Effect of Disparity Noise on Stereoscopic Surface Perception in Human and Ideal Observers

S. Palmisano[Ψ], R.S. Allison[E] & I.P. Howard[E] ([Ψ]Department of Psychology, University of Wollongong, NSW, Australia, 2522, email: stephenp@uow.edu.au; [E]Centre for Vision Research, York University, Toronto, Ontario, Canada, M3J 1P3)

Abstract: Stereoscopic surface detection of human and ideal observers was assessed using a signal detection paradigm. Signal displays were disparity defined sinusoidal or square wave corrugations in depth containing various amounts of additive disparity noise. Distracter displays were created by scrambling pure signal stimuli along the vertical dimension - destroying surface representation while leaving the depth range intact. Additive disparity noise was found to interfere with stereoscopic surface detection for both human and ideal observers. Efficiencies found for stereoscopic surface detection were similar to those found previously for detection of a single step edge in depth (a supposedly easier task).

1 Introduction

Due to their two horizontally separated eyes and overlapping visual fields, humans simultaneously receive two different perspective views of the same scene. The term *stereopsis* refers to the processes by which these different 2-D views are combined and used to extract information about the 3-D layout of the environment. Random dot stereograms (RDS) are a useful tool for examining these processes (Julesz, 1960). The 3-D surfaces represented by them are defined solely by positional shifts of corresponding dots in the two half-images (referred to as binocular disparities). RDS produce the following problems for stereopsis: (1) Corresponding dots in the two images must be matched; (2) The binocular disparity of each matched dot pair must be accurately extracted; (3) Disparity information from across the visual field must be combined; (4) Depth, surface slant, inclination and curvature must then be calculated based on this 'disparity field'. The human visual system appears to be able to perform these processes under a variety of difficult conditions. For example, observers can detect 3-D surface structure in RDS which: (1) have very low or very high dot densities (4-40% of the area - Tyler, 1974); (2) have large numbers of unpaired or monocular dots (Julesz, 1960; 1971; Christophers & Rogers, 1994; Cormack et al, 1991; Cormack et al, 1994; Cormack et al, 1997; Palmisano et al, 2000; Wilcox & Hess, 1996); and (3) have substantial amplitudes of additive Gaussian distributed disparity noise (Harris & Parker, 1992; 1994a; 1994b; Lankheet & Lennie, 1996; Palmisano et al, 1999a; 1999b).

In a series of studies, Harris and Parker (1992; 1994a; 1994b) examined the statistical efficiency of these stereoscopic processes. RDS stimuli always represented a vertically oriented step edge in depth – produced by an appropriate disparity shift – with various amounts of Gaussian distributed additive disparity noise. The task for human and ideal observers was to determine which side of the stimulus "stood further out towards them in depth". Unlike human observers, the ideal observer was able to use all of the available stimulus information to perform this task. A comparison of the performance of these two types of observer yielded the efficiency measure (Rose, 1942; 1948). In their first study, Harris and Parker (1992) examined detection of a step change in disparity (of 23''- 120'') in the presence of additive disparity noise (σ = 2' - 4') as the number of dots was increased. They found efficiencies fell dramatically (down from ~30% to ~2%) as the number of dots in the RDS increased from 4 to 350. Unlike the ideal observer who used all of the dots, the effective numbers of dots used by human observers on either side of the depth discontinuity never exceeded 5 (as many as 170 dots were available in some conditions). Harris and Parker argued that this poor performance could be attributed to either difficulties in dot matching (Stage 1 of stereoscopic processing) or to inefficiencies in combining disparity samples (Stages 3 and 4).

