A Sensitivity Study of Factors Influencing Real-Virtual Object Alignment Performance in Stereoscopic Augmented Reality Environments

Ming Hou
Defence R&D Canada

Paul Milgram
University of Toronto, Canada

A sensitivity study involving a real-virtual object alignment task was performed in a stereoscopic augmented reality environment, in which the known conflict between binocular fusion and object interposition cues was expected to play a major role. The object was to evaluate subjects' sensitivity to visual texture of a real hemisphere surface and to target position at designated probe points on that surface. Consistent with earlier experiments, the results indicate that: a) both surface texture and target position had significant effects both on real-virtual object alignment and on the estimate of surface normal direction; b) fusion breakdown caused by the conflict between occlusion and binocular disparity could have been used as an extra depth cue to detect virtual and real object interactions. In addition, a practical solution to improving remote 3D measurement accuracy is proposed.

INTRODUCTION AND MOTIVATION

Augmented Reality (AR), a display technology combining computer-generated (modelled) virtual objects with (unmodelled) real-world images, provides the potential for using stereoscopic displays not only as a means of information visualisation, but also as a tool for remote manipulation (Milgram et al., 1991; Cannon & Thomas, 1997). One of the great potentials of stereoscopic (as opposed to monoscopic) video based AR (as opposed to AR displays based on optical superposition) is its use as a virtual tape measure (VTM), which permits low cost but accurate 3D measurements within remote 3D video images (Kim, Milgram, Drake, 2000). Such measurements typically involve overlaying and aligning a stereographic virtual pointer (i.e., a calibrated computer generated 3D graphic cursor) with designated features of real-world objects within the stereoscopic video scene.

Due to the nature of such interfaces, where combining and aligning virtual and real objects within a single medium one frequently encounters perceptual ambiguities about the exact location of the real objects (Drascic & Milgram, 1996; Ellis & Menges, 1998). To address these ambiguities, and their potential effect on alignment performance, and to provide guidance for 3D AR interface design, two psychophysical experiments have been conducted (Hou, 2003; Hou & Milgram, 2000; 2001). Both involved interactive manipulation of a stereographic virtual pointer (VP) relative to designated target points on a cylindrical real object surface viewed via a dual camera stereoscopic video system.

In those earlier studies, it was found that, whenever the computer-generated VP went behind the surface of a real object, it remained visible. This was due to the fact that the display computer had no knowledge of the presence of any of the (unmodelled) real objects, thereby precluding occlusion of the VP as it went behind those surfaces. 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. Two possible consequences result from this failure to occlude when virtual objects are placed *behind* real ones. One of the results is a double image, due to the perceptual conflict between consistent binocular disparity information and inconsistent occlusion information. This occurs because the brain is no longer able to reconcile the (absence of) occlusion information while at the same time fuse the left and right images for both the real object (video) and the virtual object (graphic VP). The opposite result occurs when the conflict is apparently not too large. In that case, the brain continues to fuse both the real and virtual object images while at the same time accepting the presence of the unoccluded virtual object behind the real object surface. This results in an impression that the real object surface is somehow *transparent*.

Our earlier studies (Hou, 2003; Hou & Milgram, 2000; 2001) suggested that the fusion breakdown effect can be used as an extra depth cue for detecting the interaction between real and virtual objects under such circumstances. Especially when a VP is placed behind a *highly textured* surface, observers are less able to resist the tendency to fuse the surface texture features stereoscopically. In such situations it is more difficult to perceive transparency, that is, to reconcile the fact that the fused pointer is behind the fused surface yet still visible – a "perceptual impossibility", resulting in the breakdown of either the VP or the object surface into a double image. Although this conflict between occlusion and binocular disparity can bring some discomfort to perception of such AR environments, subjects can in fact exploit this breakdown phenomenon to their advantage when judging proximity interactions between virtual and real (video) surface images.

This explains why surface texture density played an important role for localising real object surfaces in the previous experiments. However, one question that remains to be addressed is: *how sensitive* will subjects be to different texture settings for performing accurate alignment tasks? In other words, is there any *optimal* value for a certain type of surface texture density to generate better performance?

