depth is crucial for accomplishing certain tasks. For example, a neurosurgeon may need to estimate the dimensions of an aneurysm or tumour intraoperatively (rather than using pre-operative scanning), so that it can be rendered harmless with a properly sized surgical clip. Alternatively an operator may need to align a remotely controlled robot relative to a rock jam or ore seam inside an automated mine.

Although placing a simple two-dimensional scale in the frontal-parallel (X-Y) plane may somewhat facilitate distance measurements between objects within the same depth plane, the problem of measuring distances between structures at different depth planes remains. To overcome this difficulty and address such absolute measurement and specification problems in 3D, a *Virtual Tape Measure* (VTM), was developed, based on Augmented Reality through Graphic Overlays on Stereo-video (ARGOS) [13].

The VTM is an easy-to-use 'no touch' measuring device intended to enable users (e.g., surgeons) to make 3D measurements of any structure in the operating field. Based on prior calibration of the camera system, absolute distances/dimensions can then be computed, using the camera frame of reference. The VTM therefore does not have to be registered to the real world reference frame and does not rely on the availability of any pre-operative imaging. It can thus be used with any real-world stereo video image without having to determine or track the location and orientation of the cameras' co-ordinates in relation to that real world. The VTM needs only to be calibrated in real-world units in order to give accurate measurements of dimensions and distances between real objects in that world.

At the end of the VTM is a *virtual stereoscopic pointer,* which must be interactively manipulated and aligned with features of interest in the stereo video image to make measurements of dimensions and distances between real objects in the 3D video scene. An earlier laboratory experiment has shown that people can accurately align such *virtual* pointers with *real* targets in the stereo video image as well as they can align *real* pointers with *real* targets [4]. In another investigation of the VTM for assisting micro-surgery, performance evaluation has suggested that the VTM for the operating microscope is an accurate and precise measurement device under certain conditions [10].

In order to make accurate measurements in stereoscopic AR environments, the VP has to be precisely aligned with designated real object surface features. However, whenever such surface features are not well defined, and/or of high visual contrast, such alignment is much more difficult. This may result in ambiguity in alignment, and thus inaccurate 3D measurements. This is especially true, for example, for anatomical objects which one typically encounters in surgical environments, where surfaces are rounded, shading is uneven and textural cues are ambiguous.

Although stereoscopic displays provide the general advantage of enhanced (relative) depth perception, it has been found that the perceived locations of virtual objects depend on their locations relative to adjacent real objects in AR displays [6] [9] [11]. In our own ARGOS display, which has been developed for measurements in unstructured environments, whenever the virtual pointer (VP) goes behind the surface of a real object, it fails to disappear, due to the fact that the display system has no knowledge of the presence of the (unmodelled) real object, which would otherwise occlude the VP. In conflict with this phenomenon is the mechanism of binocular fusion, which is necessary for the observer's perception of a single fused image in depth. The result of failure to occlude when virtual objects are placed 'behind' real ones is frequently a double image, because the brain is no longer able to reconcile the (absence of) occlusion information and at the same time fuse the left and right images for both the real object (video) and the VP (graphic). In other words, this loss of fusion, when it occurs, is due to the perceptual conflict between consistent binocular disparity information and inconsistent occlusion information, resulting in some kind of double images, especially when viewed statically.

The practical effect of this *pseudo-occlusion* phenomenon is spatial ambiguity when determining the VP location relative to real objects. It is interesting to note, however, that this effect does not occur when the images are displayed monoscopically, since maintaining binocular fusion is not an issue in that case. (One must not forget, however, that in the monoscopic case reliable 3D measurements are not possible.)

Besides the 'pseudo-occlusion' surface effect, other depth cues can also affect visual perception of the surface of real objects in stereoscopic AR environments, such as binocular disparity, display resolution, VP size and form, etc. In addition, surface characteristics such as texture, shading and luminance are also likely to influence the object perception, and thus pointer placement.

In this paper, two psychophysical experiments are reported. The general goal of this research is to test users' sensitivity to surface features for virtual-real object alignment tasks in stereoscopic AR environments. Ultimately we hope to determine whether a method can be developed for improving current VP alignment performance for arbitrarily oriented 3D curved surfaces. Although one may distinguish among three display modes for Augmented Reality – 'direct view see-through', 'video see-through' and 'monitor-based' video systems – the experiments reported here were conducted using a monitored-based stereoscopic video AR system.