A subsequent study by Harris and Parker (1994a) attempted to separate these two sources of efficiency loss. In their first experiment, they manipulated the amplitude of the additive disparity noise (rather than density, which was held constant at 240 dots). They also varied the step change in disparity (0.7'-2.1'), to keep the human observers' d' values as close as possible to 1 (to minimize sampling errors). They found that efficiencies in detecting a step edge in depth declined (from ~10% to ~0.1%) as the standard deviation of the additive disparity noise increased from 1' to 6'. Based on the results of two followup experiments, which used RDS stimuli that minimized or eliminated the correspondence problem, they concluded that: (1) both dot matching difficulties and inefficiency in combining disparity samples led to independent declines in efficiency; (2) of the two components, difficulties in dot matching were responsible for the most dramatic falls in efficiency.

However, it must be noted that this particular pattern of results might have been due to the simplicity of the stereoscopic task used in the Harris and Parker studies. It is possible that for more complicated tasks, such as stereoscopic surface detection, disparity combination difficulties would prove as disruptive as dot matching difficulties. The current experiments examined this possibility. Human and ideal observers had to decide whether or not the stimulus presented was a noisy version of a corrugated surface (a 3-D sinusoid or square wave) or a distracter. Based on human and ideal observer performance, efficiencies were calculated for these surface detection tasks.

2 Human Observer Experiments

This experiment examined the effect of additive disparity noise on human observer's ability to detect surfaces with periodic corrugations in depth. In principle, the stereoscopic detection of a step edge in depth could be achieved with information obtained from a relatively small subset of dots – if they were matched correctly and their disparities were accurately extracted. This task only requires the observer to detect a significant difference in disparities in the two regions of the display. Difficulties combining disparity samples could potentially have little effect on this task, since disparity information from only a few matched dots would be required. However, the stereoscopic detection of a periodic surface - such as a 3-D square wave or a 3-D sinusoid - is a potentially more complex process. Not only is accurate dot matching essential to this task, accurate disparity combination and calculation is also required. As a result, we expected that additive disparity noise would have a greater impact on surface detection than it had on detection of a step edge in depth (Harris & Parker, 1992; 1994a).

Human Observers. Three observers (aged between 24 and 39 years) participated. SAP (the first author), XF and MH (naive to the experimental hypotheses) had participated in many previous experiments on stereoscopic surface detection. All had normal or corrected-to-normal vision and had been given several hundred test trials before their data was collected.

Stimuli & Procedure. RDS were generated on a Macintosh G3 Power PC and later presented on a 17 inch Apple Vision monitor. A display splitter was used to present RDS to observers wearing CrystalEyes liquid crystal shutters. The splitter alternated the presentation of the left and right eyes' views on the screen in synchrony with the shuttering of the glasses (60Hz), which ran at half the video card refresh rate (120Hz). RDS were two stereo-half images – each consisted of 4665 bright blue dots on a black background and subtended 9˚ H x 9˚ V at the viewing distance of 110cm. Each "dot" subtended an angle of 4 $arcmin^2$ and had a luminance of 0.15 cd/m^2. RDS were of two kinds. (1) Signal Displays were stereoscopically defined 3-D surfaces (horizontally oriented square wave or sinusoidal surfaces in depth) with one of three spatial frequencies - 0.22cpd, 0.44cpd, or 0.88cpd. Surface phase was varied randomly from trial to trial. The pattern of horizontal disparities defining such a surface was produced by shifting dots in opposite horizontal directions in the left and right half-images (disparity ranged from +2' to –2'). Gaussian distributed disparity noise was then added to these dots (σ of 0, 2', 4', 6' or 8'). (2) Noise Displays were created by scrambling signal stimuli along the vertical dimension. This destroyed surface representation while preserving stereoscopic information about depth. For each signal or noise display, observers indicated whether or not they saw the signal. The display was presented until the observer pressed one of two buttons ("yes" and "no"), and then display turned black. This was followed by a 2s intertrial interval - to reduce afterimages and disparity aftereffects. Observers performed eight experimental runs of 600 stimuli.

