An Application Driven Comparison of Depth Perception on Desktop 3D Displays

Nick Holliman[a], Barbara Froner[a], Simon Liversedge[b],

[a]Department of Computer Science, Durham University, United Kingdom.
[b]School of Psychology, University of Southampton, United Kingdom.

ABSTRACT

Desktop 3D displays vary in their optical design and this results in a significant variation in the way in which stereo images are physically displayed on different 3D displays. When precise depth judgements need to be made these differences may become critical to task performance. Applications where this is a particular issue include medical imaging, geoscience and scientific visualization.

We investigate perceived depth thresholds for four classes of desktop 3D display; full resolution, row interleaved, column interleaved and colour-column interleaved. Given the same input image resolution we calculate the physical view resolution for each class of display to geometrically predict its minimum perceived depth threshold.

To verify our geometric predictions we present the design of a task where viewers are required to judge which of two neighboring squares lies in front of the other. We report results from a trial using this task where participants are randomly asked to judge whether they can perceive one of four levels of image disparity (0,2,4 and 6 pixels) on seven different desktop 3D displays. The results show a strong effect and the task produces reliable results that are sensitive to display differences. However, we conclude that depth judgement performance cannot always be predicted from display geometry alone. Other system factors, including software drivers, electronic interfaces, and individual participant differences must also be considered when choosing a 3D display to make critical depth judgements.

Keywords: Interleaving Patterns, Aliasing, Depth Perception, Empirical Evaluation, Stereoscopic Displays

1. INTRODUCTION

As 3D displays become increasingly available they are being adopted for use in applications where accurate depth judgements are critical to outcomes. In medicine and geo-science the ability to judge the co-location in depth of scene features is particularly important for operators in making domain specific judgements. For example this may involve judging the depth in the retina of anomalies in the image-based diagnosis of diabetic retinopathy[1] or the interpretation of 3D fault structure using LIDAR scanned rock outcrop data.[2]

Our concern in this paper is how well different 3D displays reproduce the depth present in an input image and particularly whether it is possible to predict human depth perception thresholds for a display from its published specifications. To analyze displays we group them into four classes according to how they physically represent the input stereo image, that is; in the original full resolution, using a row-interleaved pixel pattern, using a column-interleaved pixel pattern or using a colour-column interleaved pattern. We then use the display specifications for each class of display to predict the threshold level of perceived depth for a specific display in terms of the image disparities in the input stereo image.

We design an empirical experiment to test these predictions using a randomized within-subjects trial where participants are required to judge which of two neighboring squares lies in front of the other as shown in Figure 5. Our aim is to establish a robust and sensitive methodology for detecting depth perception differences so that the results can inform both display users and display designers in the choices they make.

Further author information: Send correspondence to Nick Holliman.
Email: n.s.holliman@durham.ac.uk, Telephone: +44 191 334 4287, Web: http://www.durham.ac.uk/n.s.holliman/

2. BACKGROUND

2.1. Previous Comparative Studies

Previous studies of human depth perception using stereoscopic displays have almost exclusively studied single displays and have investigated fusion limits, i.e. the highest levels of perceived depth a 3D display can support before fusion breaks down.

Both Yeh & Silverstein[3] and Woods[4] studied fusion limits for stereoscopic desktop 3D displays. The results showed that the total range of depth comfortably viewable on a 3D display is limited. Similar results were demonstrated for auto-stereoscopic displays by Jones et al[5] who suggested a working perceived depth range of as little as 60mm behind and 50mm in-front of the display surface. Whatever the precise value of fusible range for desktop 3D displays, it is clearly limited and we believe it is increasingly important to understand how this limited range is represented on different 3D displays.

Previous investigations into the threshold limits of perceived depth[6–8] have studied real world limits because the primary interest has been to investigate the acuity of human vision rather than the minimum perceived depth supported by different display technologies. One exception is Yeh & Sliverstein[3] who demonstrated that subjects can use minimal levels of screen disparity effectively when presented on an LC shutter glasses 3D display.

A recent study[9] compared task performance on a 2D display, a LC shutter glasses stereoscopic display, a two-view auto-stereoscopic display and a multi-view auto-stereoscopic display. This investigated interaction performance in a trial where participants had to manipulate a 3D object to be in the same depth plane as a target object. The results suggest a better performance, in terms of number of correct answers, was obtained using the LC shutter glasses, however the study did not provide a hypothesis predicting this nor explain why this might be.

