Video Texture Cues Enhance Stereoscopic Depth Perception In A Virtual Reality-Based, Telerobotic Interface

Geb Thomas
David Cannon
Joseph Goldberg
Department of Industrial and Management Systems Engineering
The Pennsylvania State University.

ABSTRACT

A major challenge in our telerobotic research is helping operators understand complex spatial relationships at the remote robot work site. Many researchers have reported that the technology of field-sequential displays, such as stereoscopic goggles, can be useful when depth perception is important. In applying this technology to our work, we met and explored a major limitation to field-sequential displays: the effect of target orientation.

Stereoscopic depth cues do not define the depth of large, homogeneous, horizontal objects. The variance of operator's depth estimates for the same target increases as the orientation of the target's major edge is rotated from a vertical position to a horizontal position. In some applications, this effect could eliminate the expected benefit of field-sequential displays. Fortunately, drawing an appropriate texture on the surface of the feature completely overcomes the orientation factor. Further experiments explore some critical features of the texture, including texture coarseness and orientation. A final experiment compares the benefit of etching textures directly onto the surface of the graphical model compared to superimposing a texture on a window looking into a three-dimensional model.

INTRODUCTION

The power of mapped textures to enhance depth perception became apparent during our development of the Virtual Environment, Point and Direct (VEPAD) telerobotic workcell. This cell is designed to clean nuclear waste from underground storage silos. By integrating a virtual reality model of the robot's gripper with a planar image of the video, this workcell bridges the gap between working with live video and relying on an abstracted, three-dimensional model of the remote environment. The cell's operator uses a joint monitoring glove to designate grasping positions viewed in the live video scene. A major advantage of this approach is that the operator constantly receives information from the real scene and does not need to rely on models of the environment that are hours or even years old. However, live video views are not always sufficient, since three-dimensional models of the environment are important for path planning, collision avoidance and other autonomous subsystems. Consequently, techniques integrating the video information with the model information are advantageous.

One integration approach maps a video image onto the solid model object, using the camera location and the perspective transformation to associate regions in the image with vertices of the modeled objects. This technique produces a visual effect of a solid model etched with the image from the real scene. Not only does the operator receive recent information about the remote environment within the virtual reality, but the textures enhance his or her perception of depth. This depth enhancement is predicted by theory and confirmed by experiment.

This paper briefly develops the benefits, mathematics and special considerations of stereoscopic displays before describing four experiments that investigate the effect of target orientation and texture. Experiment 1 examines the effect or target orientation on stereoscopic depth perception. Experiments 2 and 3 demonstrate how textures can overcome the limitation of target orientation, within certain limitations. Experiment 4 compares "etching" the texture onto the model surfaces with a rendering approach that maps the texture onto a planar "windshield" inside the model environment. The VEPAD cell presently employs the later approach [1][2] and in the JPL "phantom robot" telerobotic display [3], although neither of these interfaces use stereoscopic displays. Etching the figures on model surfaces enhances the operator's depth perception, whereas the windshield approach negates the benefits of the stereoscopic model display.

BACKGROUND

The human visual system constantly compares the two slightly different images of the world, called stereo image pairs, viewed by each eye. It analyzes the retinal disparity in these image pairs to estimate the relative depth of the viewed objects. Many other visual cues can also provide depth information including: interposition, linear perspective, texture gradient, shading, motion perspective, and motion parallax. Computer interfaces can provide stereoscopic cues by rapidly generating and displaying stereo-image pairs. Field-sequential displays alternately render each half of the stereo pair while

0-7803-2559-1/95 $4.00

synchronized shutter goggles expose each eye with an oscillating electronic shutter.

Stereoscopic displays, and field-sequential displays in particular, have become an important interface tool in the last decade [4]. Several excellent papers review the subject of human perception of stereoscopic information including: [5] and [6]. [7] and [8] established application guidelines for field-sequential displays.