2 Hypotheses and Methodology

The specific objective of the research is to study the influence of a particular set of visual characteristics of curved real object surfaces on the ability to align a computer-generated (virtual) stereoscopic pointer with real stereo video objects in an AR interface. Based on a series of exploratory studies, we propose four hypotheses about one's ability to perform such tasks:

- it is possible to exploit the breakdown of fusion phenomenon to more easily localise targets on curved surfaces of real objects which contain textures with relatively high density;
- the orientation of a curved real object surface, in terms of the direction of the normal to the surface relative to the observer (that is, relative to the stereo video cameras), will affect alignment performance;
- the form of the virtual pointer (VP), in terms of its dimensions (i.e. 1D vs 2D vs 3D) will have effect on performance; and
- the polarity of binocular disparity (i.e. crossed vs uncrossed) will also influence alignment performance.

Two experiments were conducted to test these hypotheses. A psychophysical *method of adjustment* was used for the task of aligning the VP with designated targets on the surface of a real cylinder image, all of which were displayed using stereoscopic Augmented Reality. The following conditions were manipulated as independent variables:

- surface texture,
- VP orientation relative to the real target surface,
- target position on the surface,
- angular displacement of the surface normal relative to user's viewpoint, and
- binocular disparity.

As Figure 2 illustrates, the target stimuli comprised a set of alternating field stereoscopic images of a 46 cm diameter cylinder, pre-recorded using a calibrated pair of JVC cameras and displayed on a Silicon Graphics Indy workstation. The stereo images were viewed through synchronised IMAX liquid crystal shutter glasses. The subjects' viewing distance was 48 cm from the screen. Two target cylinders were used, both with textures consisting of white dots randomly dispersed on a black background (generated using a random dot Stereogram software package), but with different texture densities.

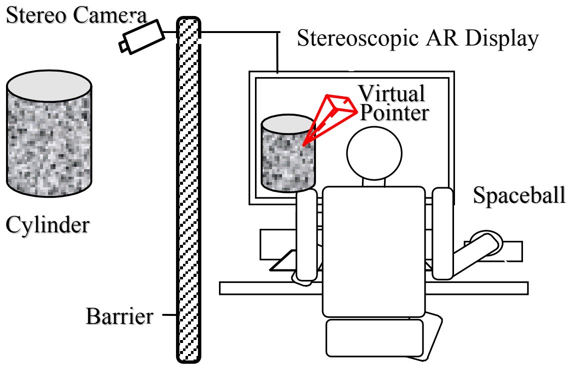


Figure 2. Experimental Set-up

3 Experiment 1

This experiment was designed to investigate the breakdown of fusion phenomenon, by examining the effects of surface texture, target position, VP orientation, and binocular disparity on real and virtual object alignment.

3.1 Method

A 2x2x3x3 factorial experimental design was used, comprising a combination of two textures (with high and low density), two target positions (on the centre of the surface facing the observer and on the right side along the normal lateral plane), three VP orientations relative to the surface (vertical, horizontal and diagonal), and three image disparities (crossed, 0, and uncrossed). The dependent variable measured was the longitudinal error between the final VP placement and the actual position of the target on the real object surface.

3.2 Stimulus and Apparatus

The stereo cameras were located 92 cm from the front surface of the cylinder. Three different camera convergence distances were used:

- 6 cm in front of the surface,
- at the surface of the cylinder, and
- 6cm behind the surface of the cylinder.

Alignment of the VP with a target at the proximal surface of the cylinder (i.e. correct placement) therefore corresponded to crossed, no disparity, and uncrossed disparity, respectively.

The VP was a three-dimensional computer generated wireframe arrow that appeared to hover within the stereo image upon which it was superimposed, as illustrated in Figure 2. Three different orientations of the pointer were used:

- vertical,
- horizontal and
- diagonal.

For the first two of these, the pointer remained within a 2D plane tangential to the cylinder surface, as shown in Figure 3. The diagonal pointer was located within a plane that was at 45º to both the tangent plane and the normal vector.