Results of Experiments. "Yes" responses in the presence or absence of a stereoscopically defined 3-D surface were converted into hit rates (H) and false alarm rates (FA). These then were converted into z-scores and used to calculate d' - the measure of sensitivity used in signal detection theory:

$$d_e' = z(H) - z(FA),$$

$$var(d_e') = H(1-H)/N_H[\phi(H)]^2 + FA(1-FA)/N_{FA}\,[\phi(FA)]^2,$$

where N_H = number of hits, N_{FA} = number of false alarms, $\phi(H) = 2\pi^{-1/2}\exp[-0.5z(H)^2]$, and $\phi(FA) = 2\pi^{-1/2}\exp[-0.5z(FA)^2]\}$.

Stereoscopic sinusoid detection was remarkably robust in the presence of substantial RMS amplitudes of disparity noise (see Figure 1). Sinusoid detection performance was similar to that reported by Lankheet and Lennie (1996) - even though static RDS were used in the current experiment, while dynamic RDS were used in the latter.

In both experiments, disparity noise with RMS noise amplitudes greater than 4' tended to reduce detection performance to chance ($d_e' = 0$). On the basis of the current findings, Lankheet and Lennie's claim that depth of both the additive disparity noise and the grating is more difficult to resolve in dynamic RDS appears questionable. Sinusoid detection performance was more sensitive for low spatial frequencies (0.22cpd) than for high spatial frequencies (0.88cpd). While stereoscopic square wave detection was also robust in the presence of substantial RMS amplitudes of additive disparity noise (see Figure 2), it was less tolerant than stereoscopic sinusoid detection. Square wave detection first fell to chance levels with RMS noise amplitudes of 4', whereas sinusoid detection first fell to chance with RMS noise amplitudes of 6'.

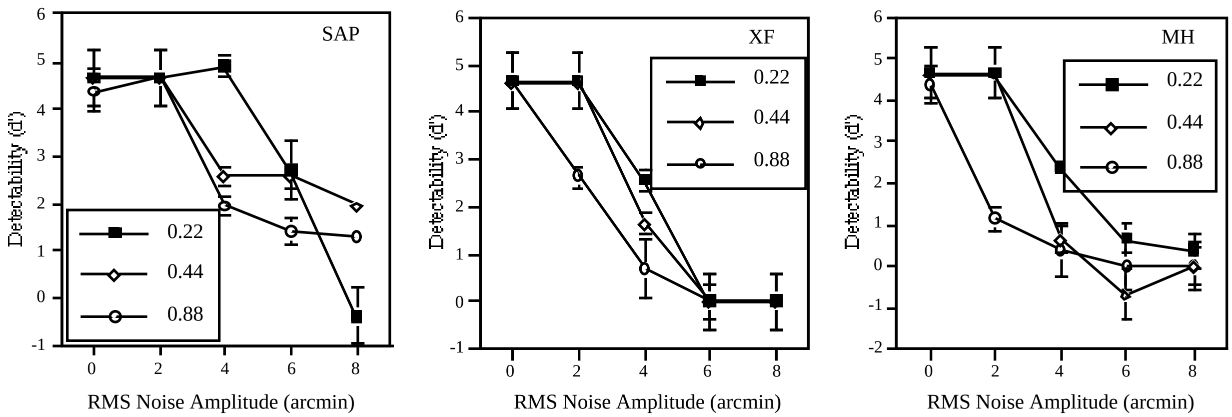

Figure 1: Sinusoid Detection (de') as a function of RMS amplitude of disparity noise and surface spatial frequency (0.22, 0.44, 0.88cpd) for 3 human observers (SAP, XF, MH). Error bars represent standard errors of the mean.

