On the Utility of Stereoscopic Displays for Simulation Training

Charles J. Lloyd
Visual Performance LLC
Charles.Lloyd@VisualPerformance.us

ABSTRACT

In 1990 Tidwell declared "Stereopsis Takes Off in Flight Simulation" and he and his contemporaries asserted stereoscopic displays were "required" for a range of training tasks including air refueling, formation flight, and low level missions. But, some 22 years later, we find stereoscopic displays in a very small fraction of simulation trainers in current use. A survey of suppliers revealed experienced display systems engineers do not believe the stereopsis cue provides benefit for working distances required in the air refueling task. This belief is consistent with the conclusions published in several papers in the "virtual reality" and "head mounted display" literatures.

While developing requirements for air refueling trainers, we noticed an apparent correlation between the recommended working distance of stereopsis and the use of electronic displays. The results of many "vision science" evaluations set the threshold for depth discrimination at about 5 arcsec, corresponding with +/- 6 inches at the working distance of the KC-10 boom operator. In none of these evaluations were spatially sampled electronic displays used. Our literature review revealed those authors who concluded stereopsis is not useful at longer distances used low resolution, spatially sampled displays and made no mention of antialiasing.

This paper describes an evaluation of the effects of spatial resolution and antialiasing on stereoscopic disparity thresholds and ratings of visual comfort. The evaluation revealed eye limited stereoscopic disparity thresholds (5 to 10 arcsec) are attainable on electronic displays with a pixel pitch as coarse as 2 arcmin, but only if sufficient antialiasing is applied. This paper provides a quantitative model of the design trade space for these practical design variables. Our results imply the utility of the stereopsis cue in simulation training applications has been substantially undervalued due to inattention to the antialiasing function.

ABOUT THE AUTHOR

Dr. Charles J. Lloyd is President of Visual Performance, LLC and has 26 years of experience in display systems and applied vision research at such organizations as Honeywell's Advanced Displays Group, The Lighting Research Center, BARCO Projection Systems, FlightSafety International, and the Air Force Research Laboratory. Charles has published/presented 70 papers in this arena. His recent papers are available at http://www.charleslloyd.us.com.

On the Utility of Stereoscopic Displays for Simulation Training

Charles J. Lloyd
Visual Performance LLC
Charles.Lloyd@VisualPerformance.us

INTRODUCTION

The Promise of Stereoscopic Displays

In a 1990 paper titled "Stereopsis takes off in flight simulation" Tidwell described the state of the art in training display systems and concluded "Stereopsis is a practical reality today in flight simulation" (Tidwell, 1990). He went on to say "in the quest for ever-greater realism in flight simulation, stereopsis is emerging as a significant visual cue, and he predicted the advent of high quality stereoscopic simulators would provide more effective training. In making his case he cited studies done at NASA Langley (Parrish & Busquets, 1990) that led to the conclusion "The applications of stereo 3-D to pictorial flight displays within the program have repeatedly demonstrated increases in pilot situational awareness and task performance improvements." Tidwell also cites evaluations conducted at CAE in which the authors conclude "Stereo imagery had a stronger effect on performance in formation flight than we expected. In tasks involving close approach to other objects (i.e., formation, air to air refueling, landing, and perhaps close air combat) it may be of significant value" (Kruk & Runnings, 1989).

It would seem the case for the use of stereoscopic displays for simulation training had been made by 1990. But, some 22 years later, we find stereoscopic displays in only a very small fraction of simulation trainers in current use. This paper examines an important design challenge that may be a primary reason stereoscopic displays have not gotten off the ground for simulation training applications. In this paper we describe an evaluation of the effects of antialiasing and spatial resolution and show that these variables are strong determinants of stereoscopic effectiveness and visual comfort. While few would argue that increasing the display resolution would improve stereoscopic performance, we have not yet found any papers that quantify the effect of the antialiasing variable on stereopsis.

Conventional Wisdom

In the process of preparing display system requirements for the next generation KC-10 boom operator trainer (BOT) upgrade, we took a look at the requirements and specifications for the recently awarded visual system upgrade for the KC-135 boom operator weapons systems trainer (BOWST). Discussions with suppliers and experienced users revealed that the use of a stereoscopic display system had not been considered for the program, it was believed that the stereopsis cue would provide no benefit at the 20-ish meter working distance of the boom operator. Further discussion revealed that this assertion was based on display design experiences and rules of thumb commonly used in the broader virtual environment industry.

