EFFECT OF SURFACE CHARACTERISTICS ON ALIGNMENT OF GRAPHIC AND REAL OBJECTS IN A STEREOSCOPIC AUGMENTED REALITY ENVIRONMENT

Ming Hou, Paul Milgram
Ergonomics in Teleoperation and Control (ETC) Laboratory
Department of Mechanical and Industrial Engineering, University of Toronto
{ming, milgram}@mie.utoronto.ca

A virtual pointer (VP) alignment task at designated probe points on a cylinder real object surface was carried out in a stereoscopic augmented reality environment. The object was to evaluate subjects' sensitivity to surface texture, target position on the curved surface, VP orientation relative to the surface, and binocular disparity. The main findings were: a) surface texture had a significant effect, with highly textured surfaces facilitating less error than low textured surfaces; b) target position had a significant influence, with the central position relative to the observer being better for locating surface positions than the off-centre position. Results are discussed in terms of hypothesised visual perceptual interactions.

INTRODUCTION AND MOTIVATION

How can a neurosurgeon accurately measure the dimensions of an aneurysm intraoperatively so that it can be rendered harmless with a properly sized surgical clip? How can an operator tell a robot where "there" is for execution of a "put that there" kind of instruction? To address such absolute measurement and specification problems, we have developed a "Virtual Tape Measure" (VTM), based on augmented reality through graphic overlays on stereo-video (ARGOS). To measure dimensions and distances between real objects in a 3D video scene, a *virtual stereographic pointer* is interactively manipulated and aligned with features of interest in the stereo video image. Based on prior calibration of the camera system, absolute distances/dimensions can then be computed, using the camera frame of reference. Earlier experiments have 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 (Drascic and Milgram, 1991).

Although the stereoscopic displays provide the general advantage of enhanced depth perception, it has been found that, for augmented reality (AR) displays such as ARGOS, which has been developed for measurements in unstructured environments, whenever the virtual pointer (VP) goes behind the surface of an object yet fails to disappear, perceptual conflicts frequently occur between the consistent binocular disparity information and the inconsistent occlusion information, resulting in some kind of a double image (Hou, 1999). Note that, in an unstructured – and thus unmodelled – environment, the computer generating graphic image does not generally have sufficient information to detect interactions between real and virtual objects, thereby making it difficult to adjust the graphic image to occlude portions of the VP which should properly be hidden. The double image happens 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 surface (video) and the VP (graphic). Certainly, this problem does not occur when the images are not displayed stereoscopically. Based on these results, the task of aligning a virtual pointer with real object surfaces when using 3D AR displays can be difficult. This is especially true, for example, for the kind of anatomical objects one typically encounters in surgical environments, where surfaces are rounded, shading is uneven and textural cues are ambiguous.

The general goal of our research is to determine what factors affect the ability to align virtual and real objects in 3D AR displays for making accurate measurements, and ultimately to determine whether a method can be developed for improving current VP alignment performance for arbitrarily oriented 3D curved surfaces. The specific objective of the research reported here is to study the influence of a particular set of visual characteristics of curved surfaces of real objects on the ability to align a virtual stereographic pointer with real stereo video objects.

Based on a series of exploratory studies, we propose four hypotheses about one's ability to perform such tasks:

a) it is possible to exploit the breakdown of fusion phenomenon to more easily localise points on curved surfaces of real objects which contain textures of relatively high spatial frequency;

b) the orientation of the curved 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;

c) in terms of orientation of the virtual pointer (VP), there exists a conical volume centred around an axis defined by the normal to a curved surface, within which the VP, when oriented within this volume, can be used to localise positions on the surface more accurately and with less fusion difficulty than for orientations outside the cone; and

d) the direction of binocular disparity (i.e. crossed vs uncrossed) will also influence alignment performance.

To test these hypotheses, the following variables were manipulated as independent variables: surface texture, VP orientation relative to the real target surface, angular displacement of the surface normal relative to user's viewpoint, and binocular disparity.

