**Volume perception: disparity extraction and depth representation in complex three-dimensional environments**

Julie M. Harris

School of Psychology and Neuroscience, University of St. Andrews

Email: jh81@st-andrews.ac.uk

*Accepted for publication in Journal of Vision, 2014.*

## Abstract

Our sensitivity to binocular disparity is exquisite under the best conditions, typically in uncluttered scenes with few small objects. Yet binocular vision can deliver a very strong impression of depth for complex, cluttered scenes with lots of objects and overlaps. How good is disparity processing under these conditions? Here we explored a novel task: depth volume perception, to study how a global representation of depth is obtained using binocular disparity information. We found that the human visual system is sensitive to depth volume, but that the volume perceived is dependent on the local and global arrangement of scene content. We also show how a model of early disparity extraction and combination can account for some of the biases found. Our work shows that the visual system is not able to correctly represent and interpret disparity for all locations in a complex three-dimensional scene.

## Introduction

Binocular disparity can be an extremely sensitive cue to depth and shape. Under optimal conditions, we can perceive just a few seconds of arc of binocular disparity. The best conditions appear to be simple stimuli, where we view small disparate elements at an optimal separation (McKee, 1983) and for smooth surfaces, with relatively slow changes in depth across space (e.g. Tyler, 1974). When there are more elements present in a scene, and multiple depths, distorting effects occur, such as disparity attraction and repulsion (Westheimer & Levi, 1987). Perceived depth can also be reduced or enhanced by altering the position or connectedness of other scene elements (e.g. McKee, 1983, Deas & Wilcox, 2012). If we consider more complex environments, very little research has been conducted on the utility of binocular disparity. For example, if one looks into the foliage of a tree, there is a rich pattern of depth, with the trees and branches representing a dense volume, where leaves can be at many orientations and scattered through the volume in depth. The role of disparity in perceiving depth in such scenes has barely been touched on by current research.

There is only a small literature that has explored binocular disparity processing using stimuli representing depth profiles that are not single flat planes or smooth surfaces. Perhaps the best of these is on disparity transparency, where a pair of flat surfaces are typically defined by dots, both surfaces being frontoparallel, and one closer to the observer than the other. Akerstrom and Todd (1988) used a rating procedure to decide whether a pair of disparity-

defined surfaces appeared as two separate surfaces. They found that apparent surface separation falls with increasing disparity or dot density, and that it increases with viewing time. Parker and Yang (1989) measured the disparity at which the transparency percept was replaced by that of a single surface, perceived at the average depth between the disparity-defined surfaces. Gephstein and Cooperman (1998) established that, for higher dot densities, transparency could only be perceived veridically for smaller depth separations. McKee and Verghese (2002) also explored stereo transparency for surfaces composed of pairs of dots on a front and a back plane. They found that apparent depth reduces as the local disparity gradient increases. Further, stereoscopically transparent surfaces produce higher discrimination thresholds than stimuli delivering a single opaque surface (Wallace and Mamassian, 2004). More recently, Tsirlin, Allison and Wilcox (2012) have discovered that segregation of two surfaces is easier if the front plane contains more dots than the rear plane.

Clearly, variables like the element density and spatial arrangement have an effect on performance in these studies. We will suggest here that some of these effects might be explainable as limited via the basic processes that extract and then combine disparity to form a depth percept.

A recent model of early disparity processing, that encapsulates a number of aspects of human performance, comprises a population of local cross-correlators that compare information between the two eyes' views (Banks et al, 2004; Neinborg et al, 2004; Filippini & Banks, 2009). This model has been refined to exploit the 'size-disparity-correlation', the idea that small disparities are processed by mechanisms with small spatial extents (Allenmark & Read, 2010, 2011, see also Smallman & MacLeod, 1994; Tyler, 1973, 1975). Thus the current understanding is that human stereo discrimination is carried out by an array of cross-correlators, using a range of different correlation window sizes, used to process different magnitudes of binocular disparity.

Models like these have not yet been applied to stimuli representing complex patterns of binocular disparity. There is potentially a problem with applying such models to complex depth environments, because a correlation-like extraction of disparity is ideally suited to representing patches of the world as locally flat regions with a specific binocular disparity. Correlators will fail when scenes contain locally very different disparities. We reasoned that the scale and range of disparity cross-correlation could limit human performance for stimuli containing complex depth profiles, for tasks requiring use of a broad range of binocular disparities. Here we sought to explore how human observers perceive depth in a complex 3D environment, and relate performance to the early stages of disparity extraction, via a computational model based on disparity cross-correlation. We also considered the next stage of processing: what rules are used to combine those extracted disparities for a specific task involving a population of depth elements.

In this study, we investigated volume perception, the ability of human observers to judge the thickness of a volume defined by a number of thin lines, of varying orientation, and a range of depths. Judging the thickness of a volume requires the processing of depths of a number of separate elements, and we can manipulate the spatial and depth arrangement across a range of parameters (number of elements, global depth arrangement, local disparity gradient). Only a small number of preliminary studies have specifically studied depth volume (Harris et al, 2013, Goutcher, O'Kane & Wilcox, 2012; Keeble, Harris & Pacey, 2006). We conducted three different stimulus manipulations. In the first, we manipulated the shape of the volume distribution itself. In the second, the number of elements representing the depth was varied. In the third, we altered local disparity gradient, the ratio of spatial separation to depth separation between elements. All three of these manipulations could have an impact on perceived depth volume if there are processing limitations on how disparity is extracted or combined. Our aim was to test the limits of

disparity extraction and explore the subsequent disparity combination needed to achieve volume perception.

A correlation-based model, akin to previous modelling (e.g. Allenmark & Read, 2011, Goutcher & Hibbard, 2014) was developed here, where we could manipulate the size of the correlation window to explore model performance across various ranges of depths and spatial scales. To summarise what we discovered, we found that human volume perception was dependent on the disparity arrangement, and the local spatial arrangement of elements within scenes. Our models were able to emulate human performance for most of our experimental manipulations, and demonstrated that some effects are limited by early disparity extraction, and others by later depth combination.

## Methods

### Apparatus

Stereoscopic images were presented on alternate video frames (120 Hz frame rate) on a Iiyama Visionmaster Pro 21inch CRT monitor, run from an Apple G3. Crystaleyes active shutter goggles were use to display alternate video frames to the right and left eye. We restricted stimulus display to the red video gun to reduce the effects of cross-talk between video frames (which was unmeasureably small on our display using a Minolta LS-110 photometer). Responses were recorded by asking observers to press left or right arrow keys on a standard Apple Mac keyboard. Viewing distance to the display screen was 1.56m.

### Observers

Ethical approval was given by the University Teaching and Research Ethics Committee (UTREC), University of St. Andrews. Observers were recruited using the online participant recruitment system (SONA) at the School of Psychology and Neuroscience, University of St. Andrews, and via poster advertisements displayed around the University. Observers were compensated for their participation and gave informed consent prior to taking part in experiments. Ethical arrangements adhered to the Declaration of Helsinki.

All observers were naïve to the purpose of the experiment, and were given one session of training, where stimuli could be viewed for longer than the 2 seconds used in the experiments. During training, observers were informed whether their responses were correct or incorrect. No feedback was given during the main experiments. Observers were screened for normal stereopsis by using the TNO test. Any observers who were unable to see correct depth of 120 min arc disparity or larger on the test were excluded from further study.

### Stimuli: General

Stimuli were composed of a number of lines elements, 1.42 min arc wide and of a length of either 14.2 or 42.6 min arc, presented in a square region of size 2.76 x 2.76 deg. Element luminance was 20.1 cd/$m^2$ (measured through the active stereogoggles), presented on a dark background (1.6 cd/$m^2$, as measured through the active stereogoggles). Each element was presented at a specific depth location (i.e. there was no variation in disparity-defined depth within a single element: it did not slant or tilt in depth), and was given a random orientation between -45 and 45 deg (where 0 deg was defined as vertical). The location of each element was defined according to the specific experiment. Binocular disparity was added to the display by shifting elements by an equal and opposite amount in the right and left eye views. Figure 1a shows a cartoon illustrating the 3-D form of the stimuli and figure 1b depicts the stimuli via right and left stereo half-images.

For all experiments, a pair of stimuli were presented, side by side on the screen, each of which could have a different physically defined depth volume. One was always a test stimulus, the other a standard stimulus. There were 7 test stimuli, each comprised a pair of planes (where half the elements were presented on one plane, half on the other), with a disparity separation between the planes ranging from 2.84 to 19.88 min arc. In the test stimuli, element length was fixed at 14.2 min arc. There were 50 elements in each test stimulus. These values were chosen so that the test stimulus had enough elements to clearly define two planes, whilst having few elements close to one another or overlapping, to reduce possible violations of the disparity gradient limit.