None of these studies attempt to quantify human depth perception threshold levels across a range of representative 3D displays. We believe our study is the first to do this and that this is important to investigate empirically because, as the analysis in the next section predicts, there are potentially significant differences between different classes of 3D displays.

2.2. Classifying 3D Displays

To predict threshold values of perceived depth we classify the displays into four groups based on the physical interleaving pattern of pixels presented to the viewer. This will allow us to calculate a geometric prediction of performance from the published display specifications and the input image disparity.

2.2.1. Full Resolution Displays

These displays show two full resolution views, one to each eye. They provide double the number of pixels of an equivalent 2D monitor and may be implemented using two displays or by temporal multiplexing of a single display. We use three full resolution displays; a time sequential stereoscopic CRT display using CrystalEyes LC shutter glasses,[10] an auto-stereoscopic Kodak[11] display and an auto-stereoscopic IRIS-3D[12] display all driven using standard graphics card settings.

2.2.2. Row Interleaved Displays

These displays spatially interleave alternate rows from left and right images. The total number of pixels seen is unchanged compared to an equivalent 2D display and half of the total are seen by each eye. We have a single example of a row-interleaved display in this study; a ColorLink linearly polarized stereoscopic display. This is driven using a graphics card with a time-sequential video signal which on-board electronics decodes interleaving left and right images appropriately in alternate rows for display.

2.2.3. Column Interleaved Displays

These displays spatially interleave alternate columns of pixels from left and right input images. The total number of pixels seen is unchanged compared to an equivalent 2D display and half of the total are seen by each eye. Two column interleaved displays are used in our study; the DTI 2018 LCD display[13] and the SeeReal C-i display.[14] The DTI is driven using a time-sequential signal and on-board electronics generate the spatial interleaving required. The SeeReal display was used here with its head tracking feature switched off and driven with a graphics card that interleaved the left and right images appropriately in alternate columns.

2.2.4. Colour-column Interleaved Displays

These displays spatially interleave left and right pixels in alternate colour columns at a sub-pixel level. The total number of pixels seen is unchanged compared to an equivalent 2D display and half of the total are seen by each eye. One colour-column interleaved display is used in our study; the Sharp LL151D[15] auto-stereoscopic display. The colour-column interleaving was generated using a graphics card to interleave the left and right images appropriately.

3. GEOMETRIC PREDICTIONS

In this section we analyze the viewing geometry of the four classes of 3D display and from this derive a prediction of the performance we expect for human depth perception on each in terms of the smallest unit of input image disparity that it should be possible for a human to perceive on each class of display.

To predict the performance of the four classes of display we will use a generic display specification for calculations in this section although in practice the displays vary in resolution and size. We assume the base displays are flat panel displays with the following characteristics:

- *Screen resolution* 1280x1024 pixels, SXGA resolution.
- *Screen size* 17.1in; W: 337mm, H: 270mm
- *Screen pixel* size 0.264mm x 0.264mm

In addition to simplify our comparison we will assume that each class of display display is viewed at the same nominal viewing distance of 650mm by a viewer with an eye separation of 65mm.

We define some basic terminology that we use to identify key features of the images we wish to display and the displays we wish to display them on. We generally follow the terminology defined by Holliman,[16] but extend this to distinguish between the input stereo image pair that we wish to display on each class of display and the capabilities of the displays themselves. Where we are considering the input image characteristics we use the prefix *image*, where we are considering display characteristics we use the prefix *view* to identify the characteristics of one or more viewing channels on a physical display.

- *Image pixel* we use to refer to the pixels in the input image that we wish to display. We consider this as the basic unit of addressable colour in one channel of the input image.
- *Image resolution* the resolution of a single channel image we wish to display. We will define this in image pixels, to be the same resolution for each channel in the input signal.
- *Image disparity* the disparity in image pixels between two homologous points in the stereoscopic input image we wish to display.
- *View pixel* we use this to define the basic addressable unit in a single view on a specific 3D display. This can have an aspect ratio and size different to the underlying display screen, for example on a column interleaved display a single view pixel is effectively twice the width of the underlying physical screen pixel.
- *View resolution* the resolution we can display per view on a specific 3D display. We define this in view pixels and this varies depending on the optical design of the display.
- *View disparity* the physical disparity of two homologous points shown on a 3D display. This can be measured in view pixels or, as it is a physical quantity, in mm. Because view disparity is defined in view pixels different 3D displays will have different minimum values.
- *Geometric perceived depth (GPD)* the geometrically calculated perceived depth predicted to be due to a specific view disparity.[16]