Designers of interfaces that rely on stereoscopic information should be wary of its limitations. Stereoscopic information can be overlooked and misinterpreted. For instance, the factors affecting stereoacuity and the threshold of diplopia (double vision) are complex and highly dependent on application details. In practice, illumination level, background content, and target orientation can cause the range of perceivable depths to shift. Even if operators perceive the disparity information, they may interpret it incorrectly because of stimulus size, target depth or target movement [9].

The Mathematics of Disparity Although the limits of stereoscopic perception may vary, depending on experimental conditions, the mathematics of disparity is constant. Koenderink and van Doorn provide an elegant derivation of disparity [10]. Their development includes the creation of a polar coordinate frame located at the nodal center of an imaginary "cyclopean eye" centered between the nodal points of the left and right eyes. Using linear transformation matrices, they deduce the location and projection system of each eye when the cyclopean eye fixates on a stimulus. They then deduce the angular disparity vector, χ, in the cyclopean system, as a function of the separation of the ocular nodal points, δ, fixation position $f(\rho_0, \theta_0, \gamma_0)$ and the point of interest $p(\rho, \theta, \gamma)$, where ρ θ and γ represent distance, azimuth and elevation, respectively.

$$\bar{\chi} = \left(\begin{pmatrix} -\frac{\delta}{\rho}\cos\theta\cos\gamma - \frac{\delta}{\rho_0}\sin\gamma\sin\gamma_0\cos\gamma_0(1-\cos\theta_0) \\ +\frac{\delta}{\rho_0}\cos\gamma(\cos^2\gamma_0\cos\theta_0 + \sin^2\gamma_0) \end{pmatrix} \bar{e}_\theta + \begin{pmatrix} \frac{\delta}{\rho}\sin\gamma - \frac{\delta}{\rho_0}\cos\theta\cos\gamma\sin\gamma_0\cos\gamma_0(1-\cos\theta_0) \\ -\frac{\delta}{\rho_0}\cos\theta\sin\gamma(\cos^2\gamma_0\cos\theta_0 + \sin^2\gamma_0) \\ -\frac{\delta}{\rho_0}\sin\theta\frac{1-\cos\theta_0}{\sin\theta_0}\sin\gamma_0 \end{pmatrix} \bar{e}_\gamma \right)$$

Superscript bars indicate vector quantities. The vectors e_θ and e_γ are unit vectors in the azimuth and elevation directions, respectively. When θ_0 and γ_0 are zero, the fixation direction is directly forward.

Orientation in frontal plane

Ebenholtz and Walchli found that stereoscopic thresholds changed by a factor of between 5 and 8 in depth estimations of thin horizontal lines rotated in increments between vertical and horizontal orientations [10]. The thresholds ranged from 10 to 140 arcsecs. The shape of the threshold function follows the form

$$threshold = a - b\cos\theta$$

where θ is the angular tilt of the lines from the vertical in the frontal plane.

Ogle [12] and Blake, Camisa and Antonetti [13] also found the cosine function.

The target orientation factor is critical to the success of stereoscopic displays because it causes the depth of large horizontal edges to be ambiguous. In a stereoscopic telerobotic workcell application, for example, robots might regularly collide with the front edge of a worktable because the operator consistently misperceives the depth of its long horizontal edges. As the experiments described here show, If the surface of the table is textured, the operator can accurately perceive its depth and successfully redirect the robot.

Ghosting

One artifact of the field-sequential display technology is the cross-talk between images. A faint ghost of the image intended for the other eye can sometimes be seen with the intended image. This gives the effect of a halo, or blurriness around the edges of the displayed objects. The ghosting, which may influence the development of asthenopia (eye strain), can be measured by the extinction ratio: the ratio of luminance of the correct eye image to the luminance of the ghost image.