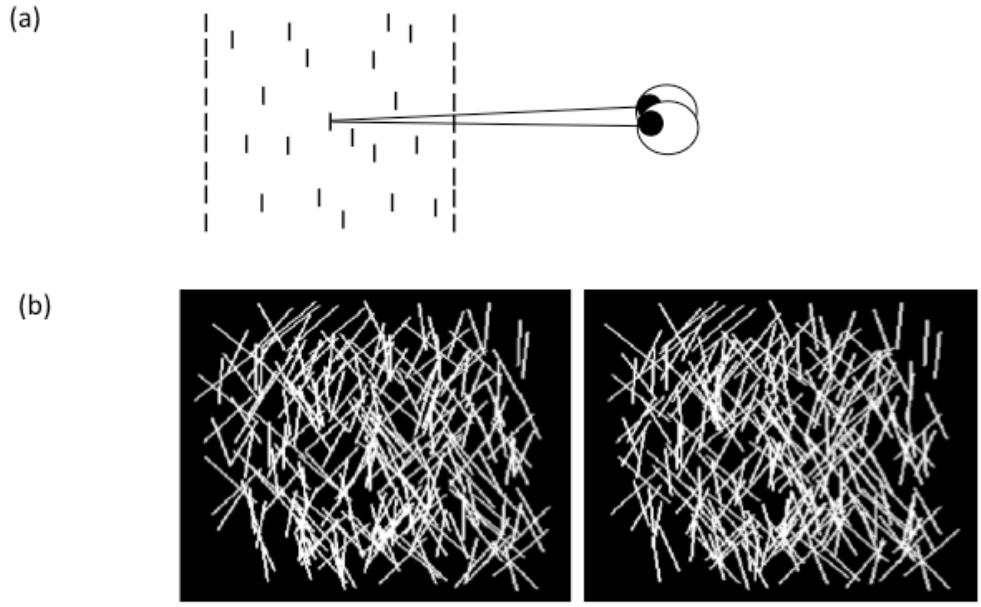



Figure 1. (a) Cartoon showing the form of the 'volume' stimulus. Observer views elements within a volume, 25% on rear surface, 25% on front surface and 50% at random positions in between. (b) left and right eye stereo-halves depicting what stimuli looked like (in the actual experiments we used the red-gun of the CRT so stimuli were presented in shades of red, not black and white).

**Standard Stimuli: Experiment 1 volume**

In the first experiment, we compared perception of the depth separation between a pair of planes with that from a volume. The test stimulus was a pair of planes, as described above. In condition 1, the standard stimulus was also a pair of planes, with a disparity separation of 11.4 min arc. Half of the 250 elements were presented on one plane, and half on the other, each was 42.6 min arc long. In condition 2, the standard stimulus comprised 250 elements spread throughout a volume. We wanted to constrain the volume, so that all elements had depths less than those in the pair-of-planes test stimuli. Thus, 25% of the elements were presented on a back plane, 25% on a near plane, with a disparity separation of 11.4 min arc between the planes. The other 50% were assigned depths at random, anywhere between the front and back planes.

**Standard stimuli: Experiment 2 element number**

In the second experiment, we compared perception of the depth separation between a pair of planes with that from either a volume, or a pair of planes, for different numbers of elements within the volume. The test stimulus was always a pair of planes, as described above. Here, the standard volume stimulus was comprised either of 250 or 50 elements (each 42.6 min arc long), spread throughout the volume. As before, 25% of the elements were presented on a back plane, 25% on a near plane, with a disparity separation of 11.4 min arc between the planes. The other 50% were assigned depth at random, anywhere between the front and back planes, drawn from a uniform distribution. The standard planes stimulus was a pair of planes, and could have either 50 or 250 elements. For the 250 element conditions, we used the data from Experiment 1, hence this experiment required us to collect fresh data only for the 50 element conditions.

**Standard stimuli: Experiment 3 disparity gradient**

In the third experiment, we compared perception of the depth separation between a pair of planes with that from a pair of planes that had high or low disparity gradients, where disparity gradient is defined as the disparity difference between a pair of points, divided by their lateral angular separation. The test stimulus was a pair of planes, as described above. Condition 1 was the low disparity gradient condition. Here, the standard stimulus was a pair of planes, with a disparity separation of 11.4 min arc. Half of the 250 elements (42.6 min arc long) were presented on one plane, and half on the other. Pairs of elements were generated with the same centre location but with different random orientations. In this condition each element in the pair was given the *same* disparity, so that the pair sat either on the back or the front plane. The disparity difference between each point on each element was zero, hence the disparity gradient between each element pair was zero. Of course, disparity gradients between each element and elements on the other plane, which could be nearby, were non-zero, but because 50% of the disparity gradients were zero, we define this as a 'low' disparity gradient stimulus.

Condition 2 was the high disparity gradient condition. Again, half of the 250 elements were presented on one plane, and half on the other. Elements were again generated with random orientation, and were paired so that every two elements had the same centre location. In this condition each element in the pair was given the *opposite* disparity so that one sat on the front plane and one on the back plane. Thus, points on each element were close in their lateral position but with a large disparity separation, and therefore had high disparity gradient. Hence we defined this the 'high' disparity gradient condition. We did not estimate the population of specific disparity gradients; for our purposes it is enough that the range of disparity gradients is substantially different in the two conditions. Further, the disparity gradients between pairs of elements would frequently violate the disparity gradient limit (where the ratio of disparity/distance is greater than 1, resulting in local diplopia, eg see Burt & Julesz, 1980).

**Procedure and Data Analysis**

The procedure was the same for all three experiments. Observers viewed two stimuli, side by side on the computer screen for 2s. One of these was one of the test stimuli and the other was the standard. In each trial, observers were asked to indicate whether the left or right stimulus had the greater volume, or depth thickness. After the stimuli were viewed, a fixation cross appeared in the centre of the display, at zero disparity, and remained in view until the observer pressed the appropriate response button. Then the next trial was initiated. No feedback was given.

For each of the experiments the stimuli from the two conditions were randomly interleaved. Observers viewed a total of 420 trials per experiment, consisting of 30 repeats of the 7 test stimuli for each of the two conditions. Blocks of trials were typically presented across 3 runs of the experiment, each self-paced, but usually lasting around 10-15 minutes. Observers could rest whenever they wished during a run, and for as long as they wished between runs.

We recorded the proportion of occasions on which observers perceived the test stimulus as deeper than the standard stimulus, as a function of the depth of the front test plane (eg. for planes separated by 11.4 min arc, the front test plane was at 5.2 min arc in front of the plane of the screen). A cumulative normal psychometric function was fit to each data set (using Psignifit; Frund, Haenel & Wichman, 2011). The point of subjective equality (disparity corresponding to 50% deeper responses) was obtained (see figure 2a for an example) and a 95% confidence interval was calculated for each PSE, based on the fitted psychometric function.

**Results**

**Experiment 1: volume**

10 observers took part in this experiment. Figure 2a shows a sample data set and corresponding psychometric function from one observer, plotting percent test thicker as a function of the disparity of each plane with respect to the central screen plane. Red triangles show data from the volume condition, blue squares data from the planes condition. The PSE of the psychometric function would be 5.7 min arc if there were no bias between the perception of test and standard stimuli. Notice that the fitted psychometric functions are offset for the two conditions, producing a smaller PSE for the volume condition (red triangles) compared with the planes condition (blue squares).

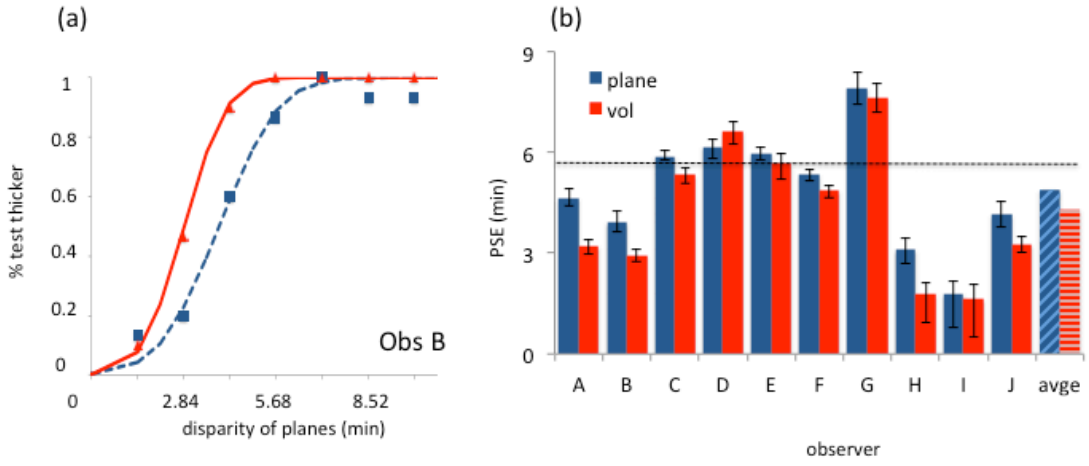



Figure 2. (a) Fitted psychometric functions for planes (blue dotted squares) and volume (red solid diamonds) for one observer. Disparities on the x-axis are those between each plane and the central fixation plane. (b) Points of subjective equality (PSE) for all observers for planes (blue) and volume (red) conditions. Average data from all observers is shown the rightmost pair of bars. Errors bars show 95% confidence intervals. The horizontal dotted line shows the expected PSE if there were no bias.