Expectations from the Vision Science Literature

Surprised by the assertion the stereopsis cue would provide little benefit for the air refueling task; we conducted a literature review to establish a defensible estimate of the precision with which people can discriminate depth from binocular disparity at the working distance of the boom operator. The results of this review produced a number of papers that indicate that the majority of observers with normal stereoscopic vision can discriminate between 3 and 10 arcsec of disparity under photopic levels of illumination, for objects with a spatial extent of at least a half degree, and for objects that are separated by no more than about one degree (Howard, 1918; McKee, 1983; Amigo, 1963; Palmer, 1999; Teichner 1956; and Allison, 2009). Using the 8 arcsec threshold described by Howard as representative of the "normal" observer, we expect boom operators to be able to discriminate distance differences of 0.24 m (9 inches) at 20 m using binocular disparity. Thus, we expected the stereopsis cue to be quite useful to boom operators in the refueling task.

To validate our expectations, we conducted an evaluation of the effects of stereopsis, collimation, and head tracking on the precision with which an experienced boom operator could estimate the distance between the nozzle on the refueling boom and the receptacle on approaching receiver aircraft (Lloyd & Nigus, 2011). As expected, the results of this evaluation showed a large and highly reliable improvement in distance estimation performance with the use of the stereopsis cue. Eight of the eight boom operators who participated in the testing indicated they preferred the stereoscopic display over the non-stereoscopic. The standard deviation of the distance estimates for the stereoscopic (and collimated) display conditions was 0.61 ft which corresponds with a disparity threshold of 6 arcsec at 20 m. This precision is entirely consistent with the disparity thresholds of 3 to 10 arcsec reported in the vision science literature.

Experience with Practical Stereoscopic Displays

While the research results from the vision science literature make us confident boom operators can use the stereopsis cue to their advantage in the aircraft, the results of our evaluation with the ground based simulation displays are clearly at odds with the findings of many others. For example, two separate papers published in 1989 report a stereoscopic threshold of about 2.3 arcmin (140 arcsec) for evaluations using electronic displays (Tidwell, 1989, Yeh & Silverstein, 1989). This estimate of the disparity threshold has been accepted as reasonable by a number of researchers since (i.e., Pfautz 2001). Note that this threshold estimate is 17 times higher than the threshold estimate we derived from the vision science literature. It would appear these authors assumed that the *practical* stereoscopic threshold attainable with electronic displays far exceeds the threshold that can be measured in the laboratory with the use of real objects. We suspect the assertion stereopsis provides little benefit at the boom operator working distance is based more on experiences with electronic displays than it is on vision science.

The idea that practical electronic displays are incapable of producing stereopsis cues comparable with the real world has been implicit in the work of many researchers. For example, in a commonly cited paper summarizing the relative utility of potential depth cues for television displays, Nagata (1991) states “binocular parallax is most effective at distance of less than 1 m.” and provides a figure indicating the relativity sensitivity provided by the binocular parallax (disparity) cue to drop below the motion parallax, texture, and brightness cues at a distance of 10 m. In a similar paper Cutting & Viston (1995) show the relative utility of the stereopsis cue to falling below the motion parallax, relative size, and visual field height cues at a distance of 10 m.

In a report describing the effects of stereoscopic display on task performance, Singer (1995) cites a 1993 unpublished Army report by Rinalducci that indicates "the stereoscopic disparity cue is functional only for a rather limited range, at a maximum of ten to twenty meters from the observer." This position is consistent with that of Gregory (1965, p. 16) who concluded “For near objects the different views of the two eyes are used to compute depth, but the baseline between the eyes is too small for distances beyond 50 ft or so, when we are effectively one eyed.” The results of Singer’s own evaluation led him to conclude the stereoscopic cue provides benefit only at distances of less than about 10 ft and “should not be used unless required” because it can be so uncomfortable.

Within the vision science community there is a clearly stated belief that electronic displays are unsuitable for the study of depth perception because of the undesirable artifacts they produce (Durgin, et al 1995; Frisby et al 1996; Proffit and Kaiser 1986). These authors have recommended the use of “real world objects” rather that computer displays as they are free of artifacts such as “sub-optimal spatial or temporal sampling…” (Bradshaw, 2000).