Ghosting in field-sequential displays depends on three factors: light transmitted by liquid crystal shutter when closed, phosphor persistence, and vertical screen position [7]. For a particular set of goggles, extinction ratios range from (61.3:1) at the monitor top to (41.1:1) at the bottom for red stimuli and from (17.0:1) to about (11.0:1) top to bottom for white stimuli.

To control ghosting, the color green, whose phosphor has a long delay time, was not used. Although relying on a particular color may be a concern because of chromostereopsis (difference in apparent depth of coplanar colored stimuli), this is probably not a great concern in long duration stimuli [14][15]. Yeh and Silverstein compared red and white stimuli. They found that stimulus color has little effect on relative distance judgment, but that white (which uses the green phosphor) resulted in poorer display quality, less depth sensation, visible ghost images, moderate eye strain, and moderate headaches [7].

METHODS

In each experiment the subject saw the front view of the situation illustrated in Figure 1.

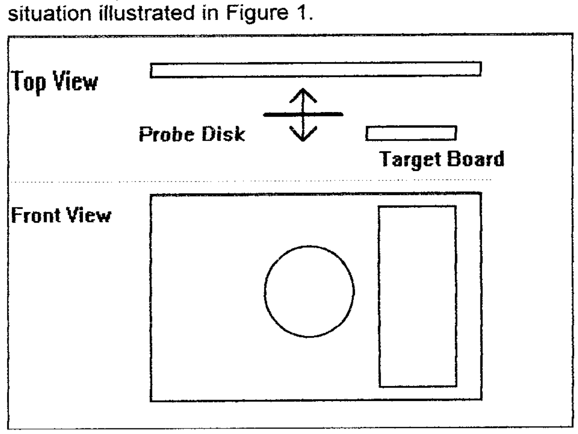

Figure 1: Model of the graphical scene.

Observers. The subjects were male members of a university community and ranged in age from 22 to

36. All participated in an eye exam using a Bausch and Lomb Orthorator. Each subject demonstrated the minimum standards listed in Table 1. Except for the measure of color vision [7] used the same standards.

Table 1: Eye Screening Criteria

Criteria	Acceptable Range
Visual acuity	20/20 Snellen equivalent or better
Near and far vertical phoria	less than or equal to 0.5 prism diopters of deviation
Near lateral phoria	-6.0 to 0.0 prism diopters
Far lateral phoria	-1.66 to 2.33 prism diopters
Stereopsis	27 arcsec or less
Color vision	3 of 4 correct in Orthorator test

Apparatus The geometry of the viewing conditions and stimulus objects are critical elements of this experiment, because they determine the size of the stereoscopic disparity.

A covered tunnel positions the shutter goggles 73.5 cm away from the Silicon Graphics HL7965Kw-SG monitor. Thus, the subject's eyes were 75 cm away from this 34.3 by 27.4 cm monitor. An 18 cm diameter circular aperture restricted the subject's view of the monitor to a central, circular region 15.2 visual degrees in diameter.

Figure 2 presents two untextured sample images, one of which shows the visual angles of each object. The outer circle is the edge of the masking aperture. The central circle with a diameter of 6 degrees represents the probe disk. The straight line represents the edge of the target board. The closest point of this line was always 4 degrees from the center of the monitor, and 1 degree away from the closest edge of the probe.

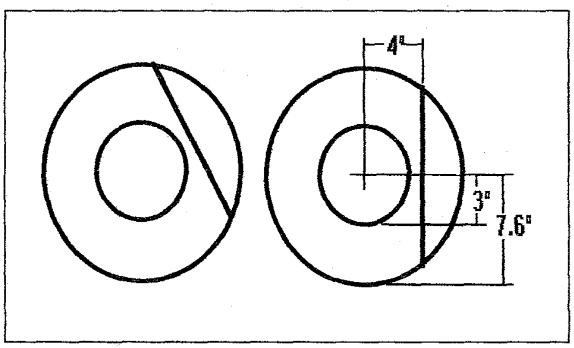

Figure 2: Two sample, untextured experimental conditions. Left, the step edge at the 45 degree position. Right dimensioned image with step edge at vertical position.