Figure 2b shows PSE's for the two conditions across all observers. PSE's for the volume condition (red bars) are typically lower than for the planes condition (blue bars). This is a small effect for each observer, sometimes falling within the 95% confidence intervals, but across all observers, a 2-tailed t-test revealed the difference to be significant ($t(9)=3.298$, $p=0.009$). The mean difference between PSE's (4.9 min for the planes condition, 4.3 min for the volume condition, rightmost columns in figure 2b) indicated that a smaller depth thickness was perceived in the volume condition than in the planes condition.

The dotted horizontal line shows the expected PSE if there were no bias between test and standard stimuli. PSE's for both conditions tended to be biased away from this 5.7 min standard, but the bias was not significantly different from the 5.7 min 'no bias' line for either the planes condition ($t(9)=1.175$, $p=0.27$), or the volume condition ($t(9)=1.79$, $p=0.11$). In this experiment, standard stimuli contained only 50 elements, compared with 250 in each test stimulus, and we suspected that this might be the cause of any bias. We explored this potential effect of dot number in more detail in the second experiment.

**Experiment 2: element number**

The same 10 observers that took part in Experiment 1, also took part in this experiment. Here we used volume stimuli to compare volume perception, using both plane and volume stimuli, when stimuli contained 50 or 250 elements. Figure 3a shows PSE's for the 50 versus 250 element conditions across all observers, for volume stimuli, and figure 3b for plane stimuli.

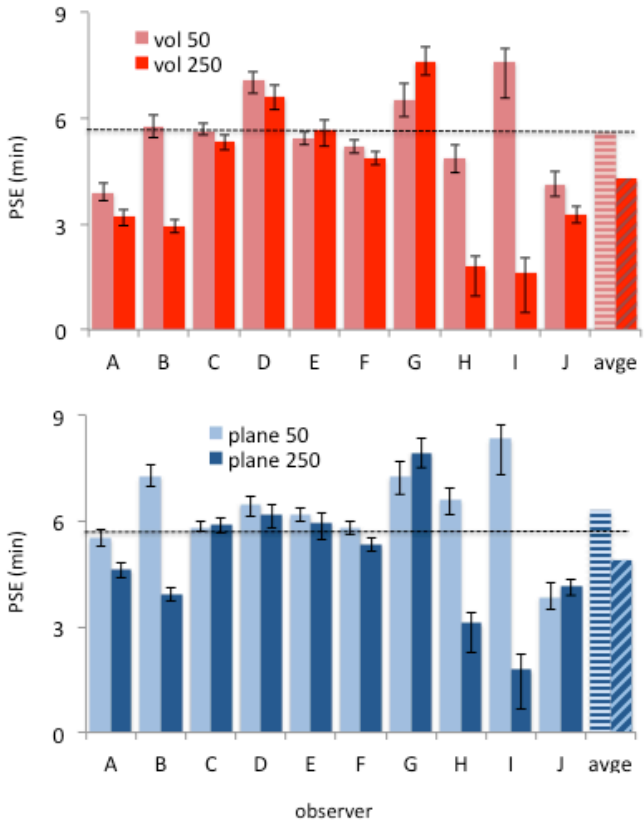



Figure 3. (a) Points of subjective equality (PSE) for all observers for volume stimuli with 250 (dark red) or 50 (pink) elements. (b) Points of subjective equality (PSE) for all observers for planestimuli with 250 (dark blue) or 50 (light blue) elements. For both graphs, Average data from all observers is shown by the rightmost pair of bars. Errors bars show 95% confidence intervals. The horizontal dotted line shows the expected PSE if there were no bias.

The dotted horizontal line shows the expected PSE if there were no bias between test and standard stimuli. For volume, PSE's for the 250 element condition (red bars) tend to be lower than for the 50 element condition (pink bars). The pattern of data, however, is somewhat different from Experiment 1. Here, there is a large effect for some observers, and almost none for others, with one observer showing a bias in the opposite direction. A similar pattern of data was found for the plane stimuli (figure 3b). The results of a two-way repeated measures ANOVA (element number and element distribution as factors) showed no interaction between distribution and number of elements ($F(1,9)=0.36$, $p=0.56$). There was a significant main effect of distribution ($F(1,9)=10.5$, $p=0.01$), but not element number ($F(1,9)=3.96$, $p=0.078$). The consistent (but not significant) difference in average PSE when stimuli contained 50 elements, compared with 250, suggests that there is a trend for smaller depth thickness to be perceived in the 250 elements conditions than in the 50 elements conditions. This trend was also suggested in the results of Experiment 1, where both mean PSE's tended to be (but also not significantly) below the zero bias line.

**Experiment 3: disparity gradient**

10 observers took part in this experiment, 3 of whom also participated in Experiments 1 and 2. Here we compared performance for stimuli that each consisted of a pair of planes, but where one stimulus had element pairs with high disparity gradients, and the other stimulus pairs with low disparity gradients. Figure 4 shows PSE's for the low disparity gradient condition (light purple bars) and the high disparity gradient condition (dark purple bars).

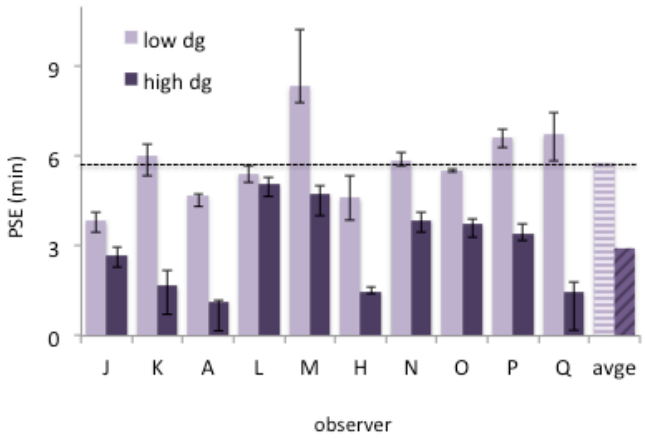



Figure 4. Points of subjective equality (PSE) for all observers for plane stimuli with low disparity gradient (pale purple) or high disparity gradient (dark purple) elements. Average data from all observers is shown the rightmost pair of bars. Errors bars show 95% confidence intervals. The horizontal dotted line shows the expected PSE if there were no bias.

The dotted horizontal line shows the expected PSE if there were no bias between test and standard stimuli. For all observers, PSE's were very much lower when disparity gradients were high than when they were low, and this effect was highly significant ($t(9)=5.991$ $p=0.0002$). The difference in average PSE (2.9 min for high condition, 5.8 min for low

condition, rightmost columns in figure 4) indicates that the stimuli containing more frequent high disparity gradients were perceived as being less deep.

## Interim Discussion

Very few studies have explored depth volume perception. The 3 experiments that we describe have all demonstrated that the perception of depth volume can be biased by specific stimulus arrangement. When the stimulus portrayed different distributions of elements (Experiment 1), volume perception was systematically different for different distributions, even though the outer extremities of the depth distributions were at identical positions in depth (recall that 25% of elements were located on the front or rear planes, respectively). Further, for a specific distribution of depth elements, the perceived depth volume tended to be biased by varying the number of elements (Experiment 2). However, this was a rather idiosyncratic effect, with really large biases in some observers, and no bias, or opposite bias for others. Finally, as has been found in a similar study on depth transparency (McKee & Verghese, 2002), the volume perceived when viewing a pair of planes containing the same depths, but different local disparity gradients, were very different, with substantial depth compression found for high disparity gradients (Experiment 3).

As discussed in the Introduction, these kinds of effects might be expected, based on the supposition that the early stages of binocular disparity extraction involve processing akin to local cross-correlation between right and left eye views. To explore whether this intuition is correct, we implemented a disparity extraction model, at a range of spatial scales, to explore whether similar biases in perception would be found. Our model used cross-correlation to extract disparity and a simple decision stage to combine and use that disparity information for volume perception.

## Cross-Correlation Model

We based our model on recent models of disparity extraction that have used a population of local cross-correlators that compare information between the two eyes' views (Goutcher & Hibbard, 2014; Banks et al, 2004; Neinborg et al, 2004; Filippini & Banks, 2009, see also Ohzawa et al, 1990). This style of model has been primarily used to explore stereo-resolution, the smallest spatial variations in depth that can be discriminated, and thus the focus has been on finding the smallest useful correlation window. Recently, data showing that human observers can perceive depth from sinusoidal oscillations as veridically as those from square wave oscillations (Allenmark & Read, 2010) has provided a challenge to the original model (Banks et al, 2004). A modified model, where larger correlation windows are used to detect larger binocular disparities, which was developed by Allenmark and Read (2011), exploiting the so-called 'size-disparity-correlation' (Smallman & MacLeod, 1994), can better account for human performance. Smallman and MacLeod found that optimal disparities for high spatial frequency information were small, and for low spatial frequency were larger, leading to the idea that large-scale (low spatial frequency) receptive fields are involved in processing large disparities, and small-scale (high spatial frequency) receptive fields are involved in processing small disparities (eg. see Harris et al, 1997, Filippini & Banks, 2009).