Another important earlier result was that centrally located targets on a cylindrical real object surface facilitated alignment performance relative to angularly displaced targets. This was opposite to our original conjecture, that approaching a cylinder with a VP from its side would be more accurate than approaching at its centre, since the VP probe in the latter case would be along the line of sight and thus block the central target when it was aimed. Thus the question remained of whether this central line-of-sight result could be maintained for other kinds of real world objects with curved surfaces, especially those for which *the VP has to be rotated* in order to align with the direction of the normal to the curved surfaces.

In addressing these questions, the goal of the experiment reported here was to further refine our understanding of subjects' sensitivity with respect to texture density and surface curvature, and ultimately to explore the existence of potential engineering solutions for improving the accuracy of real-virtual object alignment in real-world operational environments for arbitrarily oriented curved surfaces.

METHOD

Stimuli As illustrated in Figure 1, a 60 cm diameter hemisphere was chosen as the real world stimulus, with five targets distributed across its surface (centres of white circles). Alternating field stereoscopic images of the hemisphere were pre-recorded, using a calibrated pair of JVC cameras located 158 cm from the front surface of the hemisphere. The stereo images were displayed on a Silicon Graphics Indy workstation and were viewed through synchronised IMAX liquid crystal shutter glasses. The subjects' viewing distance was 48 cm from the screen.

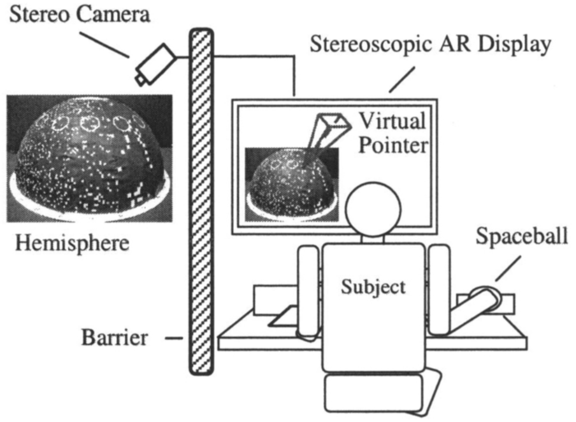

Figure 1. Experimental set-up (The centres of 5 circles in the stimulus image designate target positions for different textured sessions.)

Considering the special geometry of the hemisphere and camera viewing angle, all targets were placed within horizontally distributed circles at 50° elevation, to provide subjects an adequate view of all the targets on the hemisphere. One target was placed at the central meridian of the hemisphere surface along the line of sight (facing the observer). The other targets were distributed on both the left and right sides (relative to the observer) at spacing of 30° from the central meridian. The locations of the five targets are illustrated in Figure 2.

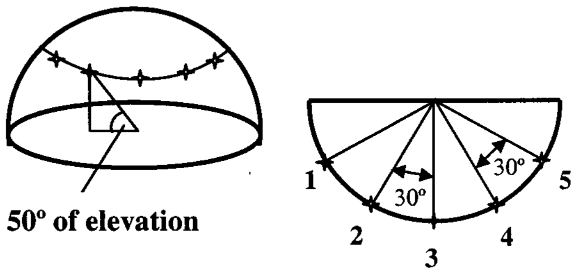

Figure 2. Target position (All five targets were centred at 50° elevation, and off-centre targets were placed at 30° intervals of longitude)

A 5x5 experimental design was used, comprising a combination of five texture densities and five target positions. Surface textures for all stimuli had the same type of random dot pattern, which consisted of white dots randomly dispersed on a black background (generated using a software package: Stereogram™). Five levels of texture density were used: {0, 10, 20, 30, 40}, where the numbers corresponded respectively to percent of the area on the grid containing *white* dot elements. The remaining area was all black on the surface background. Therefore, 0 means that the target area was all black, as there were 0% of white dots on the background. This provided a control condition for no texture at all as a staring point of the texture density continuum. (No density levels of >50 were used, as this would have introduced too much glare for the video cameras if more than half of the grid had contained white dot elements).

Procedure A psychophysical method of adjustment was used for the alignment task of the VP with the designated targets on the surface of the real hemisphere image. Subjects used the Spaceball to move the VP with six degree of freedom (three translations and three rotations) until the VP appeared to them to just touch the surface of the sphere exactly at the target position (i.e., the centre of the designated circle). As a secondary aim, they were also instructed to align the VP such that it was *normal* to the surface at the designated target position. Figure 3 illustrates how task completion might have appeared from the subject's point of view. (It is important to recall that subjects viewed the images stereoscopically, which is not possible to reproduce here.)