The overall conclusions we must draw from the literature describing practical display systems is that we cannot expect to achieve real world levels of stereoscopic disparity using practical electronic displays. The results of this review provide strong support for the position of the BOWST acquisition team and suppliers who did not consider the use of stereoscopic display for this training device. And these results indicate we cannot expect the use of stereoscopic display to provide a performance benefit for boom operators using the next generation BOT.

But… the results of our previous evaluation indicated the use of the stereopsis cue produced a large and highly statistically reliable improvement in ability of experienced boom operators to estimate the distance between the nozzle and receptacle (Lloyd & Nigus 2011). It would appear the results of our previous evaluation are “unusual.” Thus, we designed the current evaluation to replicate this finding and to address the reasons we were able to successfully demonstrate the utility of stereopsis cue where others did not.

Expectations and Objectives

Prior to conducting our previous evaluation we realized a display system that produced “eye limited” stereopsis cues would have to be capable of drawing image features such as lines and polygon edges with great geometric precision. A typical simulation training display system of today might have a pixel pitch of 130 arcsec (2.2 arcmin). To produce disparity cues as fine as 3 to 10 arcsec, the display system would have to reliably position image features with an error no greater than a few percent of the pixel spacing. We therefore surmised that to be successful, we would need to use a fine pixel pitch and a very good antialiasing function. For that evaluation we used a pixel pitch of 1.5 arcmin and the "adaptive super sampling" method available on the Vital X image generator by FlightSafety. In the previous paper we held the pixel pitch and level of antialiasing constant as these design variables were not the subject of that evaluation. In the present paper we exercise these variables over a significant range so that we can quantify their effects on stereoscopic display performance.

The evaluation described in this paper was designed to accomplish the following four objectives:

- Replicate the finding that stereoscopic disparity thresholds comparable with those reported in vision science literature (i.e., 3 to 10 arcsec) can be attained using practical electronic display systems
- Validate the expectation the disparity threshold increases with increasing display pitch
- Validate the expectation the disparity threshold increases with decreasing antialiasing
- Collect sufficient data and construct a mathematical model that describes the trade space for the pixel pitch and antialiasing design variables

METHOD

Equipment

This evaluation was conducted using the stereoscopic display shown in Figure 1. With this system, observers looked through a binocular optical system containing two mirrors and a lens for each eye. Accommodation distance of the display system was set using +0.50 diopter ophthalmic lenses (sphere) with anti reflection coatings. The display distance was set to 1.82 m (0.55 diopter), thus, net result of these two settings produced a system accommodation distance of 20.0 m (0.05 diopter).

A pair of Dell (Model 2007FPb) LCD monitors with a native resolution of 1600 x 1200 pixels were used to present the images. At the selected viewing distance the native angular pixel pitch of the monitors was 0.48 arcmin. The binocular FOV of the display system was 12.8 x 9.6 deg. Apertures were placed within the system such that the observer could see only the active display surfaces surrounded by black. Since two completely independent displays and optical systems were used, the leakage (crosstalk) between the right and left eye channels was zero.

The use of a long display-observer distance and low power lenses produced an optical system with very large exit pupils and eye relief. With this system there was no evidence of geometric distortion, chromatic aberration, or luminance roll off as a function of eye movements across the exit apertures. Both the accommodation and vergence distances were independent of the inter-pupillary distance (IPD) of the observer. The optical system was suitable for observers with IPDs ranging from 60 and 75 mm.

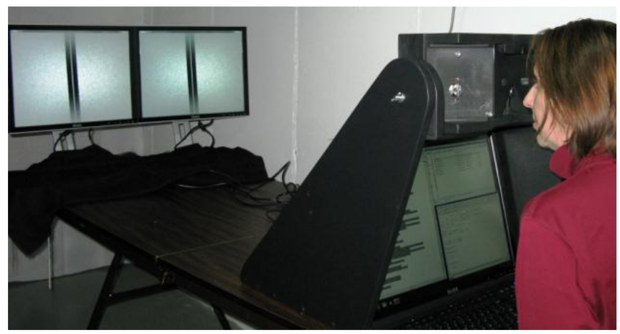

Figure 1. Photograph of the stereoscopic display system used in the evaluation. Observers viewed a pair of LCD monitors through an optical system that set the accommodation and vergence distances to 20 m.