Here we made the assumption that large correlation windows would specialise in processing low spatial frequency information (and vice versa for small windows). Note that we implement this assumption in a different way to Allenmark & Read (2011).

We started with a pair of images containing a blank background, of intensity 0.5, and disparate line elements, of intensity 1, generated in the same way as for our experiments. Images were 230*230 pixels, and disparity separation between the bounding planes was 16 pixels. Here we chose to implement a simple cross-correlation model that contains some of the key features from these other models. Some models (Allenmark & Read, 2010, 2011; Filippini & Banks, 2009) have aimed to emulate the front-end of the visual system as closely as possible, by filtering images to account for the eye's optics, as they were used to primarily address questions about the limits of stereo-resolution. Goutcher & Hibbard (2014) used a correlation-based disparity extraction model to study depth perception from ambiguous random dot stereograms. To model their data, they required an initial spatial frequency filtering stage, followed by cross-correlation. We adopted that idea here, combining an initial spatial frequency filtering stage with a constraint that we used larger cross-correlation windows for the lower ranges of spatial frequency, in line with the size-disparity correlation. Goutcher & Hibbard (2014) used a single, large, correlation window size in their model. But their stimuli contained constant disparity across the scene, making a large correlation window optimal. That clearly would not be optimal for our volume stimuli, containing elements with many different depths.

We implemented bandpass filtering of our images, $I_L(x,y)$, in the same way as Goutcher & Hibbard (2014). We conducted bandpass filtering using a filter of bandwidth *b*, around a central spatial frequency of $f_c$. To achieve this, the Fourier amplitude spectrum of the image was multiplied by a mask:

$$M(f) = 1\,, f > \frac{f_c}{2^b}\ \&\ f < 2^b f_c \qquad (1)$$

= 0, *otherwise*

Here we chose the bandwidth, b, to be ±1 octave. For each spatial scale that we explored, a window size was chosen for the cross-correlation (wsize). We next needed to choose a relationship between window size and the frequency band being explored. To fit with the size-disparity-correlation, there should be an inverse relationship between centre frequency and window size The central frequency of the bandpass filter was chosen to have one of four relationships: f=0.25/wsize, f=0.5/wsize, f=1/wsize or f=2/wsize. In the results section, we discuss the implications of these choices.

Examples of these pre-filtered images are shown in figure 5a. At this point we also added some random monocular noise independently to the right and left images, adding random luminance noise of intensity 0.001. We then ran a cross-correlator of window size *wsize*, across the left and right images, for each pixel value (x,y) within the images. The windows for each eye were at the same vertical position, but different horizontal positions. For each location in the image, (x,y), we held the left-eye window at that location. The correlation window, $L_w$, was defined as the set of image values $I_{Lf}(i,j)$ such that $|x-i|<wsize/2$ and $|y-j|<wsize/2$. The right-eye window was presented at the same vertical location as the left-eye window, but could have a horizontal offset, or disparity, *disp*. The correlation window for this eye, $R_w$, was defined as the set of image values $I_{Rf}(i,j)$ such that $|x+disp-i|<wsize/2$ and $|y-j|<wsize/2$.

The correlation, for any disparity *disp*, was then defined as:

$$C(y,disp) = \frac{\mathrm{cov}(L_W, R_W)}{\sqrt{\mathrm{cov}(L_W, L_W)\,\mathrm{cov}(R_W, R_W)}} \qquad (2)$$

We can think of the function C(y,*disp*) as representing the output of a set of disparity detectors, each centred at location (x,y), and disparity *disp*.

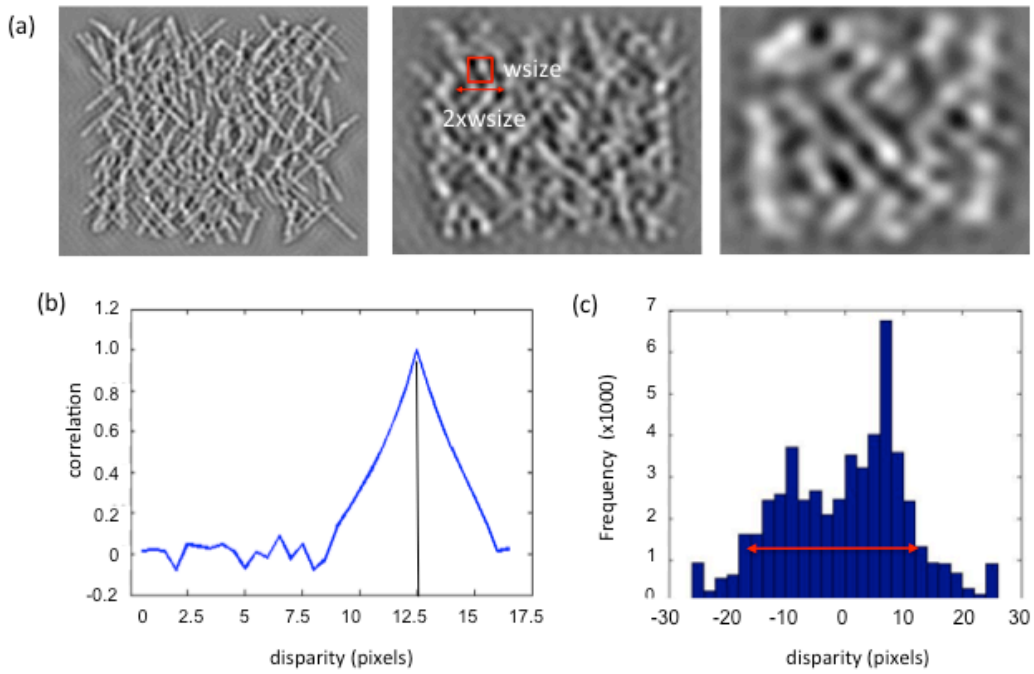



Figure 5. (a) Diagram illustrates three levels of pre-filtering of the image before cross-correlation at a specific scale. Red square illustrates an example correlation window and range. (b) Correlation as a function of disparity. The peak of the function is chosen as the disparity. (c) Histogram showing frequency of extracted disparities for all image locations. The standard deviation of the distribution is chosen as a measure of thickness of the volume of elements.

The next choice to make was what disparity range to choose. In line with the 'size-disparity correlation' described above, we chose a range proportional to the window size, in this case twice the window size. Hence, for any image location, (x,y), there were *2*wsize* correlations recorded, with possible disparity values from *-wsize* to *+wsize* (illustrated in middle panel of figure 5a).

We then used the output of the correlator to decide what disparity should be represented at each location (x,y). Figure 5b illustrates an example, showing correlation as a function of disparity across the range of disparities chosen. We found the peak of this correlation function, and chose the disparity corresponding to the peak to represent the disparity between left and right images at location (x,y). This is akin to choosing the peak response from a population of disparity-tuned neurons.

Our aim here was to represent disparity across the whole image, so we repeated the correlation process for all locations (x,y). The family of disparities produced can be presented as a histogram showing the frequency of occurrence of each disparity. Figure 5c shows an example histogram for a stimulus composed of a pair of planes, located at disparities -8 and +8 pixels. Because of the restricted window size, range and filtering, the histogram does not show disparities only at those values, but instead shows a broad range, with noticeable peaks at -8 and +8.

Finally, the overall goal was to model how human observers might represent the depth of the volume of elements. To do this one must choose a decision rule to be implemented in

the model. In other words, how is the distribution of disparities, like those in figure 5c, used to decide which of two stimuli have the deeper volume? We chose the simplest possible decision rule that avoids prior knowledge of the actual image disparity distribution, by recording the thickness of the distribution as twice the standard deviation of the distribution of disparities delivered by the disparity-extraction stage of the model (arrowed line in figure 5c).

**Model results**

As a start-point, we considered what results we would expect for a system that could achieve 'perfect' disparity matching and extraction. We considered an 'ideal' model that delivered the disparity, for each element that was specified in the stimulus. Further, we assumed that only a single disparity sample was delivered for each element (akin to ideal-observer models of disparity processing see, Harris & Parker, 1992, 1994a, b, 1995). For our planes stimulus, such a simple model would deliver 50% of elements with 8 pixels of disparity, and 50% with -8 pixels. Using the 'thickness' decision rule described above, this will deliver a standard deviation of 8 and thus a thickness of 16 pixels.

Our first aim was to explore how the correlation model behaved as the relationship between band-pass centre frequency and correlation window size was altered. As explained above, we tested a range of different relationships between centre-frequency and window size. To equate with what our human participants saw, the frequencies used, converted to cycles per degree, ranged from 0.54 to 18.5 cpd. Specific details of parameters chosen can be found in the Supplementary Materials 1.