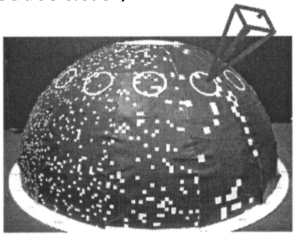

Figure 3. Example of placement task using a 3D pyramidal virtual pointer (VP), shown being aligned with a 10% density target, situated at 30° from central meridian.

Each experiment consisted of 6 randomised replications of each condition, for a total of 150 (6x5x5) judgements. The experiment, including practice, took place over a span of two days, with each session lasting approximately two hours per day. Eleven university students participated in the experiment. None of them knew about the design and the goals of the experiment. Some of them used optical correction.

Measurement

Placement Error The placement errors between the target positions estimated and the corresponding real surface positions were computed along the direction of the normal vector at the perceived target position. This corresponded to the difference between the radius of the sphere (r) and the distance from the perceived target position to the origin (d), as illustrated in Figure 4. (Note that attention was paid only to *radial error*, since differences between the actual and perceived centres of the target circles on the hemisphere surface were not of interest.) A *negative* mean error was taken to mean that the position detected was *outside* the sphere surface (r–d < 0), and a positive mean error meant that the detected position was *inside* the real target surface along the surface normal (r-d > 0).

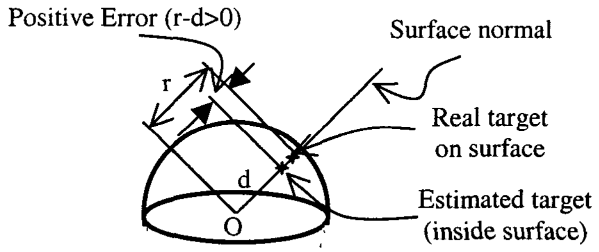

Figure 4. Measurement of placement error (Positive error shows the perceived target is *inside* the sphere along surface normal)

Angular Error There were two types of angular errors related to the estimates of surface normal at target locations: *elevation errors* and *azimuth errors*. As Figure 5 illustrates, for a target **T** on the hemisphere surface, the *actual* surface normal is the vector **ON**, which passes through **T**, and vector **TP** is the *estimated* surface normal at the target **T**. If vector **OQ** is parallel to vector **TP**, then they have the same elevation and azimuth. Thus, the elevation and azimuth errors between the estimated normal **TP** and the real surface normal **ON** can be regarded as the elevation and azimuth errors between vector **OQ** and vector **OT**. If vector **OS** and **OR** are the projections of vectors **OT** and **OQ** onto the sphere's horizon plane, then the elevation and azimuth angles for vectors **OT** and vector **OQ** are $\alpha\,(\angle TOS)$, $б\,(\angle SOZ)$, and $\alpha'\,(\angle QOR)$, $б'\,(\angle ROZ)$ respectively. Therefore, the elevation and azimuth errors between the estimated normal (vector **TP**) and the real surface normal (vector **OT**) in this experiment can be calculated as $\Delta\alpha = \alpha' - \alpha$ and $\Delta б = б' - б$. If $\Delta\alpha > 0$, the estimated normal has a larger elevation angle than the real surface normal, and vice versa. If $\Delta б > 0$, the estimated normal has a larger azimuth angle (relative to the south point on the sphere horizon plane) than the real surface normal.

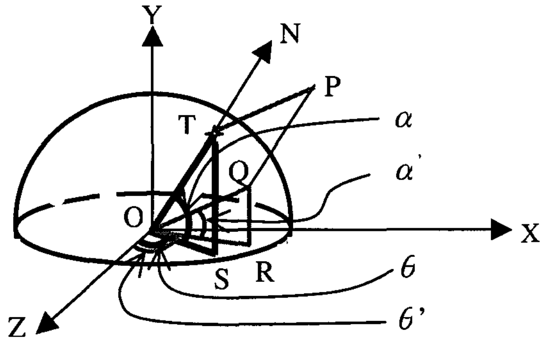

Figure 5. Definitions of elevation and azimuth errors (The difference between $\angle TOS$ and $\angle QOR$ is elevation error $\Delta\alpha$. The difference between $\angle SOZ$ and $\angle ROZ$ is the azimuth error $\Delta б$)

RESULTS AND DISCUSSION

An analysis of variance performed on the results revealed that both surface texture and target position had significant effects on alignment accuracy and target position affected estimates of surface normal, as explained below.