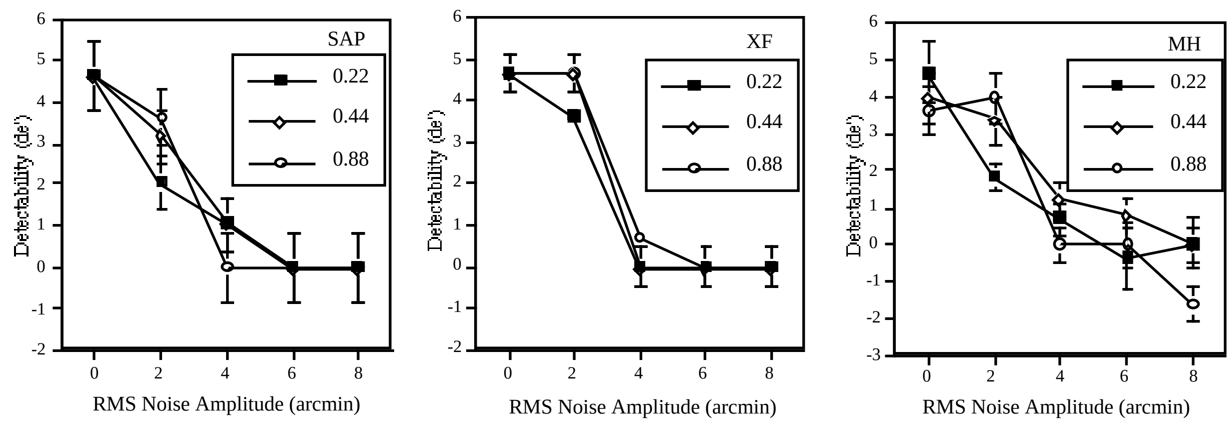

Figure 2: Square wave Detection (de') as a function of RMS amplitude of disparity noise and surface spatial frequency (0.22, 0.44, 0.88cpd) for 3 human observers (SAP, XF, MH). Error bars represent standard errors of the mean.

3 Ideal Observer Simulations

We constructed an ideal observer that utilised all of the available information about each RDS. We assumed that this model observer could perform the matching task correctly and recover the ideal disparity map. Thus, its task was to determine from this disparity map, whether or not a corrugated disparity signal had been displayed.

We wanted the ideal observer to minimise the probability of detection error. Given the disparity image map, **d**, the probability of error in detection for a disparity corrugation signal (denoted as s_i) is (Haykin, 1988):

$$
\begin{aligned}
P_e(s_i|\mathbf{d}) &= P(s_i \text{ not displayed}|\mathbf{d}) \\
&= 1 - P(s_i \text{ displayed}|\mathbf{d})
\end{aligned}
$$

Minimising the error in detection is equivalent to maximising the a posteriori probability that the signal was displayed given the data (Kersten, 1990). From Bayes' rule:

$$P(\mathrm{s_i\ displayed}|\mathbf{d}) = \frac{P(\mathrm{s_i\ displayed})\, f_d(\mathbf{d}|\mathrm{s_i\ displayed})}{f_d(\mathbf{d})}$$

Since the probability of displaying each signal is equal and $f_d(\mathbf{d})$ is independent of the signal displayed, we can dispense with the Bayesian formulation. The maximum a posteriori decision rule reduces to deciding that s_i was displayed if

$$f_d(\mathbf{d}|\mathrm{s_i\ displayed})$$

was the greater for s_i than for any other disparity corrugation signal or distracter. This is the optimal (maximum-likelihood) decision rule. An ideal, maximum likelihood observer for the detection of known signals in additive, white Gaussian noise is known to be a matched filter. Since the ideal observer knew the stimulus phase of each RDS, detection with a matched filter was equivalent to cross-correlating with the expected signal (Burgess & Ghandeharian, 1984). Since we used periodic signals - sinusoids and square wave corrugations - sampled at random positions (the dot locations), the matched filter operation was equivalent to correlating with a template corresponding to the ideal disparity corrugations, sampled at the locations of the texture elements – i.e. at a discrete number of positions.