The binocular vergence angle of the system was calibrated using fine adjustment screws on one of the mirrors in the optical path of each eye. Prior to calibrating the display system, the laser vergence gauge shown in Figure 2 was set to the indicate 20 m by shining it on a wall at that distance and carefully adjusting the angle between the red and green lasers to be coincident. This gauge was then mounted on a tripod and directed through the optical system at the LCD monitors while an alignment test pattern was displayed. The binocular display system mirrors were adjusted so the laser spots lined up with the alignment marks on the monitors. The accuracy of this alignment process is estimated to be approximately 0.5 arcmin.

The relative roll of the right and left eye images was adjusted to zero using a flat, partially silvered mirror that was temporarily placed in front of the display optical system. A bright white horizontal line positioned above the display system could be seen in reflection from the partially silvered mirror. This line was superimposed on the test patterns seen through the binocular optical system and was used to null the relative roll of the right and left images to within a tolerance of approximately 0.2 deg.

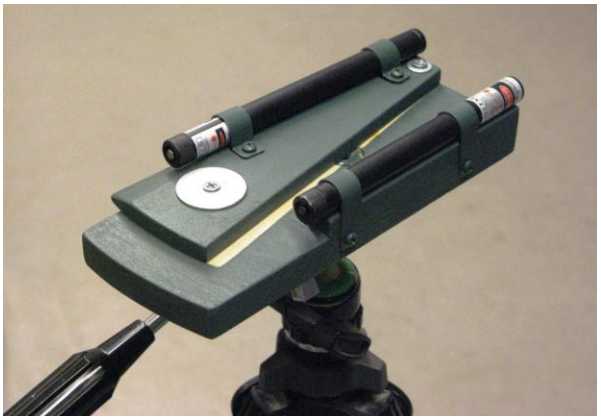

Figure 2. Photograph of the laser vergence gauge used to calibrate and check the binocular vergence distance of the display system.

Observers

Four paid observers, one female, participated in this evaluation. The observers were screened using a variant of the "Titmus Circle" test that is one of the accepted tests used to screen Air Force pilot and boom operator candidates. To participate in this evaluation observers had to respond correctly to at least 16 out of 20 trials with the disparity set to 25 arcsec. Observers had to correctly indicate the direction of the depth (near or far) as well as indicate which of the four circles (top, bottom, left, or right) differed in depth. Twelve observers were screened for this evaluation. The four who passed the test produced 20 correct responses.

Stimuli

On each trial the observer's task was to view a small rectangle positioned in a dark gap between two edges (see Figure 3) and to indicate if the rectangle was in front of or behind the edges. The task is much like the classic "Howard-Dolman" or three-bar tests of stereopsis that have been used for almost 100 years. In the version of the test used here, edges rather than bars were used for the comparison as this eliminated the possibility of using relative size (width) as a cue to distance.

The angular size of the rectangle was set to the approximate size of the receptacle on a receiver aircraft when viewed at the contact distance. The mean width of the rectangle was 0.25 deg while the mean height was 0.50 deg. The mean width of the gap between the edges was 1.25 deg, thus, the mean distance of the rectangle from an edge was 0.5 deg. The slope of the edges and rectangle was randomly varied from trial to trial over a range of +/- 1.5 deg relative to vertical. On trials for which antialiasing was insufficient and the display pitch was coarse, the observer could see definite spatial sampling artifacts (i.e., "stair stepping") on the edges and rectangle as in Figure 4. On each trial, the positions of the images were randomly jittered relative to the pixel mosaic so that a different pattern of sampling artifacts would be produced on each trial. For those conditions where sampling artifacts were present, they were spatially uncorrelated between the right and left eye images.

The luminance of the white edges and rectangle was 18 fL (60 cd/m2) and the dark gap was 2.2 fL for an edge contrast ratio of 8:1 (modulation = 0.78). On each trial the stimulus was presented for 1.4 sec, long enough to allow 2 or 3 fixations on the pattern. The results of pretesting indicated thresholds began rising as the stimulus duration was reduced below about 1 second.