Texture Density Figure 6 summarises the significant effect of surface texture on alignment accuracy ($F(4,40)=41.34$; $p<0.001$). It is important to note that the positive errors found here are consistent with results from our earlier experiments, where it was found that highly textured surfaces generated smaller placement errors than sparsely textured surfaces. In those experiments, however, only two values of texture density were used: high=33% and low=11% (Hou & Milgram, 2000; 2001).

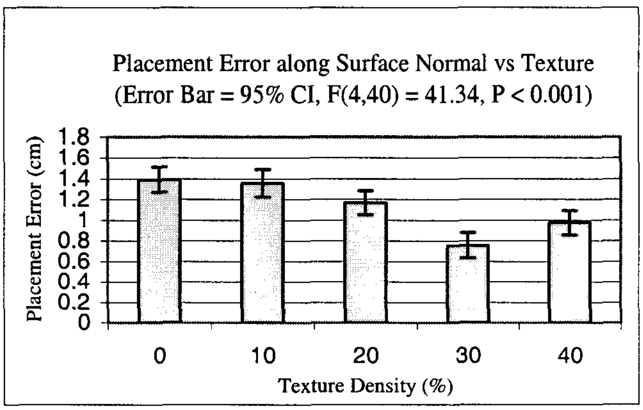

Figure 6. Effect of surface texture (Positive values indicate estimated position inside sphere surface)

Recalling that a positive error means that the perceived target was *inside* the sphere (see Figure 4), this result supports our contention that the cue conflict between the binocular disparity and occlusion was being used as an extra cue to detect the sphere surface. In other words, the fact that subjects consistently placed the cursor *behind* the target surface suggests that there was some special cue associated with the target surface itself, since, on the basis of binocular disparity matching alone, there is no reason to predict such a bias.

It should also be noted that there appears to be a limit for the higher texture density advantage, somewhere around 30%. Since the 40% texture density error appears to be on the right hand side of a U–shaped function, this suggests that there is an *optimal value* for certain textures (in this study square randomly distributed dots). In other words, when the texture density approached this optimal value, it facilitated alignment, but when the texture density exceeded this optimal value, it started to generate more alignment errors, possibly because there were too many surface features to be fused, or perhaps due to too much glare caused by the increasing proportion of white versus black elements.

Target Position

Placement Error: Figure 7 summarises the experimental effect of target position on translational alignment performance. As hypothesised, target position had a significant effect on the placement accuracy ($F(4,40)=11.90$; $p<0.001$). Consistent with our earlier experiments, performance was better when the VP was placed at a point along the longitudinal axis intersecting the centre of the surface (target #3, as illustrated in Figures 2) relative to the observer's normal straight viewing angle, as compared to placing the VP off the centre of the sphere (targets #1, #2, #4, and #5).