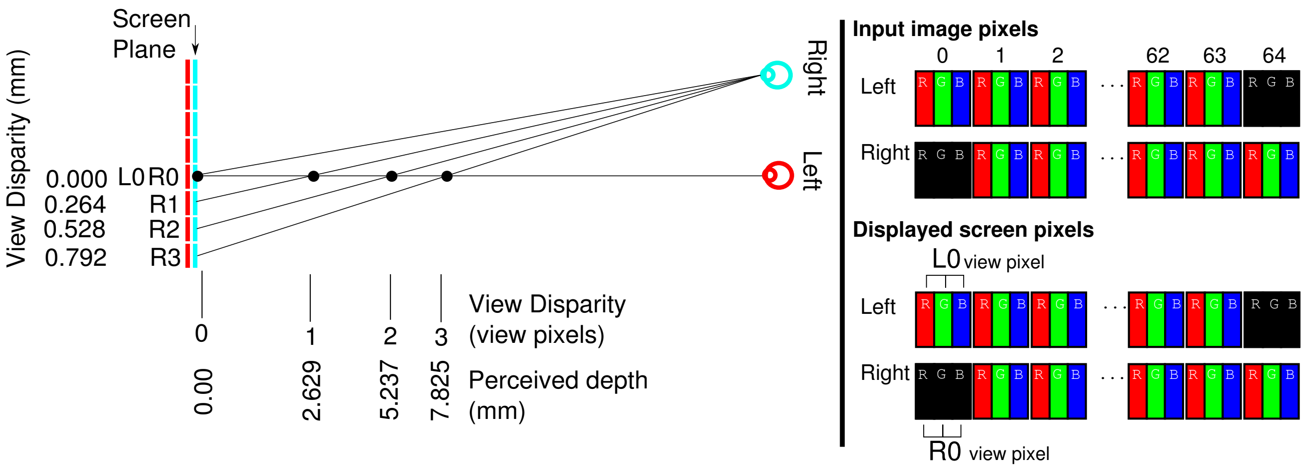

Figure 1. A full-resolution display provides a one-to-one correspondence between the input image and the view image shown on the display. On the right we illustrate corresponding rows of pixels from the left and right input images and beneath these the pattern of view pixels physically displayed on screen. To the left is illustrated the perceived depth due to physical disparities of 0,1,2 and 3 view pixels.

3.1. Full Resolution Displays

Figure 1 illustrates the generation of the view pixel pattern and the depth reproduction capability of a full resolution display using the generic display specifications we gave above. The calculation for perceived depth uses the standard equation for geometric perceived depth with crossed disparity.[16] This pre-supposes a simple geometric model of binocular vision is sufficient to provide a first order approximation of actual perceived depth.

As is clear in Figure 1 an image disparity of 0-pixels will be reproduced as 0-pixels view disparity on full-resolution displays and the minimum increment of 1-pixel image disparity will be reproduced as 1-pixel view disparity.

We predict the performance for each display in terms of which levels of input image disparity we expect to be reproduced as view disparity and hence perceived as depth by an observer. For the full resolution display this is straight-forward as we anticipate this class of display can reproduce all input image disparities as view disparity and hence all image disparities will be perceived by an observer as discrete depths. We make an assumption here that all the displays in the study show a minimum disparity above the visible threshold, hence if a display can reproduce a value of input image disparity it will be perceived as depth by an observer.

3.2. Row Interleaved Displays

For the purposes of calculating geometric perceived depth we will treat row interleaved displays as identical to full resolution displays since they have full horizontal resolution and it is therefore possible to display the entire input image disparity range. However, it is worth noting these displays have a built-in vertical offset of one view pixel which could alter the perceived depth in practice.

3.3. Column Interleaved Displays

As is shown in Figure 2 column-interleaving can be expected to have a direct effect on view disparity as the horizontal image disparity range is sub-sampled by a factor of two.