We used four window sizes to emulate different spatial scales. Because the stimuli contained disparities ranging from -8 to +8 pixels, the smallest window chosen was 9 pixels (range of 18). Windows used were 9, 19, 29 and 39 pixels with respective ranges 18, 38, 58 and 78 pixels. Clearly, during the neural processing of disparity-defined volume, at some stage information from these different scales must be combined. There have been several suggestions for how this may be done (eg. Tsai & Victor, 2003). We do not consider that issue here because our experiments were not intended to address this point. Instead, we calculate our proxy for depth volume, the thickness of the extracted-disparity distribution, separately for each scale.

We ran the model 100 times (50 times for the high-low gradient condition as differences between conditions were very large) for each combination of frequency-window ratio, window size (and range), and stimulus type, with a different random set of image elements for each run. This allowed us to measure the thickness of the disparity distribution. We plot model results in pixels, as we are interested in the relative performance across stimulus types.

We started with the 2-planes stimulus, where we expected our model to do a good job of extracting disparity distributions close to -8 and +8 pixels, and delivering widths close to 16 pixels. Figure 6a shows a sample disparity distribution (for 250 elements, wsize=29, f=1/wsize), showing that most extracted disparities were grouped around the +- 8 pixel disparity values in the stimulus. Figure 6b shows how the extracted thickness of the disparity distribution varied as a function of frequency (cycles per image), for the 4 different correlation window sizes. For the smallest window size (purple line), the model delivered an answer close to veridical for all frequencies. This is perhaps not surprising as this window size did not measure disparities larger than 9 pixels. For larger window sizes, the model performance was tuned as frequency varied. This means that some combinations of frequency and scale deliver an over-estimate of thickness. Larger windows will have many elements falling within them, each contributing to the disparity extraction and potentially

causing many ‘false matches’, resulting in many extracted disparities at the ‘wrong’ pixel values.

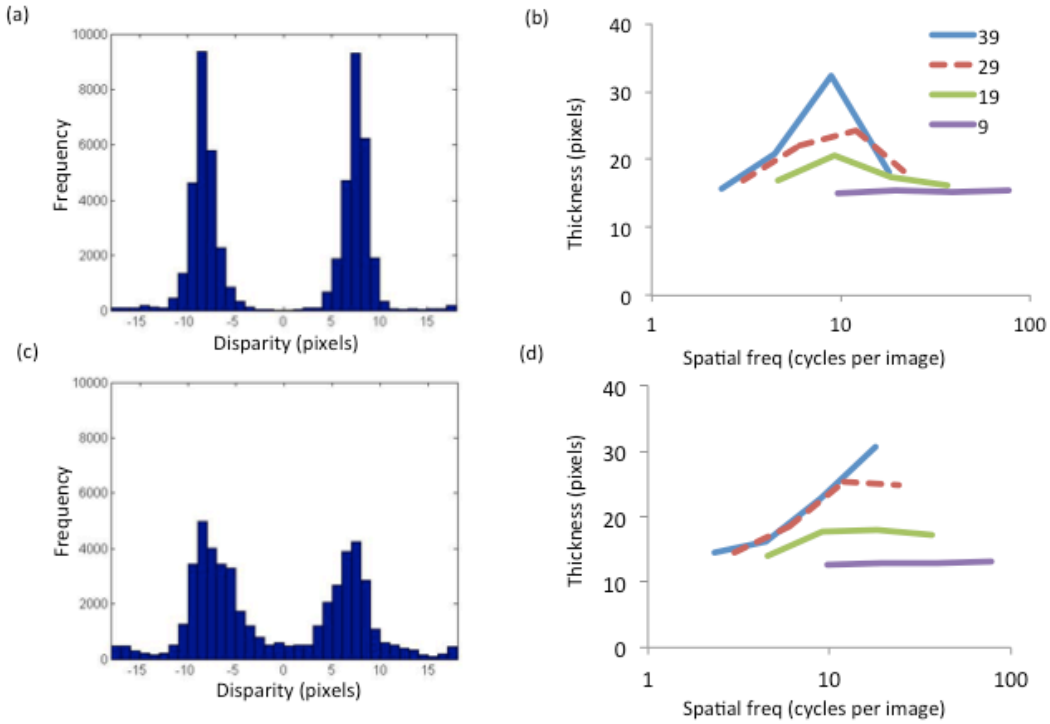



Figure 6. (a) Frequency of extracted disparities for planes stimulus, for f=1/w relationship, and a correlation window size of 19 pixels. (b) Extracted thickness of the disparity distribution for plane stimulus, as a function of spatial frequency (cycles per image), for 4 window sizes. (c) Frequency of extracted disparities for volume stimulus, same parameters as in (a). (d) Extracted thickness of disparity distribution for volume stimulus.

Figures 6c and 6d show the equivalent plots for the volume stimulus. There is a broader spread of extracted disparities now, as one would expect if there were ideal disparity-extraction, because there is a different spread of disparities in the stimulus itself. Notice here (figure 6d) that for the combination of larger window sizes and higher frequencies, the extracted thickness becomes consistently higher.

We next explored model performance for the specific stimulus conditions used in the experiments. It is important to note that we were not seeking to fit models explicitly to human performance. The key here was to look for similar patterns of performance error to those found for our human observers, without developing a complex model with many parameters. We first explored the stimuli used in Experiment 1, where we compared the perceived thickness of a pair of planes with that of a filled volume. We used the same number of elements, the same element lengths and locations, and the same depth separations. Figure 7a shows mean model-generated width of the disparity distributions, as a function of correlation window size, for the condition where f=0.5/w (the full set of data, for each f-w relationship can be found in the Supplementary Materials 2). Notice that error bars (SEM) are very small, due to the large number of model runs. Model generated thickness tended to increase as a function of window size, but there was a consistent difference between the planes and the volume condition, with thickness always smaller for the volume condition (red lines) than for the plane condition (blue dotted lines). For every window width, the difference between the 2 conditions was highly significant (window widths of 9 ($t(99)= 83.7$, $p<0.00001$) 19 ($t(99)=21.8$, $p<0.00001$), 29 ($t(99)=10.7$, $p<0.00001$), and 39 ($t(99)=7.2$, $p<0.17$). This result is consistent with what we found from our human observers in Experiment 1. This pattern occurred for the lower frequency ranges (f=0.25/w, f=0.5/w) but

not for the higher ones (see Supplementary Materials 2), suggesting that human observers may be relying on lower spatial frequency channels to perform this task.

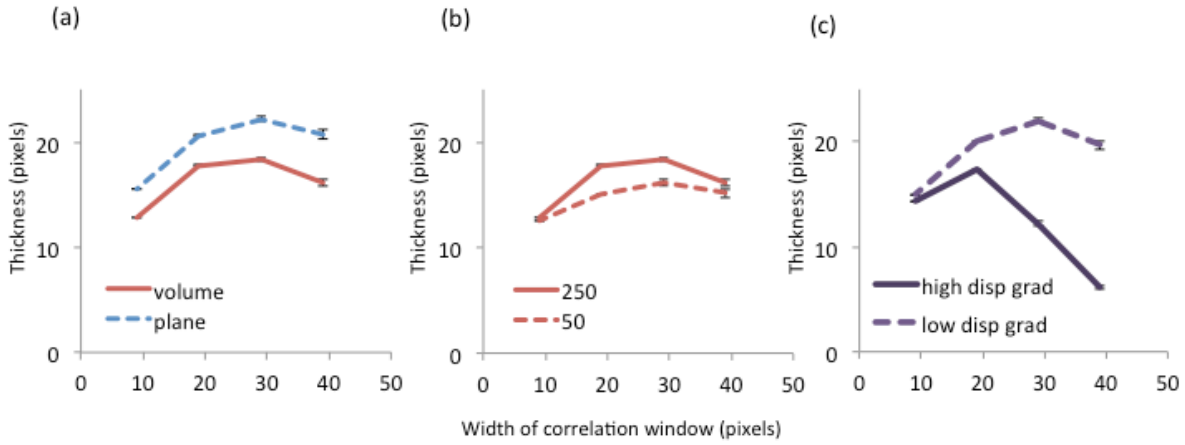



Figure 7. Output of correlation model for f=0.5/w. (a) Volume versus plane stimuli, (b) volume stimuli 50 versus 250 elements, (c) plane stimuli, high or low disparity gradients. Error bars shown standard error of the mean.

Next, we compared performance for the volume condition when there were 250 elements in the scene, and when there were 50, as for Experiment 2. This time we found a small difference between model performance across these two stimulus conditions when f=0.5/w (figure 7b), with perceived depth volume larger for the 250 element stimulus (red solid lines) than for the 50 element stimulus (red dotted lines). Thickness for the 250-element condition was significantly different (Bonferoni corrected p-values shown) from the 50-element condition for window widths of 9 ($t(99)=6.98$, $p<0.00001$) 19 ($t(99)=18.04$, $p<0.00001$) and 29 ($t(99)=6.1$, $p<0.00001$), but not for 39 ($t(99)=1.75$, $p=0.17$) pixels. This behaviour was found consistently across a range of window sizes and frequencies (see Supplementary Materials 3), though for some ranges there was very little difference between the 250 and 50 element results. The model behaviour here was unlike that found for most of our human observers. Recall, for the human observers there was a trend for PSE's to be lower when there were more stimulus elements, a result not found for the model under any of the conditions tested.