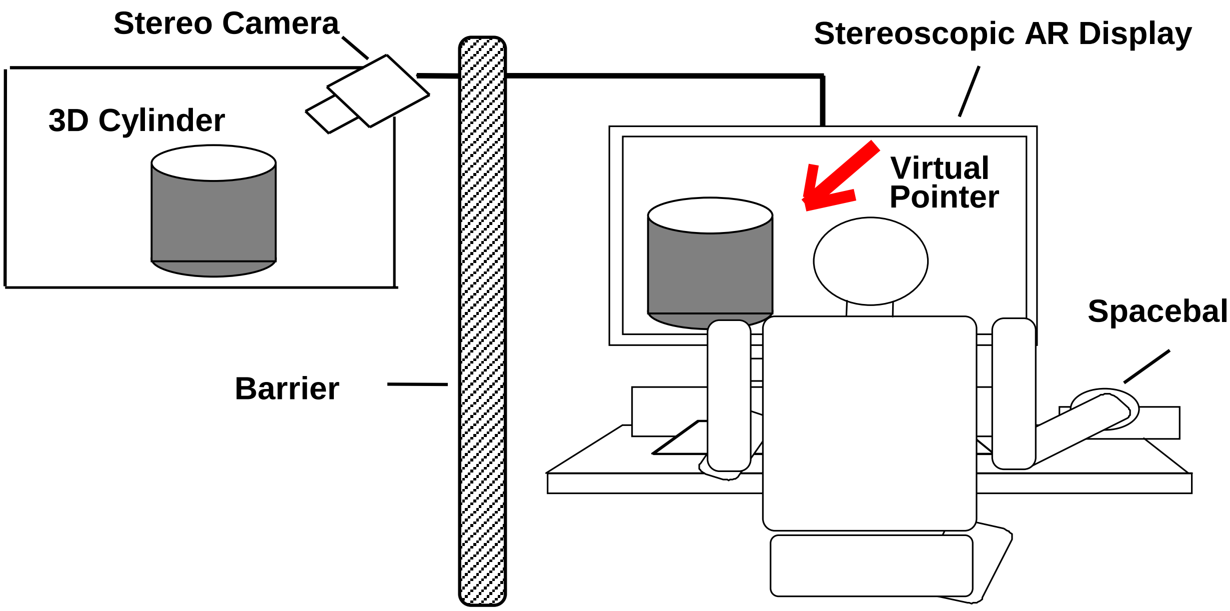

Figure 1. Experimental Set-up

METHOD

The experiment consisted of a psychophysical method of adjustment task involving the alignment of the VP with designated targets on the surface of the real object image, all of which were displayed using stereoscopic augmented reality (ARGOS). The experimental set-up is depicted in Figure 1. 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 VP orientations relative to the surface (vertical, horizontal and diagonal), 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.

Stimulus Generation and Apparatus

The target stimuli comprised a set of alternating field stereoscopic images of a 46 cm diameter cylinder, recorded by a 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 from a random dot stereogram package), but with different texture densities (see Fig.2).

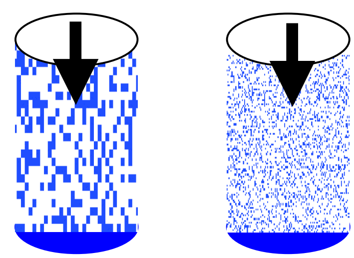

Figure 2. Low (left) and highly (right) textured cylinders

The stereo cameras were located 58 cm from the front surface of the cylinder. Three different camera convergence distances were used: 6 cm *behind* the surface of the cylinder, *at* the surface, and 6 cm *in front of* the surface. Alignment of the VP with a target at the proximal surface of the cylinder (i.e. *correct* placement) therefore corresponded to *crossed* disparity, *no* disparity and *uncrossed* disparity respectively.

The VP was a three-dimensional computer generated arrow that appeared to hover within the stereo image upon which it was superimposed. 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 illustrated 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 translation degrees of freedom.

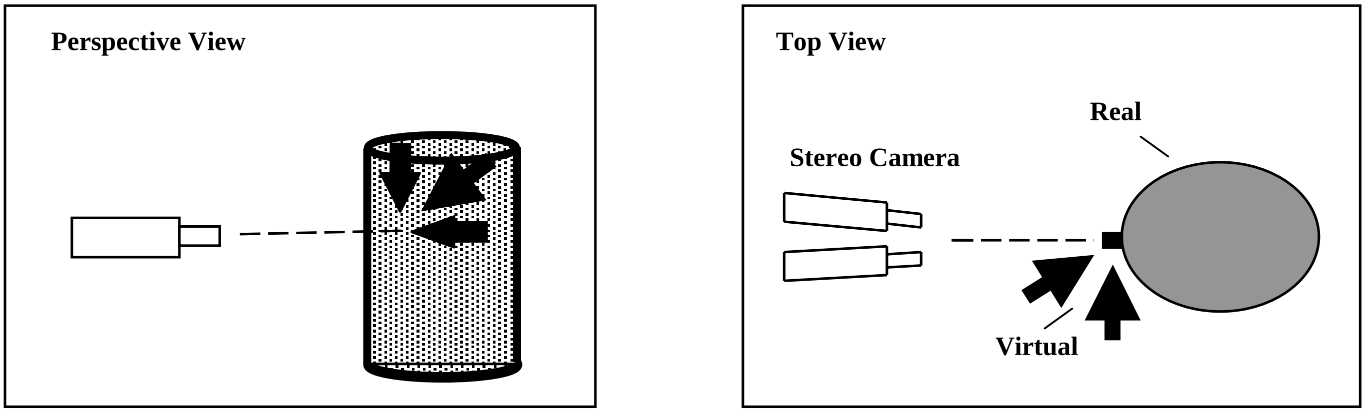

Figure 3. Different orientations (horizonal, vertical and diagonal) of virtual pointer

Procedure

The experimental task was to localise points on the cylinder surfaces by manipulating the VP, for the three orientations, two textures, three camera configurations (disparities) and two target positions. Subjects used the Spaceball to move the VP in and out along X, Y, Z axes until it appeared to them to touch the surface of the object exactly at the designated target location. 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.

Subjects

N=10 university students (6 male and 4 female) were screened using the RANDOT STEREOTESTS to participate the experiment. None of them knew about the design and the aims of the experiment. Where necessary, subjects wore appropriate optical correction.