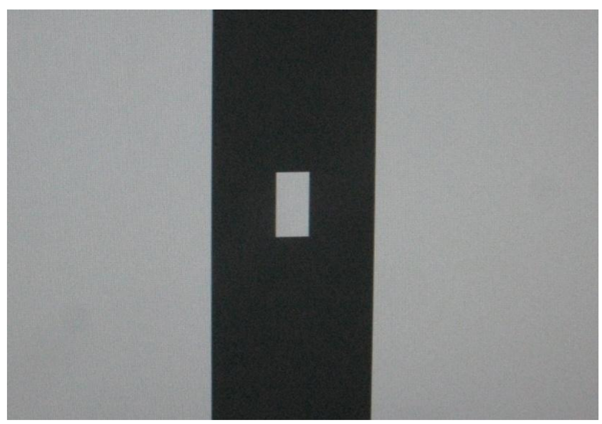

Figure 3. Close up photograph of the rectangle and edge pattern used in the evaluation. The rectangle subtended an angle of 0.25 x 0.5 deg for all experimental conditions. For this photograph the pixel pitch was set to 0.5 arcmin and the AA filter width was set to 1.5 pixels.

The antialiasing filter used to generate the images had a Gaussian shape and was applied to both the rectangle and the two vertical edges. The width parameter used in this paper refers to the half-maximum width of the Gaussian that was applied to the edges prior to spatial sampling.

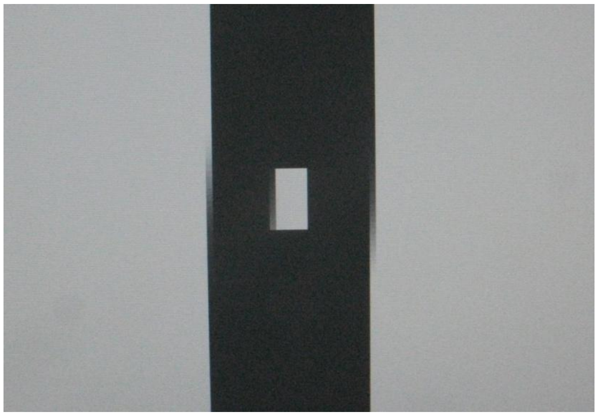

Figure 4. Close up photograph of the rectangle and edge pattern showing spatial sampling artifacts. For this photograph the pixel pitch was set to 3 arcmin and the AA filter width was set to 0.7 pixels.

Threshold Measurements

Binocular disparity thresholds were measured using a two-alternative, forced choice (2AFC) psychophysical staircase procedure that produced an average rate of correct responding of 0.75. Fixed-length sequences of 50 trials were used with the first five trials excluded from the threshold calculation. For the practice and first third of the trials, the starting level of disparity was set to 100 arcsec. For the remaining trials the starting level was set to 4 times the expected threshold based on a model fit to the initial data collected from that observer. Trial sequences were paced by the observer which allowed them to take a break at any time. The measurement of one threshold required approximately 2.5 minutes to complete and approximately 20 thresholds could be measured in an hour. Sessions were limited to one hour per day to avoid fatigue. Each observer participated in 8-9 sessions that were spread over a period of 3 weeks.

Ratings of Comfort

Immediately after completing each threshold measurement (sequence of 50 trials) observers provided a rating of the visual comfort of that condition using a rating scale ranging from 1 to 5. Two verbal anchors were used for the scale: a 1 on the scale indicated the condition was “comfortable” and a 5 indicated the condition was “very uncomfortable.” Observers were instructed to assume they would have to use a display system of that design for several hours per day.

Experimental Design

Threshold measurements were made for 42 combinations of pixel pitch and antialiasing filter kernel width. Six levels of pixel pitch were used and these were created using “pixel replication” such that the resulting pitches were integer multiples (1 to 6) of the 0.48 arcmin native resolution of the display system. Between six and eight levels of antialiasing kernel width were used at each level of pixel pitch. The specific levels of width selected were based on the results of pre-testing and were positioned to most efficiently reveal the shape of the surface.

For each observer the threshold for each experimental condition was measured at least 3 times. For those conditions that produced the highest variance, one or two additional threshold measurements were made to stabilize the estimates for those conditions. An average of 147 threshold measurements were made for each observer for an average of 3.5 estimates per condition.

RESULTS

Disparity Thresholds

Prior to model fitting, the data of each observer were reduced by calculating median threshold for each of the 42 experimental conditions. The median data were then averaged across the observers, thus, the threshold estimates at each condition were based on a average of 14 measurements.