Finally, we ran the model on our two disparity-gradient manipulated stimuli, as were used in Experiment 3. As for our human participants, there were very large differences in width between the low- and high-disparity conditions (Figure 7c), especially for the larger correlation window sizes (see Supplementary Materials 4). For the high-disparity condition, much smaller thicknesses were recorded, particularly for the larger scales. For every window width, the difference between the 2 conditions was highly significant (window widths of 9 ($t(49)= 18.5$, $p<0.00001$) 19 ($t(49)=17.2$, $p<0.00001$), 29 ($t(49)=22.3$, $p<0.00001$), and 39 ($t(49)=21.9$, $p<0.00001$)). However, for f=1/wsize and f=2/wsize, there were conditions where similar thicknesses were obtained for low and high disparity gradient conditions,

suggesting that the human visual system may be relying on lower frequency information to perform the volume task (see Supplementary Materials 4).

Overall, the modelling delivered some promising results, showing similar patterns of performance to those found in two of the experiments (Experiment 1, Experiment 3). However, the modeling suggested that when the number of elements was varied, perceived depth might be a little higher for larger numbers of elements, in contrast to some of our human participants (Experiment 2).

## General Discussion

The main aim of our experiments was to explore the extent to which the perception of depth volume would be altered by manipulation of the depth content of the scene being viewed. What we know already about first-stage disparity extraction suggested that scene depth content might have an impact, but no one had previously tested a task like this and compared it to a model of disparity extraction and combination. Our experimental manipulations did have an impact on perceived depth volume. We compared volume perceived for a pair of planes in depth, compared with the same pair of planes, but with additional depth elements scattered through the volume between them (Experiment 1). We found that the volume stimulus was consistently perceived as having less depth than the pair of planes. This result was consistent with the results generated by a model based on disparity extraction via cross-correlation (compare figures 2 and 7a).

Observers also experienced differing volume perception when different numbers of elements were used to represent depth in the scene (Experiment 2). The results of this experiment were less conclusive. Some observers experienced a sharp fall in perceived volume when more elements were present, occasionally they experienced an increase in perceived volume, but many perceived no difference (figure 3). Differences were not found to be statistically significant for the group of observers tested, but figure 3 illustrates that for some individuals there were very large differences for different numbers of elements. The model delivered little difference as element density was increased, showing a small increase in perceived volume as element number increased, for larger spatial scales of correlation (figure 7b, Supplementary Materials 3).

When local disparity gradient was manipulated (Experiment 3) we found very large effects, such that stimuli with high local disparity gradients were perceived as having very much less volume than those with more low local disparity gradients (figure 4). Again the experimental result was consistent with those found using our cross-correlation model (figure 7c), particularly when the model was run using large-scale correlation windows and lower frequency ranges (Supplementary Materials 4). This suggests that larger scale disparity extraction is specifically playing a part in volume perception.

In sum, all of our experimental manipulations resulted in observers perceiving different thicknesses of volume, demonstrating that there are limits to how binocular disparity information can be used for tasks like this, which require disparity extraction from a complex scene, followed by combination of disparity signals to perform a more global task. Our model has thus gone some way to emulating human performance.

### Why do these effects on volume perception occur?

That our model shows the same trends as human observers, for two experiments, is gratifying. However, a key issue is to understand why. The model has two independent aspects. First, there is the disparity extraction stage, which is designed to emulate disparity extraction in early stages of cortical processing. Second, to use the extracted disparities to perceive volume, one has to invoke a decision rule: how those disparities are combined to

form a representation of volume. Either or both stages may be responsible for the model's resemblance to the human data.

### *The decision rule*

We can explore the decision rule alone, by considering the 'ideal disparity extraction' model referred to earlier. This model predicts thicknesses of 16 pixels for the plane stimuli, and 13.1 pixels for volume stimuli, simply because the disparity distributions were different in those two stimuli. This difference in performance emulates that of human observers and demonstrates that Experiment 1 allowed us to tap into the disparity combination stage. In this paper, we have not chosen to explore the nature of the chosen decision rule in further detail. There are many possible rules that could lead to similar results to the behaviour obtained with our decision rule, thus we are not asserting that our 'thickness of distribution' rule is the only one possible. Our data is consistent with such a rule, and suggests that the representation of volume relies on the properties of the disparity distribution that has been extracted. Other preliminary work (Goutcher et al, 2012) hints that such a decision rule is a reasonable choice.

### *Disparity extraction stage*

The ideal model predicts identical performance for the 50 versus 250 elements volume stimuli, and for the high and low disparity gradient stimuli. This is, of course, not surprising, the distribution of disparities injected into the stimulus in Experiment 2, was the same in both conditions (it was simply the number of samples that varied), and the number and distribution was identical in Experiment 3, here it was the local arrangement of disparity samples that was manipulated. Thus some of the effects that we found experimentally are not explainable by the decision stage alone, and these effects likely rely on the way in which disparity is extracted

Others have found that our depth perception can be limited by the disparity extraction stage. Banks et al, (2004) Filippini and Banks (2009), and Allenmark and Read (2011) showed that stereo-resolution is limited by the smallest correlator size (an idea suggested psychophysically by Harris et al, 1995). They explored surfaces corrugated in depth to different extents, and found that the basic assumption of correlation-based models, that local correlators extract fronto-parallel 'patches' of depth, can explain the limits of perception of fine depth corrugations. There is also a large literature on how disparities can locally apparently 'repel' or 'attract' one another to alter the perceived depth difference between them (for example Westheimer & Levi, 1987; Stevenson, Cormack & Shaw, 1991). Some of these effects might also be explainable at the disparity-extraction stage and should be modelled.

The specific issue of how disparity gradient might impact depth perception was tackled by McKee and Verghese (2002), when they explored the perception of transparency between a pair of planes in depth. Although their task was different to ours (they asked people to judge the separation between the planes, we asked about the thickness of the whole volume), they found, as we did here, that very high disparity gradients resulted in the apparent compression of depth. Similarly, Akerstrom and Todd (1988) found compressed depth for transparent depth displays of very high density. McKee and Verghese (2002) explored how their stimuli were interpreted by another successful model of disparity extraction, the Tsai-Victor model (Tsai & Victor, 2003). This is a model that relies on basic disparity-energy detectors (Ohzawa et al, 1990; Prince et al, 2002) and considers how these respond across a number of different spatial scales. McKee and Verghese found that whether the disparities were successfully extracted by the model, or specifically not extracted, under particular conditions, could account for human performance in their transparency task. Thus they suggested a disparity-extraction limitation on the perception of

stereoscopic transparency, as we have done here for the perception of thickness of a volume.

Next, we discuss the results of volume perception using different element densities. If the disparity of each element were extracted ideally, there would be no difference in thickness found when large or small numbers of elements were used. Our model delivered slightly larger thicknesses for higher density stimuli, than lower, for some conditions. Recall, this was different from the result that we found with human observers. For some human observers, there were dramatic differences in the amount of depth perceived when stimuli contained 50 elements, compared with 250, and the trend was opposite to that of the model. This is not consistent with the transparency work of McKee and Verghese, who found no difference in perceived depth of transparent planes using between 2 and 200 elements.

One possible explanation for why we found a performance difference experimentally, comes from the idea that the visual system may down-weight noisy estimates of perceptual variables compared with estimates that are more certain. Bayesian models of perception are a classic example (e.g Clarke & Yuille, 1990; Ernst & Banks, 1992). Such models often incorporate the idea of a 'prior', where incoming estimates of visual variables are combined with prior information about the likely statistics of the world. Models like this have been used to describe a number of binocular vision phenomena including binocular correspondence (Prince & Eagle, 2000) and motion in depth (Lages, 2006; Welchman et al, 2008). In our case, increasing element density in the stimuli could make the disparity-extraction stage noisier, broadening the correlation function from which each estimate of disparity is obtained. This assumption would simply increase the range of disparities extracted. However, if there is an inbuilt prior to expect small, or zero disparities (eg see Prince & Eagle, 2000, and Lages, 2006, Welchman et al, 2008), then noisier disparity estimates would tend to increase the influence of the prior, and result in smaller disparities being chosen, and thus a compression in the estimated thickness.

Other stimulus manipulations could also result in noisier estimates of disparity. For example, for the high disparity gradient stimulus, because adjacent elements were specified to have very different disparities, disparity correlators will provide a noisier estimate of disparity than for the low disparity gradient stimulus, and hence again a compression in estimated thickness for the former stimulus would result. Thus a Bayesian account could also predict the results found in Experiment 3.