A first result of the interleaving is that 0-pixel image disparity is shown with a physical disparity of one screen pixel. The effect of this is to offset the zero disparity plane to be slightly in-front, or behind, the physical screen plane. The second result is that a column-interleaved display can only present half the disparity values in the input image. Every alternate increment in image disparity will be removed by the sub-sampling of pixel

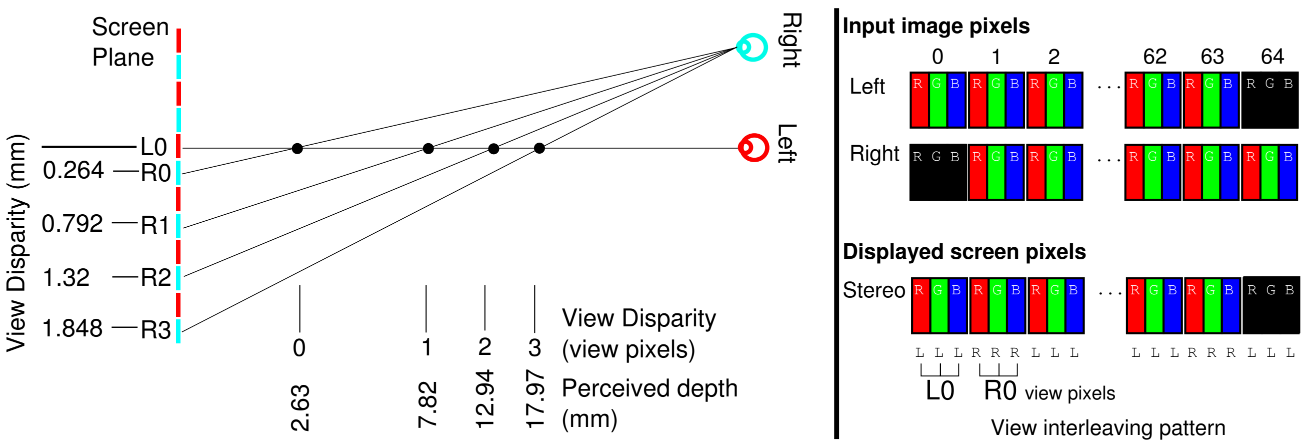

Figure 2. A column-interleaved display uses alternate columns of physical pixels to display the left and right images. As shown on the right the input image must be sub-sampled to achieve this, we assume that columns are selected alternately from left and right images. As a result a column-interleaved display can only show half the values of input disparity and the zero disparity plane is not co-incident with the screen plane.

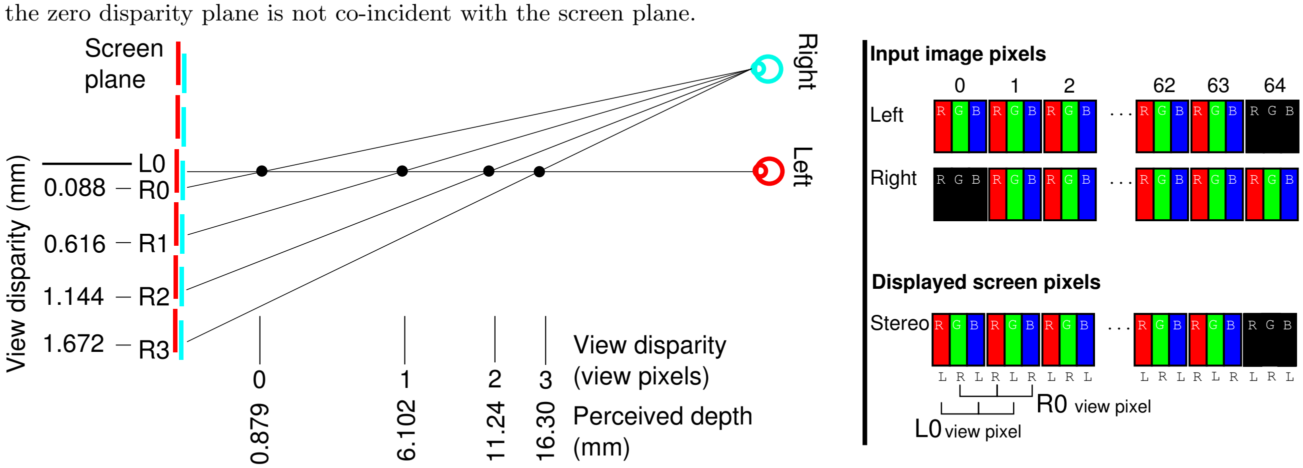

Figure 3. Colour-column interleaving samples incoming pixels at the level of sub-pixel colour components. There are several implementation choices for this sub-sampling, we assume here *pixel* columns are selected alternately from left and right images then displayed in appropriate colour-columns. A colour-column interleaved display can only show half the input disparity values and the zero disparity plane is not coincident with the screen plane.

columns. The interleaving example to the right in Figure 2 illustrates how 1-pixel disparity is aliased, producing the same interleaved pattern as 0-pixel disparity.