The VP was controlled with a Spaceball operating with only 3 translational degrees of freedom.

3.3 Subjects

N=10 university students (6 male and 4 female) participated in the experiment, following screening using the RANDOT™ Stereotest. None of the subjects knew about the aims or the design of the experiment. Where necessary, subjects wore appropriate optical correction.

3.4 Procedure

The experimental task was to localise points on the cylinder surfaces by manipulating the VP, for the three VP orientations, two surface textures, three camera configurations (binocular disparities) and two target positions on the real surface. Subjects used the Spaceball to move the VP in and out along X, Y, Z axes, until it appeared to just touch the surface of the cylinder exactly at the designated target position. They then informed the experimenter that the alignment had been completed. Each experiment consisted of 6 randomised replications for each condition, for a total of 216 judgements. The experiment, including practice, took place over a span of three days, with each session lasting approximately two hours per day.

.

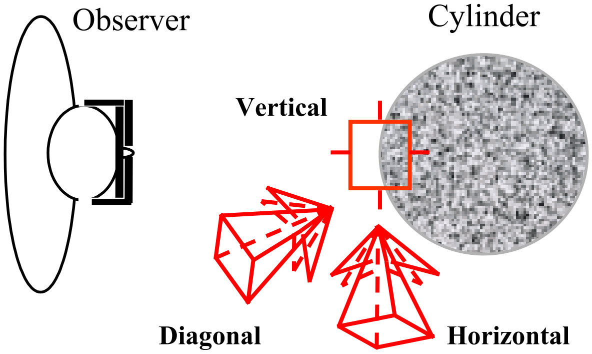


Figure 3. VP Orientations (top view)

3.5 Results and Discussion

The principal results of this experiment are summarised in Figures 4 and 5. From Figure 4, it is evident that, as hypothesised, surface texture has a highly significant effect on placement accuracy (F(1,9) = 619.70, p<0.001). These results confirm earlier observations in which it appeared that, whenever the VP is placed in *front* of the surface of a real (video) object, i.e. the case in which the binocular disparity and occlusion cues are consistent, subjects are able to shift attention easily back and forth from the pointer to the surface. Whenever the pointer is moved *behind* the surface, however, the two depth cues begin to conflict.

This experiment shows that the magnitude of this conflict is very much dependent on the visual features of the surface. Whenever the surface in question is *sparsely textured*, there are relatively fewer features to drive the stereoscopic fusion cue, so the observer is more easily able to reconcile the two conflicting cues and fuse both the real and virtual images. The result in such cases is that the object surface appears *transparent,* and it is thus more difficult to detect the transition through the surface. On the other hand, whenever the pointer moves behind a *highly-textured* surface, the observer is less able to overcome the tendency to fuse the surface features stereoscopically. In that case it is more difficult to reconcile the fact that the fused pointer is behind the fused surface yet still visible – a "perceptual impossibility". As a result of these conflicting binocular disparity and apparent occlusion cues, the tendency is to shift attention back and forth between the VP and the surface features, resulting in breakdown into either a double image of one of them or alternation between the two fused images. Because of the conspicuous nature of this conflict between the two disparate cues, subjects are ironically more easily able to move the VP in and out until the conflict disappears – at the surface of the real object. This is why, we believe, as seen in Figure 4, the placement error for the highly textured surfaces is less than that for the less textured surface.

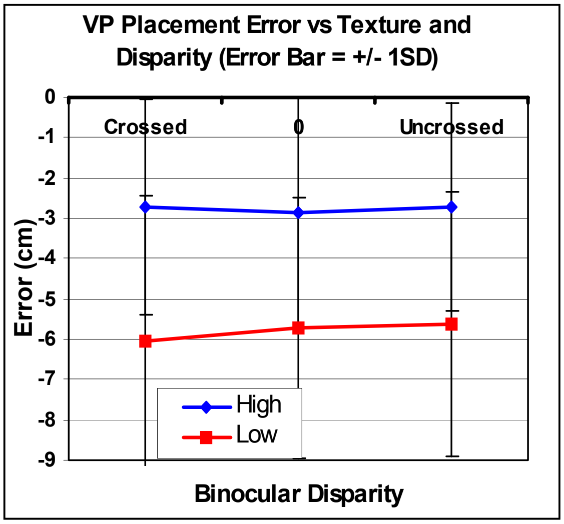


Figure 4. Effect of Surface Texture (High vs Low) and Binocular Disparity. (Negative values indicate positioning error in front of cylinder surface)