If both the right and left view of the image were overlapped, and the probe and edge were both in front of the plane of the monitor, the scene might appear like that in Figure 3. The left eye would see the probe and edge in the positions indicated by the dotted lines. The right eye would see them in the positions indicated by the solid lines. Both eyes see the masking screen in approximately the same position (when the eyes are fixating on the screen surface).

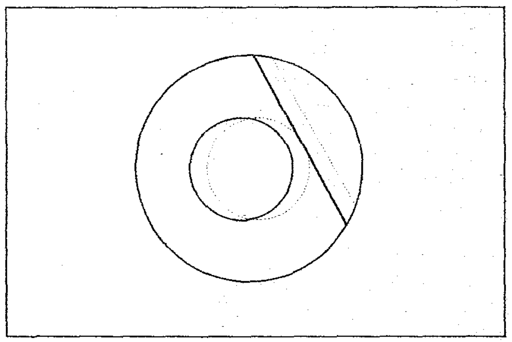

Figure 3: Stereo image pair overlaid. If objects were in front of monitor, left eye would view dotted edges. Left eye would view solid edges. If the objects were behind the plane of the monitor, the eye assignments would be reversed.

The precise location of these two, half images depends on the imaging assumptions. For all the experiments described here, the imaging geometry mimics real viewing conditions.

A coordinate system established at the center of the monitor screen extends the x-axis to the right, the y-axis upward and the z-axis toward the viewer. With the shutter glasses, the subject's left eye views the origin of the coordinate system from a position located at (0, -3.25, 75). The subject's right eye viewed the origin from a position located at (0, 3.25, 75). Thus, the model assumes that the subject has an interocular separation of 6.5 cm and fixates on the center of the screen. Note that the subject could choose to change his fixation position from the center of the screen to either a different depth or a different angular position. In either case, the projection model is not precisely valid, but is approximately accurate and conforms with standard graphic programming practices.

The viewing perspective depended not only on the viewing position, but also on the angular width and the aspect ratio, which were 0.3325 radians and 1.446, respectively.

Subjects sat and adjusted their positions until the shutter goggles pressed against the locating fixtures at one end of the viewing tunnel. The tunnel prevented ambient light or stray reflections from confusing the information on the surface of the monitor.

The probe disk, target board and background had a luminance of 1.00, 0.49 and 0.36 cd/m^2, respectively. The shutter goggles' extinction ratio for the probe was 30:1

The background was a random texture consisting of an intermediate purple with 1/2% of its pixels at full intensity and 1/2% of its pixels with 0 intensity (black).

The experiment took place in a computer laboratory with normal, office-like activity levels.

Procedure

Before beginning, each subject read the experimental instructions and became acquainted with the experimental apparatus. The experimenter answered any questions about the experiment. Next the subjects begin a training session of 8 trials. Every condition in the experiment was represented in the training session. After the training session, the subjects performed the experiment, which consisted of 4 repetitions of a fully crossed factorial experimental design. Subjects could take a break at any time, but were encouraged to continue directly through the 15 minute experiment with consistent accuracy.

In each experiment the subject moved the probe disk forward or backwards until it appeared to lie in the same plane as the target board, which was partially revealed by the viewing aperture.

The experimental program initially places the probe within 1 cm of the board in depth, according to a uniform random variable. Subjects manipulated the disk position in increments of 0.07 cm by pressing the up and down arrow keys to move it forward or backward, respectively. The right and left arrow keys provided coarse position control by moving the target in increments of 0.2 cm. The only cue to the disk's position was disparity; size and color were uniform at all positions. Depth of the stimulus plane was also cued solely by disparity information.

After manipulating the probe, the subjects indicated completion by pressing the space bar and continuing to the next trial.