We therefore predict that a column-interleaved display will only support perception of certain input image disparities. We have assumed that even values of image disparity survive the column sub-sampling and therefore we predict depth will be perceived for input image disparity values of 2,4,etc -pixels.

3.4. Colour-column Interleaved Displays

Colour-column interleaved displays also sub-sample the input image but interleave at the granularity of sub-pixel colour-columns rather than pixel-columns. There are several implementation choices for this; we assume alternate pixel columns are sampled and then displayed in appropriate colour-columns across one view pixel. The result, shown in Figure 3 is that there is a 0-pixel view disparity offset of one-third of a screen pixel and that only alternate values of input image disparity values can be reproduced on the display. In addition the colour-column interleaving means that the red, green and blue colour components in a single view pixel are no longer spatially adjacent. The interleaving example to the right in Figure 3 illustrates how 1-pixel disparity is aliased, producing the same interleaved pattern as 0-pixel disparity.

We predict that a colour-column interleaving display will have the same performance as column interleaved displays; that is it will only be able to reproduce even values of image disparity.

3.5. Hypothesis

Our prediction is that different classes of 3D display will reproduce the disparity in the input image differently and that this will have a direct effect on the threshold level of image disparity for each display. We predict that full-resolution and row-interleaved displays should have a threshold level of 1-pixel image disparity. Whereas for column-interleaved and colour-column interleaved displays, which horizontally sub-sample the input image, we predict a threshold value of 2-pixels.

To begin to evaluate this hypothesis we present the experimental design below. In this experiment we investigate participant's disparity threshold using values of 0-,2-,4- and 6-pixels input image disparity, which we expect *all* displays to be able to reproduce.

4. METHOD

4.1. Experimental Design

We designed a repeated measures trial with Display (DTI, SeeReal, ColorLink, Sharp, Iris3D, Kodak, Shutter Glasses) and Image Disparity (0-, 2-, 4-, 6-pixels) as within participants variables. The choice of increment of 2-pixels image disparity ensures we have an input signal that we would expect all displays to be able to reproduce as visible perceived depth. The dependant variable was the proportion of trials at which participants select the correct target (Score). Each subject was asked to repeat the same condition twenty-eight times for each of four levels of image disparity, giving a total of 112 conditions per display. Image disparity, and hence perceived depth, was controlled in image pixels and was randomly chosen from four possible levels (0-, 2-, 4-, 6-pixel disparity), each of which was distributed across the trials with equal probability. The position of the square that appeared to be closer to the participant was counterbalanced across trials. The order in which people performed the task on each display was also counterbalanced and followed a Latin Square design.

4.2. Participants

A total of 14 candidates (11 male, 3 female) were recruited within the Durham University population. Participant age varied between 20 and 34 while the mean age was 26 years. Participants were naive concerning the purpose of the experiment; they received a nominal sum of five pounds per hour, for a total of ten pounds.

4.3. Equipment

The earlier classification of displays described the seven displays used in the trial. The displays were driven by seven independent machines that used the same kind of graphic card (nVidia Quadro FX family) and the same software driver (nVidia ForceWare Release 80). The experiment was conducted in a dark room, with minimal light levels with equipment arranged as shown in Figure 5.

4.4. Task and Stimuli

The image used for the test consisted of two white squares on a black background, as shown in Figure 5. The squares were centered in the middle of the screen and were positioned horizontally one next to the other. Between the two squares there was a small square that marked the center of the screen and acted as fixation point; participants were asked to maintain fixation on this point throughout each trial as they were performing the task.

The square that acted as fixation point was 6 image pixels wide while the width of the other two squares was 64 image pixels each. The distance between the two internal edges of the left and right square was 20 image pixels. In each trial one square was always given 0-pixels image disparity while the position in depth of the

Figure 4. The environment used, where possible displays had chin-rests to guide participants to the ideal viewing position, the displays were placed against a blank background and any reflections of objects or lights behind participants were eliminated.