In a series of preliminary analyses, a stepwise multiple linear regression approach was taken using a wide range of 2-D polynomial models with second and third order terms in the main effects and interactions, and nonlinear transformations of both the dependent and independent variables. While some of these models produced a higher correlation with the measured data than did our final model, they were rejected as their behavior was considered to be implausible when extrapolated beyond the range of the data. Our final model was fitted by minimizing the sum of the squared residual errors using the multi-dimensional nonlinear optimizer function (fminsearch) available in the MATLAB software.

Examination of the threshold data revealed distinct regions within the overall surface created by the data. For those conditions where a wide antialiasing filter was used, there was a distinct “floor” in the threshold data that slowly increased up a “hill” as the pixel pitch increased (see left side of Figure 5). As the width of the antialiasing filter was reduced, the threshold first remained level and then climbed a steep “cliff” as it was reduce further. Given the distinctness of these regions, a multi-component model was fit to the data and the contributions of each independent component were combined to form the complete model. The nonlinear parameter optimizer was allowed to simultaneously vary all six parameters of the model. The first three model parameters were:

Kf: controls the height of the "floor"
Kh: controls the magnitude of the "hill"
Kb: controls the magnitude of the "blur" (angular width of the antialiasing filter, in arcmin)

Three separate parameters are used to set the position and slope of the cliff region in the model:

Kp: controls the position in the pitch dimension
Kw: controls the position in the width dimension
Ks: controls the slope of the cliff

The magnitude of the floor of the model is simply the first parameter which is scaled in units of arcsec. The second and third parameters are used to scale the effects of pixel pitch and the product of pitch and width:

H = Kh * pitch
B = Kb * pitch * width

The last three parameters are used to calculate the "cliff" region of the model:

C = Ks * [Kw * log10(Kp + pitch) - width]

Anywhere the quantity C goes negative it is set to zero:

C(C <) = 0

The four model components are then combined in quadrature to calculate the disparity threshold surface:

Disparity = sqrt(Kf^2 + H^2 +B^2 + C^2)

The results of the fitting process produced the following levels for the six free parameters in the model:

Kf = 5.544
Kh = 2.038
Kb = 1.112
Kp = 0.763
Kw = 3.830
Ks = 22.15

Using these levels, the correlation between the model and the mean threshold data was R^2 = 0.880 ($p <$ 0.0001, 35 df). Figure 5 shows a surface plot of the resulting model while Figure 6 shows the same model in the form of a contour plot.

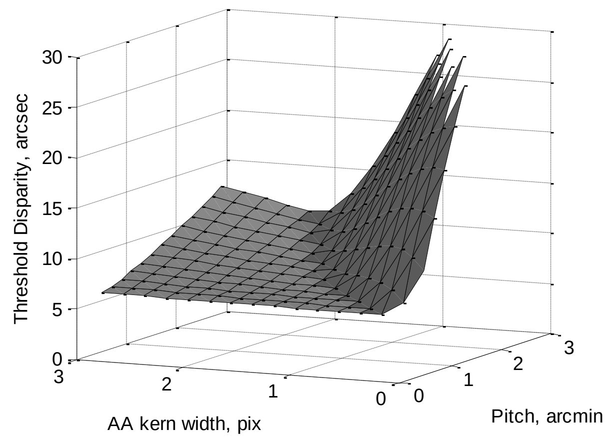

Figure 5. Model of threshold disparity as a function of pixel pitch and antialiasing filter kernel (half-max) width. The correlation between the model and the mean data of four observers was R^2 = 0.880 (p < 0.0001, 35 df).

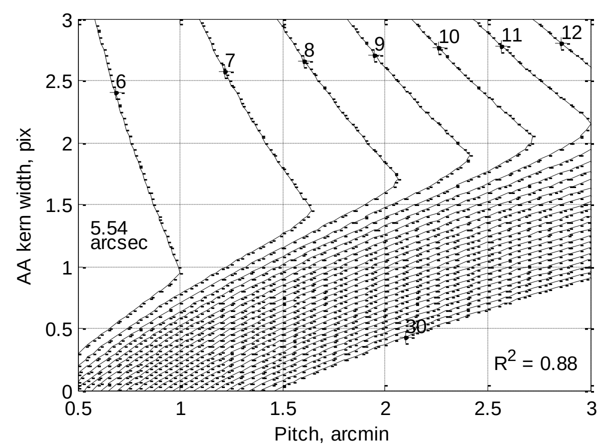

Figure 6. Contour plot of the disparity threshold model presented in Figure 5.