EXPERIMENT 1: PLAIN ORIENTATION

The first experiment explored the effect of target orientation on depth perception. The target board was a homogeneous plane 5 cm in front or 5 cm behind the plane of the monitor screen. The plane appeared at one or four orientations: vertical (0 degrees), 48 degrees, 71 degrees, and horizontal (90 degrees).

Four subjects participated in this experiment. Each completed one training session of 8 trials and 4 repetitions of the fully crossed design.

Results. Figure 4 presents the results from this experiment. This figure shows that when the target edge was nearly vertical (0 degrees), the subject depth perception was nearly veridical -- the estimates are close to the actual depth of the target. When the target edge becomes close to horizontal (90 degrees), subject estimates are inaccurate and have a large variance. Many data points lie near the -10 cm region, the location of the background plane.

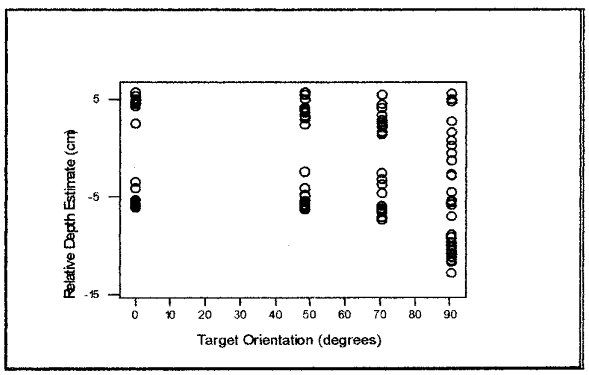

Figure 4: Results from Experiment 1.

Stereoacuity offers one possible explanation for the increasing variance. When the subject's stereoscopic acuity decreases according to the cosine relationship, his ability to estimate the target's depth diminishes. To test this hypothesis, we regressed the cosine of the orientation angle against the standard deviations of the four depth estimates made by each subject at each condition. The cosine relationship was significant ($p < .001$), but accounted for only 29% of the sum square error.

The difference between actual and estimated depth corresponds to a disparity error. An ANOVA of the disparity error as a function of depth, cosine of the target orientation, and their cross product indicated that the cosine and cross product terms were significant at the $p < .01$ level. Figure 5 presents the main effect plot. It indicates that when the target board was horizontal, Subjects tended to estimate the board as significantly farther away than in the other conditions.

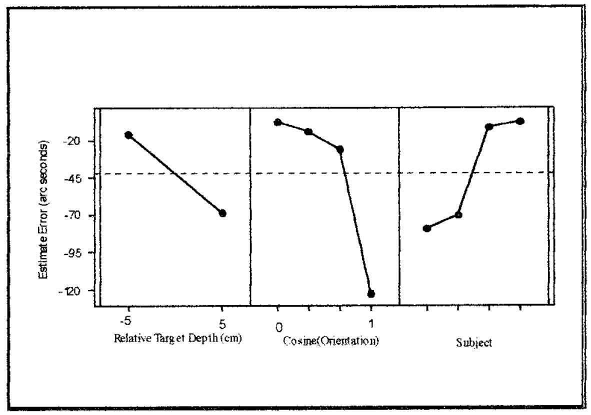

Figure 5: Main effects for disparity error in Experiment 1.

The interaction plot in Figure 6 indicates that when the orientation is horizontal and the board is 5 cm in front of the monitor, subjects tended to misperceive the stereoscopic disparity by an average of approximately -200 arc seconds.

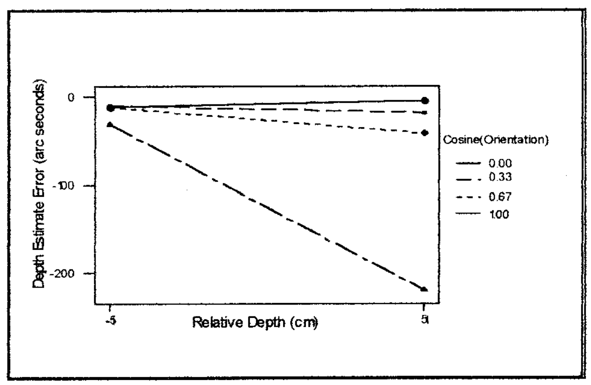

Figure 6: Interaction of depth and orientation for disparity error in Experiment 1.