We ran two sets of simulations – one set where the signal stimuli were surfaces with sinusoidal modulations in depth with one of three spatial frequencies (0.22, 0.44, or 0.88cpd), and the other set where the signal stimuli were surfaces with square wave modulations in depth with the same spatial frequencies. In each set of simulations, the ideal observer matched the disparity map from a RDS with three matched filters – corresponding to the three signal spatial frequencies. The noise was orthogonal to each signal and the signals were orthogonal to each other. Thus, the presence of a pure signal should have produced a response along one of the three axes equal in magnitude to the signal energy. Theoretically, such a signal should have produced zero responses along the other orthogonal dimensions. However, trial to trial random sampling bias in the position of the dot locations caused fluctuations in the magnitude along the principle dimension for a given corrugation and led to small positive or negative responses along the other dimensions. This variability generally cancelled across trials. The more significant noise was typically the additive, white Gaussian noise. For detection of a signal in noise, the signal to noise ratio at the output of the matched filter would have been equal to the ratio of the energy in the ideal signal to the noise spectral density (Haykin, 1988). Hence, the matched filter showed selectivity by improving the signal to noise ratio.

Each stimulus mapped on to a co-ordinate in a three-dimensional Euclidean signal space. Along the axis relevant to a given signal we would have two probability density functions (PDFs). One PDF corresponded to the signal (centred at the signal energy), while the other PDF corresponded to the distracter (centred at zero). Each had a variance proportional to the noise spectral density. The perceptual distance between noise stimuli and signal stimuli was the difference between the filter response to the signal and the filter response to the distracter relative to the variation in that difference measure. The d' value was the ratio of the difference between signal and distracter responses scaled by the standard deviation of that difference (it was basically a measure of signal to noise ratio along the relevant dimension in terms of z-scores). If the response along any one of the axes exceeded a threshold amount then the ideal observer declared that one of the signals was present. If equal noise was present in the distracter and signal stimuli then the unbiased decision rule would place the decision threshold halfway between the origin and the expected response when the signal was present. This would be the optimal maximum likelihood decision rule since the probability density functions corresponding to signal and distracter were equal here.

Results of Simulations. Monte-Carlo simulations of an ideal observer performing the experiment were run and analysed. One thousand randomly generated trials were performed for each condition and used to calculated d' from the hit and false alarm rates. A signal was declared present by the ideal observer if response on any of the signal axes exceeded one half of the expected response. The ideal observer's detection performance was far superior to that of the human observers - remaining very robust in the presence of substantial amounts of disparity noise (d' values never falling below 3). The ideal observer was still performing above chance levels with RMS amplitudes of disparity noise of 10' and 12' (which were double the RMS noise amplitudes at which human observers were performing at chance levels). These large discrepancies in performance were expected – since for example, the ideal observer would have used disparity information from all of the dots, whereas human observers would have used only a small subset of these dots (Harris & Parker, 1992). Figure 3A shows that for sinusoid corrugations the ideal observer's performance declined from d_i' values of ~15 down to ~3 as the standard deviation of the additive disparity noise increased from 2'-12'. Figure 3B shows that the ideal observer's performance was generally better for square wave corrugations. Since a square wave corrugation segment has more signal energy than a sinusoidal corrugation segment of the same peak amplitude, this finding was expected. For square wave corrugations, ideal observer performance declined from d_i' values of ~15 down to ~4 as the RMS amplitude of the disparity noise increased from 2'-12'. Interestingly, square wave detection appeared to decline in a more linear fashion than sinusoid detection with increasing RMS amplitudes of disparity noise. However, this difference might have been due to the very high d_i' values found for displays with RMS noise amplitudes of 2'-6'. In general, the ideal observer's detection performance was similar for all three stimulus spatial frequencies - for both sinusoid and square wave detection.