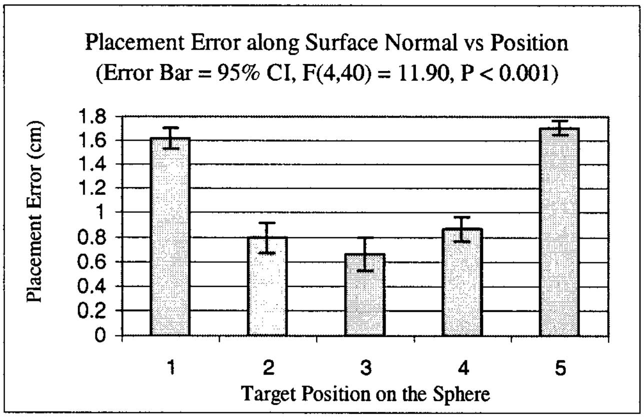

Figure 7. Effect of target position

One explanation for this result is that it was due to the relationship between the VP and the local surfaces. When the VP was aimed at the off-centre targets (especially for target #1 and #5), it was easier to focus on the VP because its major portion was outside the sphere background, thus making the fusion breakdown harder to perceive. Subjects may therefore have tended to push the VP further and deeper behind the surface until they encountered fusion difficulties, thereby generating larger placement errors in comparison with the central target results. When the VP was controlled to align with targets #2 and #4, the portion of the VP outside of the sphere background surface was less than when it was aimed at targets #1 and #5, so this contributing factor would have been smaller. Also, because targets #2 and #4 were closer to the stereo cameras, their image resolutions were higher than those of targets #1 and #5. Together, these factors may have contributed to the smaller alignment errors for targets #2 and #4. For the central target #3, the whole VP was viewed against the background of the sphere surface, and the image resolution was also the highest relative to the other four targets because it was closest to the stereo cameras, in addition to being at the convergence point of the stereo cameras. Consequently, there was minimum alignment error for the central target.

This result implies that, since the observers' viewing angle was different for the central target and off-centre targets, perception of the local surfaces at these sites was also different. This implies further that one can expect to perform better when placing a VP at a point along the centre of a surface relative to one's normal straight viewing angle (that is, looking straight at the surface) as compared to any other angle relative to the normal lateral plane. This finding is not intuitively obvious, since one might otherwise expect superior performance when one is able to watch the VP approaching a surface more from the side, rather than straight on.

Angular Error: Interestingly, the effect of target position on estimate of surface normal revealed a different perspective. Although the ANOVA for the azimuth error did not show any significant influence ($F(4,40)=0.307$; $p=0.677$), the results for the elevation error did indicate a significant impact on surface normal alignment performance. As illustrated in Figure 8, the central target generated larger elevation errors than the other targets, although it was closest to the stereo cameras, and had the highest image resolution, compared with the other targets. This position was also along the line of sight, however, and thus could easily have been blocked by the VP while approaching the surface. In other words the VP might have caused difficulty in seeing the target area clearly whenever the targets were approximately aligned with respect to placement (translational) accuracy. The fact that subjects tended to tilt the VP upwards (i.e., positive error) in order to create a better viewing angle to achieve the primary task (aligning the VP tip with the surface target), may inadvertently have introduced more angular error to the secondary task (aligning the VP with the normal direction to the hemisphere surface at the target).

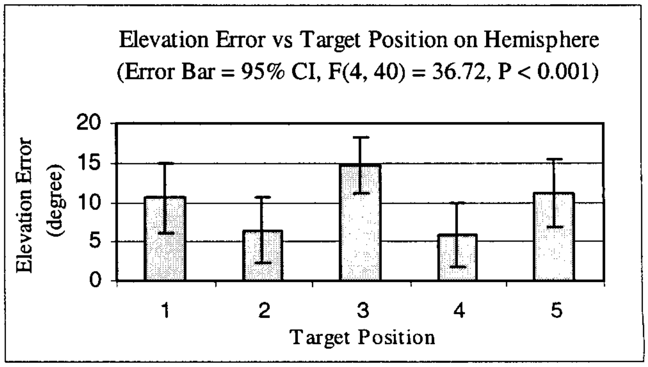

Figure 8. Effect of target position on elevation error

When the target was off-centre, the positive bias continued to be observed, but elevation errors at targets #2 and #4 were smaller than those at targets #1 and #5. Compared with the central target #3, targets #1, #2, #4 and #5 were not along the line of sight, but were located off-centre from the sphere surface (relative to the observer). From our original theorising, it had been assumed that it would be easier for subjects to control the VP when the viewing angles were off-centre. However, since the resolutions of the target #2 and #4 images were higher than those of targets #1 and #5, as they were closer to the stereo cameras, this may have generated smaller angular errors than targets #1 and #5. In other words, although somewhat speculative, the results in Figure 8 may represent the interaction between two conflicting processes: a U-shaped curve centred at target #3, due to camera resolution, and an inverted U-shaped curve centred at target #3 due to line of sight interference.