EXPERIMENT 2: THE BENEFIT OF TEXTURES

The second experiment explored the effect of mapping textures to the surface of the target board. It investigated four textures of different coarseness. One was a plain, gray, homogeneous texture with the intensity of every pixel equal to 127 on a scale of 0 to 255. The gray level of each of the other three textures varied according to a Gaussian random variable with mean zero. The variance of the Gaussian distribution was either 9, 16 or 25. Figure 7 presents examples of the three random textures.

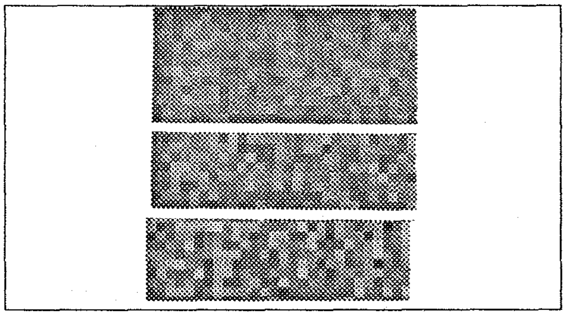

Figure 7: Textures used in Experiment 2. From top to bottom the textures have variances of 9, 16, and 25.

Each experimental block consisted of all four textures crossed with two target depths: 5 cm in front or behind the monitor surface. After an initial training block, each of the four subjects attempted four blocks.

Results. Figure 8 presents the depth estimates for Experiment 2 as a function of the standard deviation of the Gaussian generating function for the texture. This figure indicates that subjects accurately estimated the depth of the target board with the coarse textures, but were unable to estimate its depth without textures.

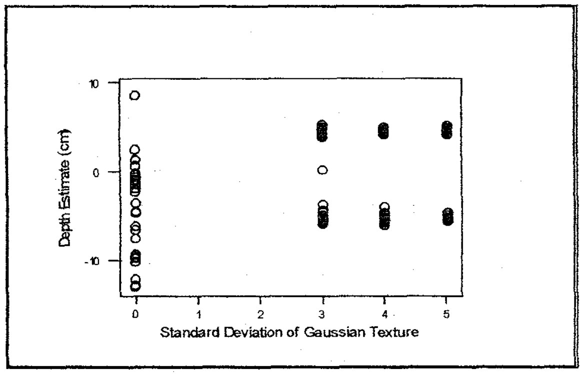

Figure 8: Depth estimates from Experiment 2.

An analysis of variance on the disparity error of the estimates as a function of depth, standard deviation of the texture pattern and their cross product indicated that all three factors were significant at the $p < .001$ level. The main effects and interactions are similar to those in Experiment 1. In a manner similar to Experiment 1, subjects' estimation of the disparity was incorrect by an average of -150 arc seconds when the target board was horizontal and had no texture.

EXPERIMENT 3: TEXTURE ORIENTATION

In Experiment 3, a symmetric, sinusoidal texture varied the color intensity of the horizontal target board. The orientation of the texture varied among vertical (0 degrees), 48 degrees, 71 degrees, and horizontal (90 degrees). Figure 9 displays these textures. Four subjects participated in a training block and 4 experimental blocks consisting of all four textures crossed with two depths (5 cm in front and behind the monitor) with the target board in the horizontal position.