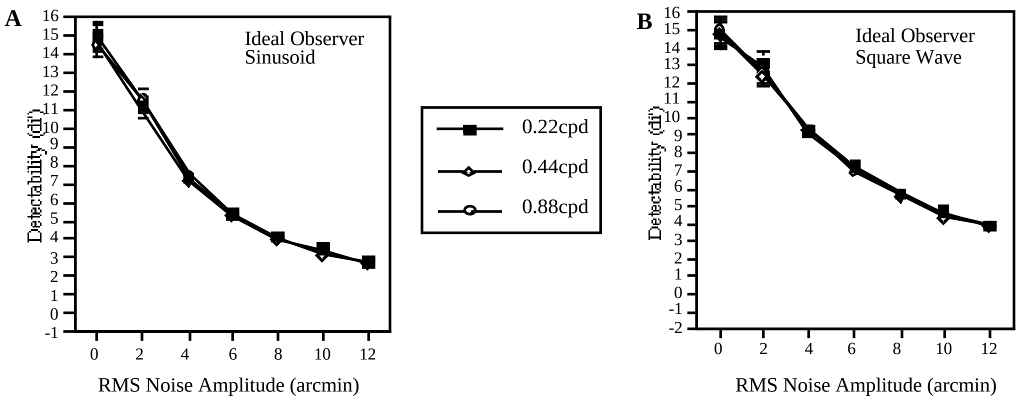

Figure 3: (A) Sinusoid and (B) Square wave detection (di') as a function of RMS amplitude of disparity noise and surface spatial frequency (0.22, 0.44, 0.88cpd) for ideal observer. Error bars represent standard errors of the mean.

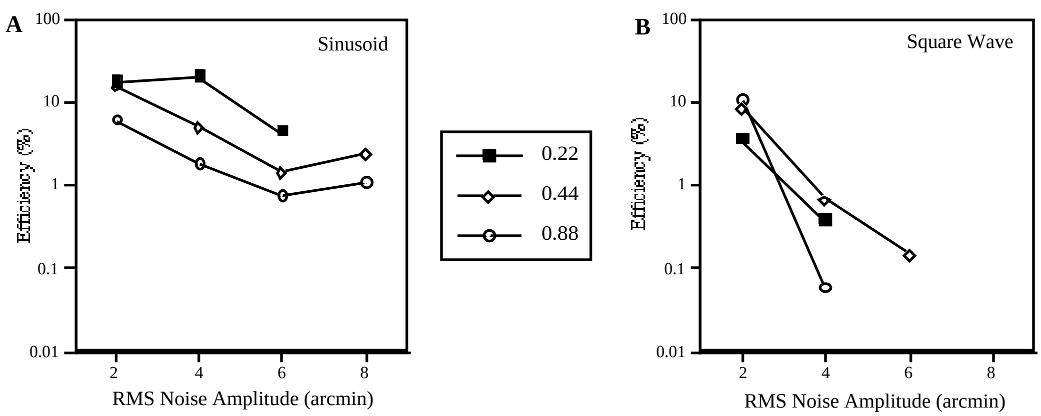

Figure 4: (A) Sinusoid and (B) Square wave detection efficiency as a function of RMS amplitude of disparity noise and surface spatial frequency (0.22, 0.44, 0.88cpd).

4 Surface Detection Efficiencies

We next compared human and ideal detection performance to calculate the statistical efficiency (F) for these surface detection tasks, which can be defined as follows:

$$F = (d_e'/d_i')^2$$

where d_e' is the d' value for human observers and d_i' is the d' value for the ideal observer respectively (Barlow, 1978). The d_e' values were averaged for the three human observers. Efficiency fell for both square wave and sinusoid detection tasks as the amplitude of the disparity noise increased. While square wave detection efficiency fell sharply, sinusoid detection efficiency declined more gradually. Figure 4A shows that efficiencies for sinusoid detection fell from 17%-6% with 2' RMS amplitudes of disparity noise down to 2.5%-1% with 8' RMS amplitudes of disparity noise.