There may also be other influences on the perceived thickness, caused by higher-level effects. Tsirlin, Allison and Wilcox (2012) recently found asymmetries in the number of dots needed on front and back planes to detect stereo transparency between two surfaces. They suggested that, for a pair of planes, the front surface signal is obtained from the dots that define it, but the back surface signal is obtained from the dots, but also the spaces between dots are interpolated as if they are part of the back surface. Thus there is effectively 'more' signal in the back plane that the front plane. This may also be related to a recent result (Schutz, 2012) showing the perceived numerosity is also higher for the rear, than the front plane. Another way of thinking about this idea is that higher-level mechanisms, perhaps linked to figure-ground segregation, are thought to be playing a part when we perceive transparent surfaces. Such a scheme is compelling, given that disparity interpolation is important in the perception of smooth surfaces (e.g. Yang & Blake, 1995; Wilcox, 1999; Wilcox & Duke, 2005). However, how interpolation would work for surfaces that are not smooth, such as our disparity volumes, is unknown. So we cannot predict the effects interpolation would have on our stimuli, and thus cannot pursue this issue further here. Exploring ideas around interpolation, for scenes composed of volumes of points, rather than smooth surfaces, might be a fruitful area for further research.

Finally, we should emphasise that in this work we have considered the extraction and utility of binocular disparity in isolation. Other work has considered how both motion and colour

contribute to the utility of disparity for tasks such as transparency. For example, Qian, Andersen & Adelson (1994) explored transparency from motion. They used a stimulus display similar in concept to the one we used for Experiment 3, except that they were studying motion transparency not disparity. They paired elements such that each pair moved in opposite directions. In such paired stimuli, motion transparency was not perceived, but it was for stimuli containing the same distribution of elements moving in opposite directions, but without local pairing. This is reminiscent of the dramatic apparent flattening of perceived thickness in our pair of disparity-defined planes in Experiment 3. Also of interest for our purposes, was that when Qian et al added binocular disparity to their motion-paired elements (so that they had different disparities), or made them different spatial frequencies, transparency reappeared. Thus, this work shows that there are interactions between disparity and motion systems, and between different scales, that our work has not addressed.

A related study, and perhaps more relevant to our own work, is that of Bradshaw and Cumming (1997). These authors explored fine-scale disparity processing using stimuli with very fine-scale disparity corrugation. They arranged for depth corrugations to be so fine that they were beyond the stereoresolution limit, and transparency was not perceived. If elements on the rear plane were given different directions of motion compared with those on the near plane, observers then could perceive transparency. Bradshaw and Cumming (1997) suggested that this provided evidence that disparity and motion mechanisms must interact early in the visual hierarchy. If this is the case, then our work, and that of others studying disparity extraction in isolation from other cues, may present a special case, applicable to disparity extraction only under static conditions.

## Summary and conclusions

Complex depth environments, like the tree example described earlier, provide compelling sensations of depth that seem to be enhanced by binocular disparity. Yet, to our knowledge there has been very little research on volume perception, though a few studies have used other stimuli and tasks to explore on complex depth environments. Here we studied volume perception in scenes containing elements at many depths. We found that both the local and global arrangement of elements in a scene can affect its global perception of depth, at least for our volume perception task. Disparity extraction appears to limit the relatively local effect of disparity gradient manipulations. Further work will be needed to explore in detail the exact rules by which extracted depths are combined, and these rules are likely to be highly task-specific.

Finally, this work has implications for the generation and use of binocular disparity content. Most of what we currently know about binocular disparity processing has been acquired using simple stimuli depicting scenes with few objects and smooth disparity distributions. Much of this knowledge will not apply to more complicated 3D environments. Our work moves towards more ecologically relevant stimuli that challenge many of the assumptions of recent cross-correlation models. It is critical to understand that, for complex scenes containing many scene elements at many different depths, the early stages of visual processing, where disparity is initially extracted, places limits on what depths, or ranges of depths can be perceived. Thus, depth is not necessarily veridically represented for every element within a scene. This can have an impact on both the local and global perception of depth within that scene.

## Acknowledgements

Much of the data was gathered by undergraduate research assistants Bryony Croucher, Sarune Savickaite and Emma Thomson. We thank Paul Hibbard for allowing us to explore and use his code for band-pass filtering disparity models, and for ideas for interpretation of our results. Thanks to Olivier Penacchio, Philip Cammack and Andy MacKenzie for reading early versions of this manuscript. The work was supported by a Leverhulme Research Fellowship, and a grant from the Engineering and Physical Sciences Research Council (EPSRC) Grant EP/G038708/1.

**References**

Akerstrom, R. A., & Todd, J. T. (1988). The perception of stereoscopic transparency. Perception and Psychophysics, 44(5), 421–432.

Allenmark F, Read JC (2010) Detectability of sine- versus square-wave disparity gratings: A challenge for current models of depth perception. J Vis 10: 1–16.

Allenmark F, Read JCA (2011) Spatial Stereoresolution for Depth Corrugations May Be Set in Primary Visual Cortex. PLoS Comput Biol 7(8): e1002142. doi:10.1371/journal.pcbi.1002142

Banks, M. S., Gepshtein, S., & Landy, M. S. (2004). Why is spatial stereoresolution so low? Journal of Neuroscience, 24, 2077–2089.

Bradshaw, MF; Cumming, BG (1997) The direction of retinal motion facilitates binocular stereopsis. Proc. R. Soc. B., 264, 1421-1427.

Burt, P., & Julesz, B. (1980). Modifications of the classical notion of Panum's fusional area. Perception, 9, 671–682.

Clark, J. J., & Yuille, A. L. (1990). Data fusion for sensory information processing systems. Boston: Kluwer Academic Publishers.

Deas L, Wilcox L M, (2012), "The role of stereopsis in figural grouping versus segmentation" Perception 41 ECVP Abstract Supplement, page 18.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature, 415, 429–433.

Filippini, H. R., & Banks, M. S. (2009). Limits of stereopsis explained by local cross-correlation. Journal of Vision, 9(1):8, 1–18, http://www.journalofvision.org/ content/9/1/8, doi:10.1167/9.1.8.

Frund, I., Haenel, A. V., & Wichmann, F. A. (2011). Inference for psychometric functions in the pres- ence of nonstationary behavior. Journal of Vision, 11(6):16, 1–19, http://www.journalofvision.org/ content/11/6/16, doi:10.1167/11.6.16. http://psignifit.sourceforge.net/

Gepshtein, S., & Cooperman, A. (1998). Stereoscopic transparency: A test for binocular vision's disambiguating power. Vision Research, 38(19), 2913–2932.

*Goutcher, R. & Hibbard, P. B. (2014) Mechanisms for similarity matching in disparity measurement FRONTIERS IN PSYCHOLOGY, 4, 1014*

Goutcher, R., O'Kane, L. & Wilcox, L.M. (2012). Representation of Stereoscopic Volumes. Journal of Vision, 12(9): 221.

Harris , J M (2013) Stereoscopic volume perception: effects of local scene arrangement across space and depth erception 42 ECVP Abstract Supplement

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

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

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

Harris, J.M. and Parker, A.J. (1995) Independent use of bright and dark information in stereopsis. Nature, 374, 808-811.

Harris, J.M., McKee, S.P. and Smallman, H.S. (1997) Fine-scale processing of human binocular stereopsis. Journal of the Optical Society of America, A., 14, 1673-1683.

Keeble D R T, Harris J M, Pacey I (2006) The perceived depth of a dot cloud is the centroid of the disparity distribution" *Perception* **35** ECVP Abstract Supplement.

Lages, M. (2006) Bayesian models of 3-D motion perception. Journal of Vision, 6, 4, 14, doi: 10.1167/6.4.14.

McKee, S.P. (1983). The spatial requirements for fine stereoacuity. Vision Res 23: 191–198.

McKee, S. P., & Verghese, P. (2002). Stereo transparency and the disparity gradient limit. Vision Research, 42(16), 1963–1977.

Nienborg, H., Bridge, H., Parker, A. J., & Cumming, B. G. (2004). Receptive field size in V1 neurons limits acuity for perceiving disparity modulation. Journal of Neuroscience, 24, 2065–2076.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. Science, 249, 1037–1041.

Parker, A. J., & Yang, Y. (1989). Spatial properties of disparity pooling in human stereo vision. Vision Research, 29(11), 1525–1538.

Prince, SJD; Eagle, RA (2000) Weighted directional energy model of human stereopsis. Vision Research, 40 (9): 1143-1155

Prince SJ, Cumming BG, Parker AJ (2002) Range and mechanism of encoding of horizontal disparity in macaque V1. J Neurophysiol 87: 209–221.

Qian, N., Andersen, R.A. and Adelson, E.H. (1994) Transparent Motion Perception as Detection of Unbalanced Motion Signals I: Psychophysics , J. Neurosci. 14, 7357-7366.

Smallman HS, MacLeod DI (1994) Size-disparity correlation in stereopsis at contrast threshold. J Opt Soc Am A Opt Image Sci Vis 11: 2169–2183.

Stevenson, S. B., Cormack, L. K., and Schor, C. M. (1991). Depth attraction and repulsion in random dot stereograms. Vision Research. 31, 805-13.

Tsai JJ, Victor JD (2003) Reading a population code: a multi-scale neural model for

representing binocular disparity. Vision Res 43: 445–466.

Tsirlin, I., Allision, R.S. & Wilcox, L.M. (2012) Perceptual asymmetry reveals neural substrates underlying stereoscopic transparency. Vis. Res, 54, 1-11.