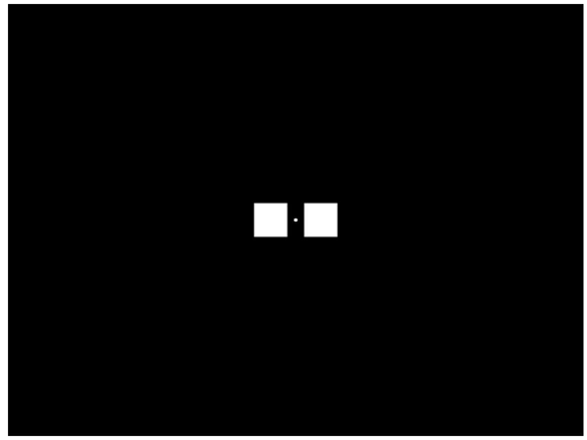

Figure 5. The trial stimulus consisted of two neighboring squares, participants were instructed to look at the fixation target between the squares and make a forced choice judgment about which square appeared to be in-front of the other.

other square was randomly chosen among a range of four different image disparity (i.e. 0-, 2-, 4-, 6-pixel image disparity).

Stimuli were presented to candidates via the stereoscopic displays at the manufacturers nominal viewing distance for the Kodak and Iris3D displays and at 65cm for all other displays. Candidates were asked to identify which square was the closest to them by pressing the letter "C" on the keyboard if the left square appeared to be closer to them or press letter "M" if instead the right square appeared to be closer, a choice was always required even if there appeared to the subject to be no difference in depth between the two squares.

4.5. Protocol

Volunteers were screened for stereovision prior the start of the experiment using the Titmus test. All participants met the minimum criteria for selection, namely, stereo-acuity at 40 sec-arc. Participants were divided into two groups of seven people each. The experiment was carried out in two separate sessions, one for each group. Prior to the start of the experiment, candidates reviewed instructions and completed practice trials with at least one of the displays.

Participants then completed the 112 experimental trials on each display for a total of seven experimental sessions. Trials started with an orthoscopic test to check that the candidates were in the correct viewing position and the display was presenting the left and right images correctly to the appropriate eye. During trials head movements were minimized via use of chin rests on all the displays except the Kodak and Iris-3D where it was impractical. Participants were instructed to be as accurate as possible in their decision but not to spend too much time on each trial, even though no time limit was imposed. Answers could not be changed and score was recorded. In each trial, candidates were assigned a score of 1 if they gave the correct answer and a score of 0 if they gave the wrong answer. In the trials where both squares had zero disparity, candidates were assigned a score of 1 if they selected the square on the right and a score of 0 if they selected the square on the left.

Finally, all candidates were debriefed and were given the chance to ask questions. The experiment lasted two hours including a thirty minute break half way through and small breaks at the end of each trial.

5. RESULTS AND ANALYSIS

We consider results from the trial in terms of the score of participants. We report data from 12 of the 14 subjects as the other two had poor average task performance (average score of 49% and 52% respectively, compared to a minimum average of 74% for the remaining participants). Overall results showed a strong effect, when participants could detect a depth difference the average score was 94%, which is close to the ideal score of 100%.

Table 1. Mean score (%) and standard deviation

Disparity	0		2		4		6	
	M	SD	M	SD	M	SD	M	SD
DTI	53.87	.31	95.83	.08	93.75	.13	95.83	.07
SeeReal	56.84	.23	52.98	.07	88.10	.22	88.69	.23
ColorLink	61.31	.22	88.39	.28	88.39	.28	89.29	.27
Sharp	63.10	.15	95.54	.07	95.24	.07	86.61	.21
Iris3D	72.02	.14	99.11	.02	100.00	.00	99.41	.01
Kodak	52.68	.24	96.43	.07	95.54	.08	95.24	.09
LC Glasses	60.71	.26	100.00	.00	98.51	.03	98.21	.05

Data were first subjected to Analysis Of Variance (ANOVA), with Disparity and Display as within-subjects independent variables and Score as the dependent variable. The ANOVA revealed that there was a significant effect of both display and disparity on performance, as well as a significant interaction between the two (all F values > 6.04 and all p values $< .001$).