Viewing Comfort

Prior to fitting the model of viewing comfort, the rating scale data were reduced by averaging across replicates and observers. Thus, the means of between 14 and 15 estimates of comfort were used for each of the 42 experimental conditions. These mean comfort data were then used to fit a model of the same form used to describe the threshold disparity data. The results of the multidimensional fitting process produced the following levels for the six free parameters in the model:

Kf = 1.352
Kh = 0.846
Kb = 0.341
Kp = 1.474
Kw = 3.315
Ks = 2.786

Using these levels, the correlation between the model and the mean comfort rating data was $R^2 = 0.964$ ($p < 0.0001$, 35 df). Figure 7 shows a surface plot of the resulting model while Figure 8 shows the same model in the form of a contour plot.

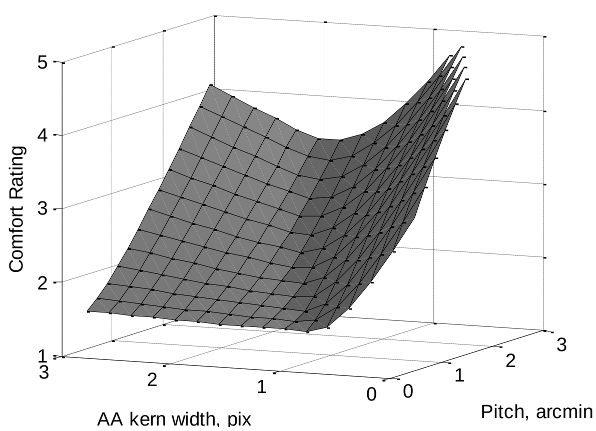

Figure 7. Model of viewing comfort as a function of pixel pitch and antialiasing filter kernel (half-max) width. The correlation between the model and the mean data of four observers was $R^2 = 0.964$ ($p < 0.0001$, 35 df).

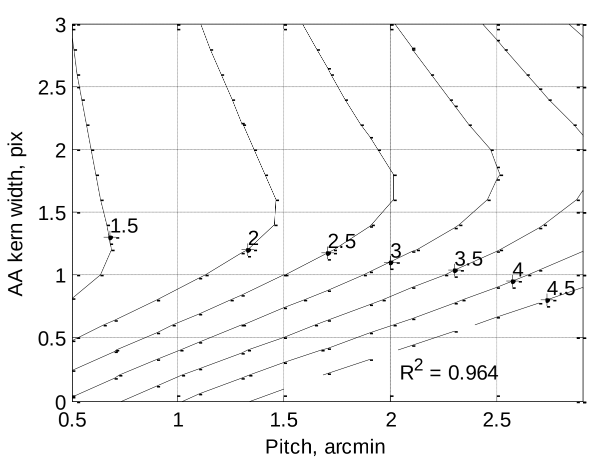

Figure 8. Contour plot of the viewing comfort model presented in Figure 7.

DISCUSSION AND CONCLUSIONS

The mean binocular disparity threshold for the four observers who participated in this evaluation was 5.5 arcsec in the floor region of the model. These data clearly confirm our expectation that thresholds in the range of 3 to 10 arcsec can be obtained using electronic displays when sufficient resolution and antialiasing are employed. Recall that the authors cited in the introduction (Tiddwel 1989, Yeh & Silverstein, 1989) reported a disparity threshold of 140 arcsec (2.3 arcmin) for the display systems they used, a level 25 times higher than the thresholds we obtained.

Perhaps the most important finding from this evaluation is that it illustrates the large difference the two design variables have on performance. Examination of Figure 5 reveals the slope of the hill due to increasing display pitch is much shallower than the slope of the performance cliff caused by insufficient antialiasing. When we consider that the cost of reducing the display pitch is much greater than the cost of increasing the amount of antialiasing applied by an image generator, we can see the performance/cost ratio of improved antialiasing is far greater than that of improved resolution. It is surprising that the influence of antialiasing on stereoscopic display effectiveness has received so little attention by scientists and applications engineers over the past two decades.

Comparison of the comfort model in figure 7 with the threshold model in Figure 5 reveals the effects of the design variables are more evenly balanced for the comfort model. With the comfort model the slope of the hill region is higher and the slope of the cliff region is lower than they are in the threshold model, leading to a less well defined crease between the regions.