RESULTS AND DISCUSSION

The principal results of the 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.

The present 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. 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.

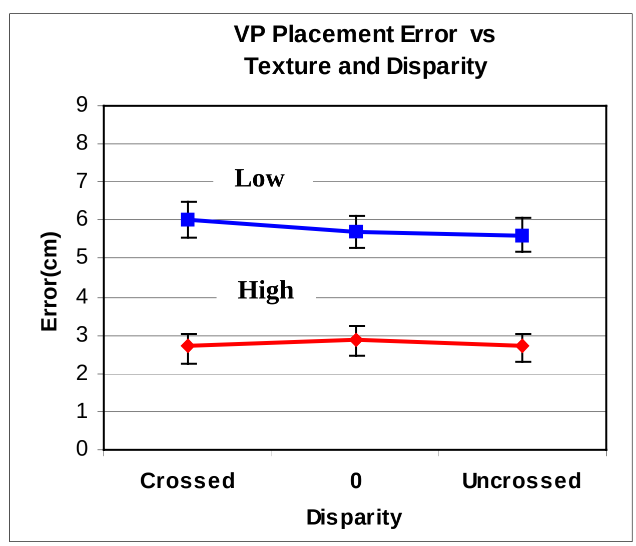

Figure 4. Effect of Surface Texture (High vs Low) and Disparity (Crossed vs 0 vs Uncrossed)

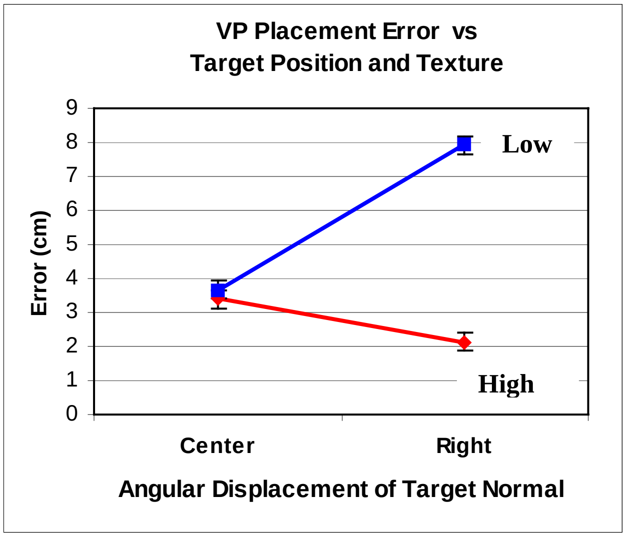

Figure 5. Effect of surface texture (High vs Low) and target position (Center vs Right)

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 to the normal straight view angle, 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 the highly textured surface error ($F(1,9)=30.22$, $p<0.001$).

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 view 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 further 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 analysis ($F(2,18)=1.015$, $p=0.38$) in this experiment.

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

The ANOVA also showed that there was a 3-way interaction, between surface texture, disparity, and the VP orientation ($F(4,36)=3.10$, $p<0.02$). This significant interaction can possibly explain partially why there were no significant effects due to either disparity or VP orientation. We are currently investigating these two factors further.

Although augment reality combining stereoscopic video and stereoscopic computer graphics has already resulted in new technical capabilities, such as virtual tape measures and virtual telerobotic control systems (Milgram et al, 1997), a number of interesting perceptual issues have arisen as well. An experiment to examine "perceptual surface effects" has indicated that surface texture and position of targets on a curved surface have significant effects on the ability to align graphic and real objects in stereoscopic AR environments. In order to facilitate the design of more efficient probing tools for AR applications, further investigations are needed, not only to elaborate on the results reported here, but also to determine the relative influences of factors such as disparity due to stereo camera configuration, VP orientation relative to the curved surface, and form (i.e. shape and volume) of the virtual pointer. At a more fundamental level, such research should also increase our understanding of human perception of surface curvature in AR environments, as well as the relative strength of binocular disparity and occlusion depth cues.

ACKNOWLEDGEMENTS

The work reported here was carried out with support from NSERC and the Institute of Robotics and Intelligent Systems (IRIS), under the project *SMART: Sensori-Motor Augmented Reality for Telerobotics.*

REFERENCES

1. Drascic, D. & Milgram, P. "Position accuracy of a virtual stereographic pointer in a real stereoscopic video world." *Proc. SPIE 1457 - Stereoscopic Displays and Applications II*, 1991.
2. Hou, M. "Perceptual localisation of surface normal." *Proceedings of ACM Conference on Human Factors in Computer Systems CHI'99* (Extended Abstracts), 1999.
3. Kim, M., Milgram, P. & Drake, JM. "Computer Assisted 3D Measurements for Micro-Surgery." *Proc. Human Factors & Ergonomics Society 41st Annual Meeting*, 1997.
4. Milgram, P., Yin, S. & Grodski, J.J. "An Augmented Reality Based Teleoperation Interface for Unstructured Environments", *ANS 7th Topical Meeting on Robotics and Remote Systems, Augusta, Georgia*, 966-973, 1997.