Figure 5 shows that there is an interaction between the surface texture and the target position on the surface along the normal lateral plane (F(1,9) = 246.33, p<0.001). When the target is at the central position on the surface, that is, facing the viewer directly, positioning error with the highly textured surface is essentially the same as with the low textured surface. However, when the target is off to one side (on the right side, 20º from the central target in our case), the placement errors for high and low textured surfaces are significantly different, with error for the low textured surface being almost 4 times as great as for the highly textured surface error (F(1,9)=30.22, p<0.001).

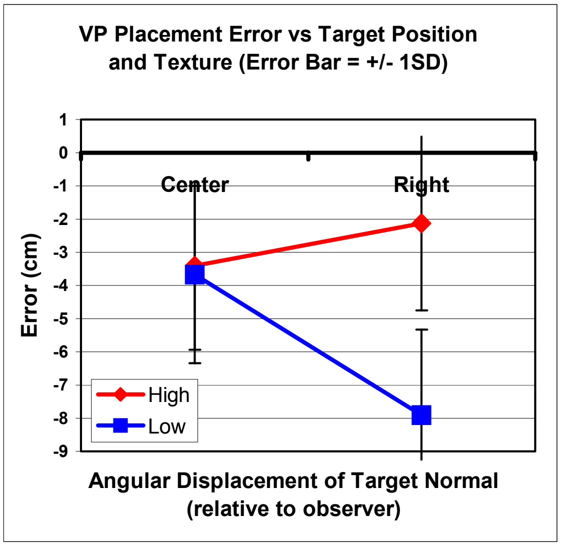


Figure 5. Effect of surface texture (High vs Low) and target position (Centre vs Right)

This result implies that, since the observers' viewing angle was different for the two targets, the perception of the local surfaces at the two sites was also different. This implies further that one can expect to perform better when placing a virtual pointer at a point along the centre of a surface relative to the observer's normal straight-on viewing angle (that is, looking straight at the surface) as compared to any other angle relative to the normal lateral plane. This finding is perhaps not intuitive, since one might otherwise expect superior performance when one is able to watch the pointer approaching a surface more from the side, rather than straight on. We believe, however, that a large part of the performance in this respect was due to the *form* of the graphic pointer, a topic of our second investigation.

Since the results from our pilot study (with 2 subjects) showed that the smallest errors in localising surface positions in the video image were obtained when the VP was diagonally oriented, we speculated that VP orientation (horizontal vs vertical vs diagonal) would have a significant effect on the alignment task. However, there was no statistical significance from the ANOVA (F(2,18)=1.015, p=0.38) for the full experiment.

Another prior hypothesis was that the disparity (crossed vs 0 vs uncrossed) would also have a significant effect; however, the ANOVA revealed no statistical significance (F(2,18)=0.40, p=0.68).

4 Experiment 2

A negative bias (where the VP was typically placed in front of the real surface), with relatively large error ranging between 2 to 6cm was noticed from the results of experiment 1 (see Figure 4). In reviewing the geometric design of the virtual pointer, it was noted that a possibility existed that the geometrical form of the pointer might have influenced this result. In the first experiment, three different orientations of the arrow-shaped VP were used: vertical, horizontal and diagonal, and each of those had four prongs on the tip of the arrow (see Figure 3). For the first two orientations, the pointer remained within a 2D plane tangential to the cylinder surface. The diagonal pointer was located within a plane that was at 45° to both the tangent plane and the normal vector. Thus, there would be at least two thirds of the trials in which one of the *prongs* touched and went behind the surface before the *tip* of the VP arrow (which is the measuring point) touched. The cue conflict caused by the prong and the surface might thus have been perceived first, causing subjects to think that the VP had already touched or had gone behind the surface, even if the tip had not in fact touched the surface yet. We believe that this could have been the reason why the localisation error was biased, causing the detected surfaces to appear to be localised in front of the real surfaces.