It is informative to compare our results with the findings of Singer (1995) who concluded the stereoscopic cue provides benefit only at distances of less than about 10 ft. In that evaluation he used a display with a pixel pitch of 14 arcmin... 4.7 times larger than the largest pitch used in our evaluation. The models we present predict that his display system would produce very poor disparity cues and that viewing comfort would be well beyond “very uncomfortable.” Our models are entirely consistent with Singer’s findings.

Considering both the threshold and comfort models, we can see that real world levels of stereoscopic performance and reasonably comfortable display systems can be achieved using pixel pitches as coarse as

1.5 to 2 arcmin, but only if sufficient antialiasing is applied. The effective width of the antialiasing function that is required depends on pixel pitch as indicated by the base of the cliff in Figure 6.

Thus far we have found no significant technical or cost barriers to employing sufficient antialiasing in training display systems. We suspect a primary reason sufficient antialiasing is not performed is that designers do not want to "waste" spatial resolution. This aversion to applying "too much" antialiasing seems to have led inadvertently to a significant underestimation of the utility of stereoscopic displays for training applications.

REFERENCES

Bradshaw, M. F., A. D. Parton, et al. (2000). "The task-dependent use of binocular disparity and motion parallax information." Vision Research **40**(27): 3725-3734.

Couvillion, W. C., Roberts, A. J., and Newell, B. D. (2003). KC-135 Boom Operator Training Hi-Fidelity Stereoscopic Display Technology Demonstrator. *Interservice / Industry Training, Simulation & Education Conference.* NTSA, Orlando.

Cutting, J. E. and P. M. Vishton (1995). "Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth." Perception of space and motion: 69–117.

Durgin, F. H., D. Proffitt, et al. (1995). "Comparing depth from binocular disparity to depth from motion." Journal of Experimental Psychology: Human Perception and Performance **21**: 679–699.

Frisby, J. P., D. Buckley, et al. (1996). "Evidence for good recovery of lengths of real objects seen with natural stereo viewing." PERCEPTION-LONDON- **25**: 129-154.

Gregory, R. (1965). "Seeing in depth." Brian Physiology and Psychology: 164.

Jameson, D. and L. M. Hurvich (1959). "Note on factors influencing the relation between stereoscopic acuity and observation distance." Journal of the Optical Society of America.

Kruk, R. and D. Runnings (1989). Low level flight performance, and air combat maneuvering performance in a simulator with a fiber optic helmet mounted display system.

Lloyd, C. J., and Nigus, S. G. (2012) Effects of stereopsis, collimation, and head tracking on air refueling boom operator performance. *Proceedings of the IMAGE Conference,* Scottsdale.

McKee, S. P. (1983). "The spatial requirements for fine stereoacuity." Vision Research **23**: p. 191-198.

Nagata, S. (1991). Chapter 36: How to reinforce perception of depth in single two-dimensional pictures. Pictorial communication in virtual and real environments, Taylor & Francis.

Palmer, S. E. (1999). Vision science: Photons to phenomenology, MIT press Cambridge, MA.

Parrish, R. V., A. M. Busquets, et al. (1990). Recent research results in stereo 3-D pictorial displays at Langley Research Center, IEEE.

Pfautz, J. D. (2001). Sampling artifacts in perspective and stereo displays.

Proffitt, D. R. and M. K. Kaiser (1986). "The use of computer graphics animation in motion perception research." Behavior Research Methods **18**(6): 487-492.

Singer, M. J., J. Ehrlich, et al. (1995). Task Performance in Virtual Environments: Stereoscopic Versus Monoscopic Displays and Head-Coupling, DTIC Document.

Teichner, W. H., J. L. Kobrick, et al. (1956). "Effects of target separation and distance on commonplace binocular depth discrimination." JOSA **46**(2): 122-125.

Tidwell, R. (1989). "Stereopsis as a visual cue in flight simulation." Journal of Aircraft: 188-194.

Tidwell, R. P. (1990) Stereopsis takes off in flight simulation. *Proceedings of the IEEE Southeastcon Conference.* Pp. 578-582.

Yeh, Y. Y. and L. D. Silverstein (1989). Using electronic stereoscopic color displays: Limits of fusion and depth discrimination. SPIE.