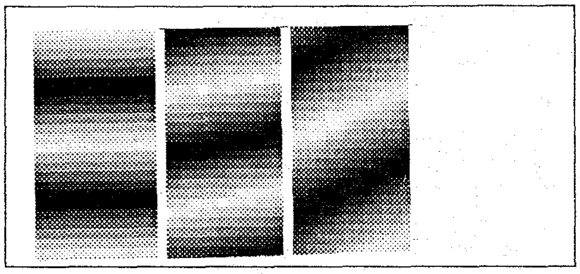

Figure 9: Textures used in Experiment 3.

Results. Figure 10 presents the depth estimates. There was considerably more variance in responses with these textures than in Experiment 2. Generally, the estimates become less precise as the textures become more horizontal. An ANOVA on the disparity error as a function of depth, texture orientation and their interaction revealed that depth was the only significant term at the ($p < .05$) level. This is probably a result of unexpectedly large variance in the vertical texture condition.

Figure 10 also indicates that when the texture bars are horizontal, subjects tended to estimate the depth of the target board to be at the depth of the monitor screen, rather than the depth of the background, as in Experiments 1 and 2.

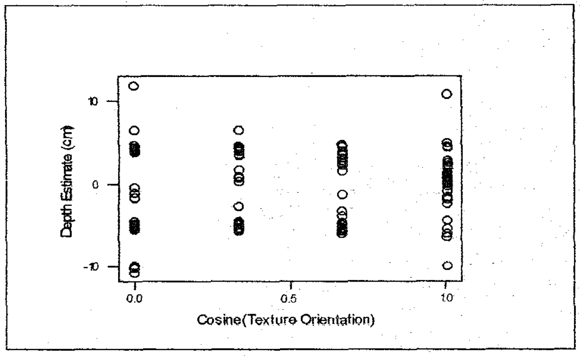

Figure 10: Depth estimates for Experiment 3.

EXPERIMENT 4: TEXTURE MAPPING STRATEGY

The final experiment compared the etched and windshield texture mapping strategies. The etched approach maps the texture directly onto the surface of the figure, in which case the disparity of the texture matched the disparity of the model. The windshield approach maps the texture to a plane at the depth of the screen, in which case the texture has effectively zero disparity. The experiment used the two methods on horizontal and vertical target blocks graphically located 5 cm in front and behind the monitor surface. Each experimental block consisted of the 8 fully crossed conditions. Four subjects completed a training block and four experimental blocks.

Results. Figure 11 presents the depth estimations for Experiment 4 as a function of the texture mapping method. Subjects were generally able to perceive the depth of etched target boards. However, with the windshield approach, subjects consistently estimated the target board at the depth of the windshield texture, rather than the depth of the target boards.

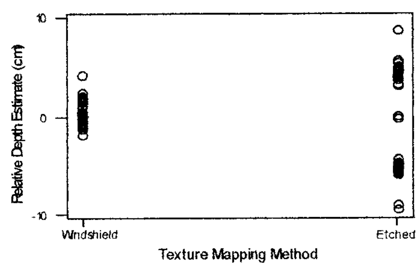

Figure 11: Depth estimation as a function of texture mapping strategy for Experiment 4.

An ANOVA on the disparity error for each estimation as a function of: orientation, depth, mapping method and their mutual interactions, indicates that depth, mapping method and their interaction are significant at the $p < 0.001$ level.

Figure 12 illustrates the interaction term. Subject disparity error is consistent for the etched figure method. With the windshield method, disparity error is consistently large for distant objects and consistently small for close objects, demonstrating the subjects' tendency to perceive the board at the depth of the texture.

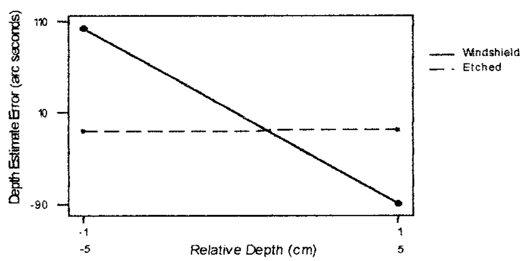

Figure 12: Interaction of depth and texture mapping method.