In order to test this hypothesis, and thus to confirm if subjects explores the surfaces by using fusion breakdown as a supplementary cue (in addition to binocular disparity matching), we removed the prongs on the VP arrow tip in the second experiment, and redesigned its form (as an independent variable). Based on the definition of stereoscopic cursor shapes proposed by Barham and McAllister [1], three types of VP, with one, two, and three dimensions respectively were designed (see Figure 6):

- LINE,
- AREA,
- VOLUME.

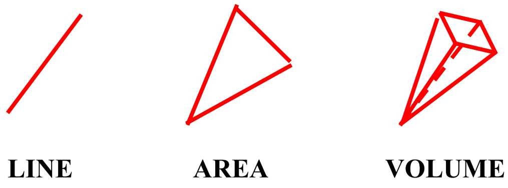


Figure 6. VP Form for the Second Experiment

With these three types of VP forms, the goal of the present experiment was to test these in conjunction with surface texture, target position, VP form, and binocular disparity effects again. In addition, we also conducted a subjective evaluation of the VP form, to investigate subjects' preferences.

4.1 Method

Similar to the first one, this experiment consisted of a psychophysical method of adjustment task, involving alignment of the VP with designated targets on the surface of the real object image, all of which were displayed using stereoscopic augmented reality. A 2x2x3x3 experimental design was used, comprising a combination of two textures (high and low density), two target positions (on the centre of the surface facing the observer and on the right side along the normal lateral plane), three wireframe VP forms (one, two, and three-dimensions), and three image disparities (crossed, 0 and uncrossed). The dependent variable measured was the error between the final VP placement and the actual position of the target on the surface of the real object. This experiment had exactly the same procedure and trial numbers as experiment 1. Twelve university students (7 male and 5 female) participated as subjects.

4.2 Apparatus

Two target cylinders of 46 cm diameter were used, as in experiment 1, with the same high and low texture densities. The front surfaces of the two cylinders were located 82cm, 92cm, and 102cm from the stereo cameras, corresponding to *crossed* disparity, *no* disparity and *uncrossed* disparity, respectively. The stereo images of these cylinders were recorded and displayed on a SGI Indy workstation, and viewed through synchronised Imax liquid crystal shutter glasses at 48 cm from the screen.

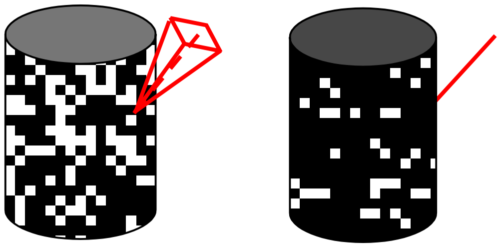
The three VPs were computer generated wireframe arrows that appeared to hover within the stereo image upon which they were superimposed (only one VP for each surface image). They were all diagonally oriented - as was the one in the earlier experiment - within a plane that was at 45° to both the tangent plane and the normal vector. The three VPs were controlled with a Spaceball operating with only 3 translational degrees of freedom.

4.3 Paired Comparisons

In order to obtain subjective assessments of the different VP forms, a paired comparison procedure was performed immediately after the trials were completed. Subjects were presented two interfaces side by side with different combinations of textured surfaces and VP forms. They were asked to manipulate the VP on both interfaces and give the ranking in terms of 'ease of use' (i.e. which interface was easier to use for localising the surface), 'transparency' (i.e. which surface appeared more transparent – in the sense of being able to see through it when the VP went behind it), and 'ease of fusion' (i.e. for which combination it was easier to fuse both the surface and the VP). There were a total of 15 randomised pairs of 6 combinations of VP forms and textured surfaces (see Table 1). Figure 7 illustrates one example of these pairs.

Table 1. **Image Combinations for Paired Comparisons**

Image #	**1**	**2**	**3**	**4**	**5**	**6**
Texture Density	High	High	High	Low	Low	Low
VP Form	LINE	AREA	VOLUME	LINE	AREA	VOLUME



High Density Low Density

Figure 7. Sample set of Paired Images

4.4 Results and Discussion

An ANOVA was performed on the experimental results, which are summarised in Figures 8 and 9. As hypothesised surface texture and target position had significant effects on placement accuracy (F(1,11)=11.14, $p<0.01$ and F(1,11)=98.19, $p<0.001$, respectively). This finding is consistent with what was found in the earlier experiment, in which highly textured surfaces elicited less placement error than low textured surfaces, and performance was better when the VP was placed at a point along the centre of a surface relative to the observer's normal straight view angle (that is, looking straight at the surface) as compared to any other angle relative to the normal lateral plane.