IMPLICATIONS FOR AR INTERFACE DESIGN

One specific goal for this research was to find an engineering solution for improving the accuracy of 3D measurements on real objects in stereoscopic AR environments, in light of the difficulties encountered when trying to align a virtual pointer in situations for which target points are located on smooth curved surfaces. For example, in an earlier test of the feasibility of the virtual tape measure for estimating the circumference of aneurysms in a neurosurgical application, using a stereoscopic operating microscope, it was found that precision and accuracy were quite acceptable for measurements performed on sharp, high contrast targets objects, but difficulties were encountered when trying to align the VP with the smooth, curved, featureless aneurysm models (Kim, Drake, and Milgram, 2000).

In light of the results reported here, which indicate that fusion breakdown can be used as an extra cue, over and above binocular disparity matching, for real-virtual alignment tasks, in the sense that performance can be improved if the surface being localised comprises a suitable density of visible features (in our experiment, 30% texture density was found to be optimal), it is now reasonable to consider the possibility of *artificially introducing an appropriate surface texture*, as a means to facilitate surface localisation. This concept has been tested in our laboratory, using projected random dot images to create texture patterns on curved target surfaces (Hou, 2003). The rationale behind this idea is that, in an environment in which natural textures are inadequate for accurate alignment, such as aligning a virtual robot with a rock face inside a mine (e.g., Hou & Milgram, 2001), one could simply add the required texture artificially, as required, through incident light projection, and thereby provide more depth and curvature information than is provided naturally by the untreated rock surfaces.

Besides visual texture influences on alignment in stereoscopic AR environments, the target position effect found here provides guidance on how to place a stereoscopic camera system relative to a known target area, and where best to aim at for accurate and efficient operations with respect to real object interactions. Since a central target is along line of sight of a VP, which can block the target area, it is suggested that the VP should aim at an off-centre target but retain a textured surface background, in order to exploit the beneficial effects of real-virtual object interactions more easily.

CONCLUSIONS

A detailed investigation was conducted of the effects of texture density and target position on VP alignment, both translational and rotational, relative to a hemispherical surface. The translational and rotational (angular) alignment errors relative to the surface normal were recorded and analysed. The findings were consistent with previous experimental results, but also added some new insights. The main findings and insights reported here are as follows:

1. Visible texture density of target surfaces significantly affected target placement (translational) performance, and the existence of an optimal level of texture density was found. This serves as evidence that the conflict between binocular disparity and occlusion cues can be used as an extra cue for locating real object surfaces when manipulating a virtual pointer in such stereoscopic AR displays. The detected surface location was overestimated along the surface normal direction, but only slightly.
2. Target position also played an important role in the alignment tasks. Consistent with previous findings, the central target position produced smallest placement errors. However, it also generated the largest errors in directional (angular) alignment, due to position along the line of sight.
3. In terms of 3D AR interface design, a practical solution for improving real-virtual object alignment accuracy, involving projected light patterns, has been proposed.

REFERENCES

1. Cannon, D. & Thomas, G. "Virtual tools for supervisory and collaborative control of robots". *Presence*, 6(1), 1-28, 1997.
2. Drascic, D. & Milgram, P. "Perceptual issues in augmented reality", *Proc SPIE, Vol 2653-19, Stereoscopic Displays and Virtual Reality Systems III*, Jan.-Feb. 1996.
3. Ellis S. R. & Menge B. M. "Localization of virtual objects in the near visual field". Human Factors, Vol. 40(3), 415-431.
4. Hou, M. "Effects of surface characteristics on alignment of real and graphic objects in stereoscopic augmented reality environments". *Unpublished Ph.D. dissertation*. University of Toronto, 2003.
5. Hou, M. & Milgram, P. "Effect of surface characteristics on alignment of graphic and real objects in a stereoscopic augmented reality environment" *Proc IEA200 / HFES2000 Congress*, San Diego, CA Vol. 3, 476-479, 2000.
6. Hou, M. & Milgram, P. "Investigation of surface effects on alignment of real and graphic objects in stereoscopic augmented reality environments". *Proc. HFES 45th Annual Meeting*, Minneapolis, MN, 1877-1881, 2001.
7. Kim, M. Y., Drake, J. M. and Milgram, P. "Virtual tape measure for the operating microscope: System specifications and performance evaluation". *Computer Aided Surgery 5*, 148-155, 2000.
8. Milgram, P., Drascic, D., and Grodski, J. "Enhancement of 3D video displays by means of superimposed stereographics". *Proc. HFES 35th Annual Meeting*, 1457-1461, 1991.