Figure 4B shows that efficiencies for square wave detection fell dramatically from 10%-3% with 2’ RMS amplitudes of disparity noise down to 0.7%-0.05% with only 4’ RMS amplitudes of disparity noise. Stimulus spatial frequency did appear to have some effect on efficiency – especially in the case of sinusoid detection, which was more efficient with low spatial frequency sinusoids (0.22cpd) than with high spatial frequency sinusoids (0.88cpd).

5 General Discussion

Contrary to predictions, we found efficiencies for stereoscopic surface detection (20%-0.1%) were similar to those found previously by Harris and Parker (1994a – Experiment 1) for stereoscopic edge detection. However, before we compare the efficiencies of these two studies, some checks need to be made. In calculating efficiencies, Harris and Parker ensured that their human d’ values were always close to 1 by changing the size of disparity step as they increased the amplitude of the disparity noise. However, we did not place the same restriction on our human observers – human d’ values ranged between ~5 and 0 as the amplitude of the disparity noise increased (the disparity range of sinusoids and square waves was always a constant 4’). The most similar stimulus in the current experiments to the Harris and Parker step edge was the 0.22cpd square wave (it had 2 step edges in depth as opposed to 1). For two observers, d_e’ values were close to 1 when 4’ RMS amplitudes of disparity noise were added to this stimulus – producing efficiencies of ~1.25% (SAP) and ~1.8%-0.5% (MH). Since these efficiencies were within Harris and Parker’s range of 10%-0.1% and close to their efficiencies found with 4’ RMS amplitudes of disparity noise, they would appear to vindicate our study – since similar tasks produced similar efficiencies.

In principle, human deviations from ideal performance could have arisen from a number of sources of error. Unlike the ideal observer, human observers had an imperfect knowledge of the stimuli (ie they had no prior knowledge of stimulus phase, spatial frequency, correct dot matches, etc). As a result, they were vulnerable to matching difficulties, imperfect disparity recovery and problems combining disparity samples – which would have been exacerbated by the externally applied disparity noise. Human observers also had an additional source of external noise – the crosstalk between the left and right stereo half-images (~8%) – which would have further impaired stereoscopic surface detection. Finally, internal noise – ranging from registration errors, to threshold criterion interactions and guessing behavior - would have also impaired performance.

Efficiencies for sinusoid detection were generally higher than those for square wave detection – which was expected since the square wave detection task by definition produces higher d_i’ values for the ideal observer than the sinusoid detection task. However, human observers higher d_e’ values for sinusoids would also have contributed to this effect. This finding suggests that stereo processes used by the visual system are specialized for sinusoidal surfaces, just as those used by the ideal observer were specialized for square wave surfaces. For example, the processes used by visual system might prefer smoothly continuous surfaces – as opposed to discontinuous surfaces. A smoothly continuous surface potentially simplifies the dot matching process (since adjacent dots would have very similar disparities) as well as the processes of combining disparity samples and depth calculation (since these disparities would be differentiable). Irrespective of the particular source of the discrepancy, this finding highlights the differences in the strategies used by the visual system and machine vision.

There was some evidence that higher spatial frequency surfaces produced lower efficiencies than lower spatial frequency surfaces. This supports the notion that detecting a single step edge in depth is a simpler task for the visual system than detecting a 3-D surface with numerous variations in curvature and depth (more dots/disparity samples are required to detect a surface than a step edge in depth). Since dot density was constant (4%) for all displays, the more variations in surface structure the fewer the dots/disparity samples there were defining each of these variations. As a result, high spatial frequency surfaces would have been especially susceptible to disparity noise. This noise would have led to difficulties matching dots – resulting in fewer disparity samples to define each segment. Even if those dots had been matched correctly, their disparities would have been quite different from those of their neighbors, which might have led them to be rejected at the disparity combination stage.