These figures also illustrate that the detected surface position tended to be behind the real surface (mean placement errors are positive), and that the placement errors were much smaller than those in the earlier experiment (less than 1cm). The post-experimental interviews also supported the hypothesis that subjects were using the fusion breakdown (as inferred through words such as: image "blurring" or "fuzzy images") as a signal to localise the surface. In other words, subjects pushed the VP into the surface until they saw a blurred or fuzzy image, then they moved the VP back until the fusion difficulty disappeared (perceiving a clear image of both the pointer and the surface), etc. Because it could be seen clearly when most part of the VP was outside (in front of) the surface (even though the tip was still inside – due to the transparency effect), they thought the VP was right *on* the surface.

Although binocular disparity statistically had a significant effect (F(2,22)=36.84, $p<0.001$) in the present experiment, it cannot really be claimed that the disparity cue has significant impact on the probing performance because two other factors were may have been introduced due to the specific experimental set-up. In order to keep the stereo cameras static, and thus to keep the same system precision and accuracy for the three different disparities, instead of adjusting the sensitive stereo cameras, the target cylinders were moved to different positions corresponding to crossed, 0, and uncrossed disparities, respectively. Thus, the images for the crossed disparity case (closest to the observer) corresponded to the largest cylinder size and highest resolution, whereas the images for the uncrossed disparity had the smallest cylinder size and lowest resolution. This may thus have been the reason why crossed disparity facilitated minimum error, and uncrossed disparity had maximum error in this case, as shown in Figure 10.

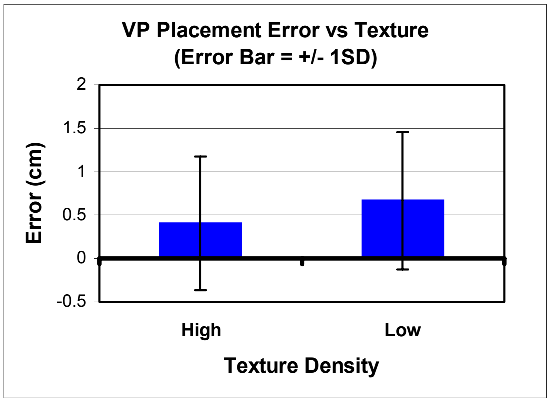


Figure 8. Effect of Surface Texture (High vs Low)

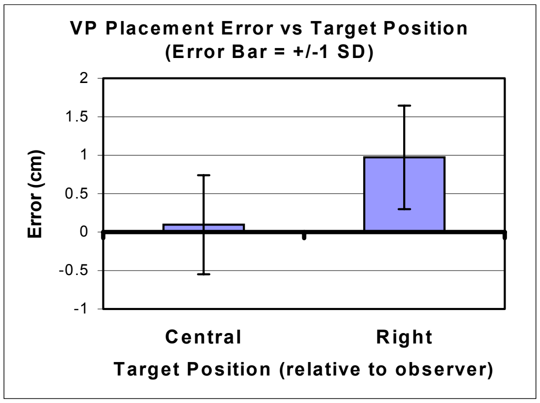


Figure 9. Effect of Target Position (Center vs Right)

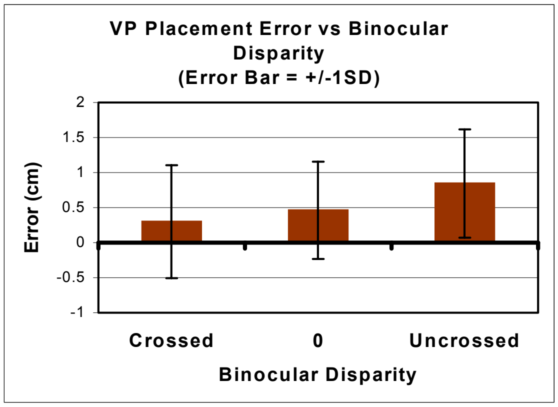


Figure 10. Effects of Binocular Disparity (Positive values indicate positioning error behind cylinder surface)