DISCUSSION

Most research into the benefit of stereoscopic displays has tested objects with clear vertical edges. In a practical telerobotic application with interactive viewing positions, perfectly vertical edges are the exception rather than the rule. Researchers must test the power of stereoscopic presentation at intermediate orientations. Consider a large, homogenous, horizontal edge extending beyond the viewport in both directions. Since the left and right images overlap, the horizontal disparity, and therefore the relative depth, is completely ambiguous.

Stereoacuity and perception accuracy may be significantly diminished, if not eliminated, for edges presented at near-horizontal orientations.

The experiments reported here indicate that displaying textured images eliminates the bias and therefore improves the quality, reliability and safety of stereoscopic displays. Several characteristics, including texture coarseness and texture orientation can affect the influence of texture on depth perception. Furthermore, when "good" texture information and simple edge information are in conflict, the texture information can dominate the perception of depth.

REFERENCES

[1] D. Cannon, G. Thomas, C. Wang and T. Kesavadas, "A Virtual Reality Based Point-and-Direct Robotic System With Instrumented Glove", *Int. J. Industrial Engineering*, Vol. 1, No. 2, 1994, pp. 137-146.

[2] G. Thomas, *A Gesture-Driven Interface For Supervisory Control of a Robot*, 1994, Masters Thesis, The Pennsylvania State University.

[3] A.K. Bejczy, W.S. Kim, and S.C. Venema, "The Phantom Robot: Predictive Displays for Teleoperation With Time Delay", *Proc. 1990 IEEE Int. Conf. on Robotics and Automation*, Vol. 1, pp. 546-551.

[4] C.D. Wickens, S. Todd, and K. Seidler, *Three-Dimensional Displays: Perception, Implementation, Applications* (Tech. Report CSERIAC SOAR-89-01). Wright-Patterson Air Force Base, OH, 1989.

[5] A. Arditi, "Binocular Vision", in K.R. Boff, L. Kaufman and J.P. Thomas (Eds.), *Handbook of Perception and Human Performance*, 1986, New York: Wiley.

[6] R. Patterson and W. Martin, "Human Stereopsis", *Human Factors*, Vol. 34, No. 6, 1992, pp. 669-692.

[7] Y.Y. Yeh and L.D. Silverstein, "Limits of Fusion and Depth Judgment in Stereoscopic Displays", *Human* Factors, Vol. 32, No. 1, 1990, pp. 45-60.

[8] R.J. Beaton, "Displaying Information In Depth", *Society for Information Display Digest of Technical Papers*, Vol. 21, 1990, pp. 355-358.

[9] R. Patterson, L. Moe and T. Hewitt, "Factors That Affect Depth Perception In Stereoscopic Displays", *Human Factors*, Vol. 34, No. 6, 1992, pp. 655-667.

[10] J.J. Koenderink and A.J. van Doorn, "Geometry of Binocular Vision and a Model For Stereopsis" *Biological Cybernetics*, Vol. 21, 1976, pp. 29-35.

[11] S.M. Ebenholtz and R.M. Walchli, "Stereoscopic Thresholds as a Function of Head and Object-Orientation", *Vision Research*", Vol. 5, 1965, pp. 455-461.

[12] K.N. Ogle, "Stereopsis and Vertical Disparity", *Archives of Ophthalmology*, Vol. 53, 1955, pp. 493-504.

[13] R. Blake, J. Camisa and O. Antonetti, "Binocular Depth Discrimination Depends on Orientation", *Perception &Psychophysics*, Vol. 20, 1976, pp. 113-118.

[14] D.A. Owens and A.W. Leibowitz, "Chromostereopsis With Small Pupils", *Journal of the Optical Society of America*, Vol. 65, 1975; pp. 358-359.

[15] J.J. Vos, "The Color Stereoscopic Effect", *Vision Research*, Vol. 6, 1966, pp: 105-107.