In conclusion, efficiencies found for stereoscopic surface detection were similar to those found previously for detection of a single step edge in depth (a supposedly simpler task). It is also possible that the higher efficiencies found in these experiments for sinusoidal surfaces indicates that the visual system has been optimized for perceiving smoothly continuous surfaces – rather than surfaces with multiple depth discontinuities.

6 References

Barlow, H.B. (1978). The efficiency of detecting changes in density in random dot patterns. *Vision Research*, **18**, 637-650.

Burgess, A.E. & Ghandeharian, H. (1984). Visual signal detection. I. Ability to use phase information. *Journal of the Optical Society of America A*, **1**, 900-906.

Christophers, R.A., & Rogers, B.J. (1994). The effect of viewing distance on the perception of random dot stereograms. *Investigative Ophthalmology & Visual Science*, **34**, 1438.

Cormack, L.K., Stevenson, S.B., & Schor, C.M. (1991). Interocular correlation, luminance contrast and cyclopean processing. *Vision Research*, **31(12)**, 2195-2207.

Cormack, L.K., Stevenson, S.B., & Schor, C.M. (1994). An upper limit to the binocular combination of stimuli. *Vision Research*, **34(19)**, 2599-2608.

Cormack, L.K., Landers, D.D., & Ramakrishnan, S. (1997). Element density and the efficiency of binocular matching. *Journal of the Optical Society of America A*, **14(4)**, 723-730.

Harris, J.M., & Parker, A.J. (1992). Efficiency of stereopsis in random-dot stereograms. *Journal of the Optical Society of America A*, **9(1)**, 14-24.

Harris, J.M., & Parker, A.J. (1994a). Constraints on human stereo dot matching. *Vision Research*, **34(20)**, 2761-2772.

Harris, J.M., & Parker, A.J. (1994b). Objective evaluation of human and computational stereoscopic visual systems. *Vision Research*, **34(20)**, 2773-2785.

Haykin, S. (1988) *Digital Communications.* New York: Wiley.

Julesz, B. (1960). Binocular depth perception of computer generated patterns. *Bell System Technical Journal*, **39**, 1125-1162.

Julesz, B. (1964). Binocular perception without familiarity cues. *Science*, **145**, 356-362.

Julesz, B. (1971). *Foundations of cyclopean perception.* Chicago, Ill: University of Chicago Press.

Kersten, D. (1990). Statistical limits to image understanding. In *Vision: Coding and Efficiency.* C. Blakemore ed. Cambridge: Cambridge University Press. pp 32-44.

Lankheet, M.J.M., & Lennie, P. (1996). Spatio-temporal requirements for binocular correlation in stereopsis. *Vision Research*, **36(4)**, 527-538.

Palmisano, S., Allison, R.S., & Howard, I.P. (1999a). Effects of decorrelation and disparity noise on stereoscopic surface perception. *Investigative Ophthalmology & Visual Science*, **40(4)**, 2199.

Palmisano, S., Allison, R.S., & Howard, I.P. (1999b). Stereoscopic detection and segregation of noisy transparent surfaces. *Perception*, **28(suppl.)**, 137a.

Palmisano, S., Allison, R.S., & Howard, I.P. (2000). What is the problem with binocular correspondence? *Australian Journal of Psychology*, in press.

Solomon, F. (1987) *Probability and stochastic processes.* Englewood Cliffs, NJ: Prentice-Hall.

Rose, A. (1942). The relative sensitivities of television pick-up tubes, photographic film, and the human eye. *Proceedings of the Institute of Radio Engineering*, **30**, 295-300.

Rose, A. (1948). The sensitivity and performance of the human eye on an absolute scale. *Journal of the Optical Society of America*, **38**, 196-208.

Tyler, C. (1974). Depth perception in disparity gratings. *Nature*, **251**, 140-142.

Wilcox, L.M., & Hess, R.F. (1996). Is the site of non-linear filtering in stereopsis before or after binocular combination. *Vision Research*, **36(3)**, 391-399.