Another prior hypothesis was that the form of VP (1D, 2D, 3D) would also have a significant effect; however the ANOVA revealed no statistical significance (F(2,22)=1.655, p=0.214). On the other hand, the results from the paired comparisons did show subjects' preferences in terms of 'ease of use', 'transparency' and 'ease of fusion' for the 6 image combinations of VP forms and textures. Based on the proportions of choices of images in the paired comparisons over all 12 subjects, the probabilities for each choice were converted to z-scores [7]. The mean z-scores were then transformed linearly to represent the 'psychological distances' between each of the 6 image combinations, as shown in Table 2. It can be seen that subjects preferred low textured surface rather than highly textured surface (images # 4, #5 and #6 all have the highest z-scores), probably because it was less likely to encounter fusion difficulties whenever the VP was placed behind the low textured (more transparent) surface. The subjects apparently felt that it was easier to fuse both VP and surface features and control (use) the VP for low textured surfaces, even though they did not know that this tactic was accompanied by larger placement errors.

Furthermore, mean z-scores for 'ease of use' and 'ease of fusion' indicated that subjects preferred the three-dimensional (VOLUME) pointer rather than the one and two-dimensional VPs, for both low and highly textured surfaces (in other words, the z-scores of images #3 and #6 are larger than for images #1, #2 and images #4, #5, respectively). Since the VOLUME VP has more features along the X, Y, and Z axes, this may have facilitated more depth perception with the stereoscopic display relative to the other two VPs, resulting in higher ratings. From the 'transparency' rating, it can also be noticed that the LINE VP is facilitates the transparency effect (for highly textured surfaces, the z-score for image # 4 is larger than for images #5 and #6; and for low textured surface, image #1 is greater than for #2). Overall, the VOLUME VP is the most favourable one, and the LINE VP is the least favourable one.

This experiment re-examined the effects of surface characteristics, and confirmed not only surface texture and target position effects on the alignment between real and graphic objects, but also supported the postulated exploitation of the conflict between binocular disparity and occlusion depth cues. Subjective judgement of the VP forms also provided users' preferences for virtual object design for AR interfaces.

Table 2. **Results of Paired Comparisons**

	Linearly Transformed Mean Z Score					
Ease of Use	0	1.15	1.71		2.26	3.16
Image #	1	2	3	4	5	6
Transparency	0	0.25	0.89	2.97	3.82	4.37
Image #	2	1	3	6	5	4
Ease of Fusion	0	0.69	0.73	2.08	2.14	2.26
Image #	1	2	3	5	4	6

5 Conclusion

This paper has endeavoured to illustrate a number of perceptual issues related to stereoscopic Augmented Reality interface design. In the two experiments reported, the effects on a real-virtual alignment task of surface texture, target position, VP orientation and form, and binocular disparity were reported. Due to the interaction between real and virtual object images, perceptual conflicts between object occlusion and binocular disparity cues cause spatial ambiguity to occur. This can make it difficult to make accurate 3D measurements in stereo AR environments. Ordinarily one would assume that such conflicts would be considered a bad thing. However, it has been shown here that in our particular case the fusion breakdown can be regarded as an extra cue for detecting interactions between real and virtual objects, and thus for localising the positions of real objects. From a practical point of view, this is a positive feature of such interfaces. Coupled with subjective evaluation of the VP forms, user experiences suggested that use of a 3D virtual pointer is enhanced in the vicinity of highly textured (curved) surfaces, for targets which are central to the surface. Future studies should investigate such alignment performance for AR interfaces as a function of other factors, such as motion parallax and perspective cues.

Acknowledgements

The author would like to thank Professor Paul Milgram for his guidance and support on this research. The author is also grateful for the support of Natural Sciences and Engineering Research Council (NSERC) and the Institute of Robotics and Intelligent Systems (IRIS).