Tyler CW (1973) Steroscopic vision: cortical limitations and a disparity scaling effect. Science 181: 276–278.

Tyler CW (1974) Depth perception in disparity gratings. Nature 251: 140–142.

Tyler CW (1975) Spatial organization of binocular disparity sensitivity. Vision Res 15: 583–590.

Wallace, J. M., & Mamassian, P. (2004). The efficiency of depth discrimination for non-transparent and transparent stereoscopic surfaces. Vision Research, 44(19), 2253–2267.

Welchman AE, Lam JM, Bülthoff HH. Bayesian motion estimation accounts for a surprising bias in 3D vision. Proc Natl Acad Sci USA. 2008;105:12087–12092

Westheimer, G., & Levi, D. M. (1987). Depth attraction and repulsion of disparate foveal stimuli. Vision Research, 27(8), 1361–1368.

Wilcox, L. M. (l999). First and second-order contributions to surface interpolation. Vision Research, 39, 2335-2347.

L.M. Wilcox, P.A. Duke (2005) Spatial and temporal properties of stereoscopic surface interpolation Perception, 34 (11), 1325–1338.

Yang, Y., & Blake, R. (1995). On the accuracy of surface reconstruction from disparity information. Vision Research, 35, 949–960.

**Volume perception: disparity extraction and depth representation in complex three-dimensional environments**

**Julie. M Harris**

**Supplementary Materials**

## 1. Range of stimulus parameters used

Table 1 shows the matrix of frequencies (in cycles per image) corresponding to each window size (columns) and frequency-window ratio (rows).

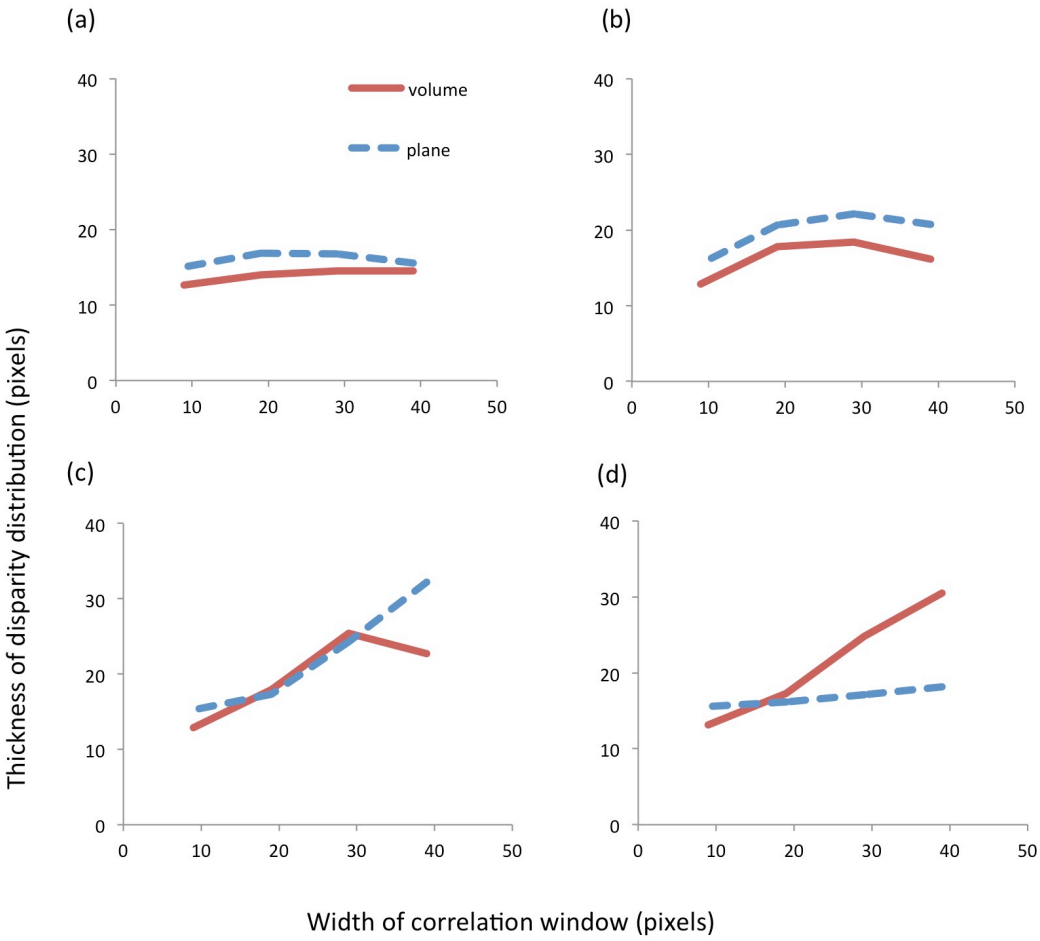
| **Fc /wsize** | **Window size -> (pixels)** | **39** | **29** | **19** | **9** |
|---|---|---|---|---|---|
| **0.25/wsize** | | 2.3 | 3.0 | 4.6 | 9.7 |
| **0.5/wsize** | | 4.5 | 6.0 | 9.2 | 19.4 |
| **1/wsize** | | 9 | 12.1 | 18.4 | 38.9 |
| **2/wsize** | | 18 | 24.1 | 36.8 | 77.8 |

Table 2 shows the matrix of frequencies (in cycles per degree) corresponding to each window size (columns) and frequency-window ratio (rows).

| **Fc /wsize** | **Window size -> (pixels)** | **39** | **29** | **19** | **9** |
|---|---|---|---|---|---|
| **0.25/wsize** | | 0.56 | 0.72 | 1.1 | 2.31 |
| **0.5/wsize** | | 1.07 | 1.44 | 2.19 | 4.63 |
| **1/wsize** | | 2.13 | 2.88 | 4.38 | 9.25 |
| **2/wsize** | | 4.26 | 5.76 | 8.76 | 18.5 |

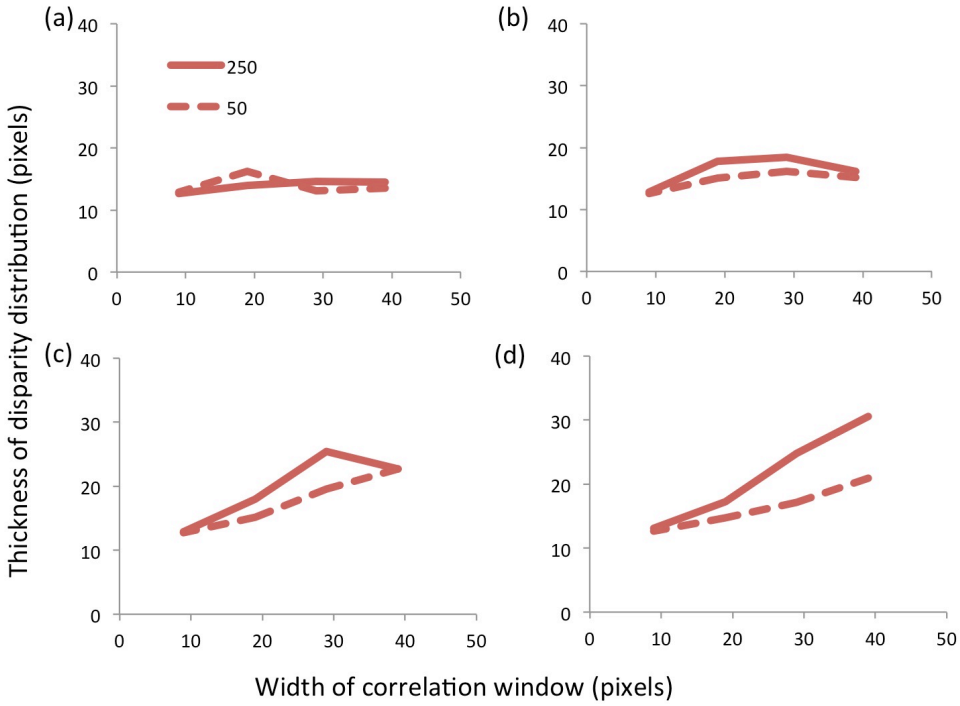
**2. Model performance for plane versus volume stimuli (as per Experiment 1)**

Graphs show model performance for the volume (solid red line) and plane (dotted blue line) stimuli for each f/w ratio used: (a) f=0.25/w, (b) f=0.5/w, (c) f=1/w, (d) f=1/2w.



**3. Model performance for volume stimuli for 250 and 50 elements (as per Experiment 2)**

Graphs show model performance for the 250 (solid red line) and 50 (dotted red line) stimuli for each f/w ratio used: (a) f=0.25/w, (b) f=0.5/w, (c) f=1/w, (d) f=1/2w.



**4. Model performance for high disparity gradient versus low disparity gradient stimuli (as per Experiment 3)**

Graphs show model performance for the volume (solid red line) and plane (dotted blue line) stimuli for each f/w ratio used: (a) f=0.25/w, (b) f=0.5/w, (c) f=1/w, (d) f=1/2w.