About the Author

Mr. Ming Hou is a PhD candidate (anticipated December 2001) working in the Ergonomics in Teleoperation and Control (ETC) Laboratory, in the Department of Mechanical and Industrial Engineering, at the University of Toronto. His research focus is on perceptual issues associated with stereoscopic augmented reality displays as well as design issues of 3D interface. He is a member of ACM and HFES. Further information about him can be located via the website: http://www.etclab.mie.utoronto.ca.

References

[1] P. T. Barham and D. F. McAllister. *A comparison of stereoscopic cursors for the interactive manipulation of B-splines*. *Proc. of SPIE 1457- Stereoscopic Displays and Applications II*, pages 18-26, 1991.

[2] W. Barfield and T. Caudell. *Fundamentals of Wearable Computers and Augmented Reality*. Lawrence Erlbaum Associates, Publishers, 2001 Mahwah, New Jersey.

[3] M. Billinghurst, H. Kato, and I. Poupyrev. *MagicBook: Transition Between Reality and Virtuality*. Demonstration at CHI 2001, ACM Conference on Human Factors in Computing Systems, Seattle, WA, pages 25-26, April 2001.

[4] D. Drascic and P. Milgram. *Position Accuracy of A Virtual Stereographic Pointer in A Real Stereoscopic Video World*. In *Proceeding of SPIE 1457 – Stereoscopic Displays and Applications II*, pages 302-313, 1991.

[5] D. Drascic & P. Milgram. "Perceptual issues in Augmented Reality", *Proc. SPIE, Vol. 2653-19, Stereoscopic Displays and Virtual Reality Systems III*, Jan.-Feb. 1996.

[6] S. R. Ellis and U. J. Bucher. *Distance Perception of Stereoscopically Presented Virtual Objects Optically Superimposed on Physical Objects by A Head-mounted See-through Display*. In *Proceeding of Human Factors and Ergonomics Society 38th Annual Meeting*, pages 1300-1304, 1994.

[7] T. Engen. Psychophysics: II. *Scaling Methods*. Chapter 3 in J. W. King and L. A. Riggs (ed's), Woodworth & Schlosberg's *Experimental Psychiology*, third edition, Holt, Reinhart & Winston, pages 47-86, 1971.

[8] Feiner, B. MacIntyre, and D. Seligmann. *Knowledge based Augmented Reality*. In *Communication of the ACM*, Vol.36, No.7, pages 53-62, 1993.

[9] M. Hou and P. Milgram. *Effect of surface characteristics on alignment of graphic and real objects in a stereoscopic augmented reality environment*. *Proc. of IEA200 / HFES2000 Congress*, Vol. 3, 476-479, 2000.

[10] M. Kim, P. Milgram, and J. M. Drake. *Computer Assisted 3D Measurements for Micro-Surgery*. *Proc. of Human Factors and Ergonomics Society 41st Annual Meeting*, 787-791, 1997.

[11] M. Y. Kim, J. M. Drake, and P. Milgram. *Virtual tape measure for the operating microscope: System specifications and performance evaluation*. *Computer Aided Surgery* 5, 148-155, 2000.

[12] C. L. MacKenzie, K. S. Booth, J. C. Dill, K. Inkpen & S. Payandeh. *Evaluating human goal-directed activities in virtual and augmented environments*. *Proceedings of IEA 2000/HFES 2000 Congress.* Santa Monica, CA, 546-551, October 2000.

[13] P. Milgram, D. Drascic, and J. Grodski. *A virtual stereographic pointer for a real three dimensional video world*. In *Human-Computer Interaction – INTERACT'90*. Elsevier Science Publishers, 1990.

[14] P. Milgram, and H. Colquhoun,. *A taxonomy of real and virtual world display integration*. Chapter 1 (pages 1-16) in Y. Ohta and H. Tamura (Ed's), *Mixed Reality – Merging Real and Virtual Worlds***.** Ohmsha (Tokyo) & Springer Verlag (Berlin), 1999.

[15] Y. Ohta and H. Tamura (ed's). *Mixed Reality: Merging Real and Virtual Worlds.* Ohmsha & Springer-Verlag: Tokyo, 1999.Y.

[16] Wang, & C. L. MacKenzie. *The role of contextual haptic and visual constraints on object manipulation in virtual environments*. *Proc. of Conference on Human Factors in Computing Systems CHI 2000*, ACM Press 532-539, April 2000.