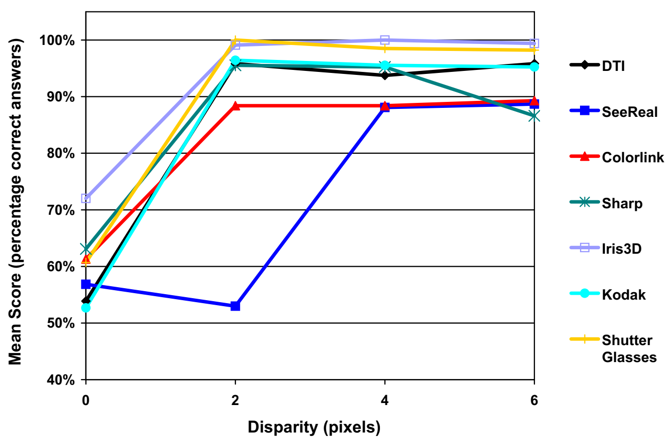

Figure 6. Graph illustrating the mean score(%) against image disparity (image pixels). We predicted that all participants would achieve an ideal score of 100(%) when the image disparity is non-zero, otherwise they should achieve a score of 50%.

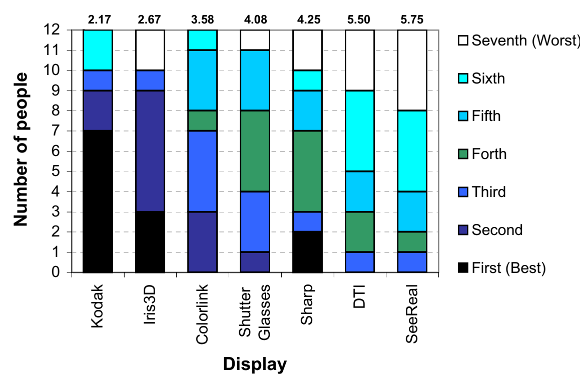

Figure 7. A subjective comparison of the seven displays. Using a post-trial questionnaire participants were asked to rank the displays in order of preference. The cumulative results are sorted by the average rank of each display.

5.1. 0-pixel Disparity

Figure 8 shows the mean score and standard deviation from Table 1 for each display when there was zero image disparity in the input image. It is important to confirm our prediction of a performance at chance (score 50%) given an image disparity of zero. If we find a reliable detectable depth difference in the 0-pixel disparity case then this may indicate problems with participants viewing position during the trial or a display problem such as an optical or mechanical misalignment.

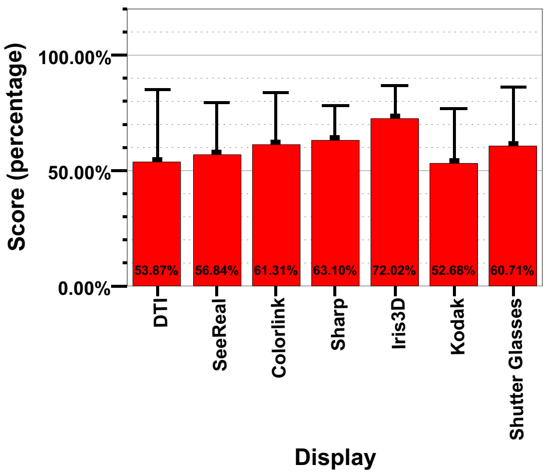

Figure 8. Mean score and standard deviation for 0-pixel disparity.

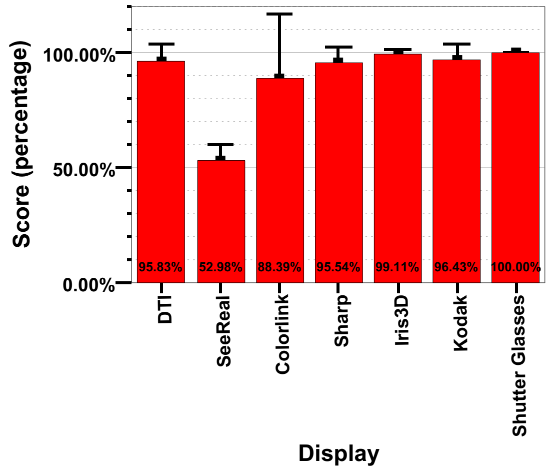

Figure 9. Mean score and standard deviation for 2-pixel disparity.

To investigate this we conducted a series of pairwise t-tests where we compared the mean score associated with each display against chance (i.e. score = 50%). Even though participants seemed to be slightly biased towards choosing the right square (all mean values > 50%), the tests showed that the mean scores for the first five displays (DTI, ColorLink, SeeReal, Kodak and Shutter Glasses) were not significantly different from chance (all $t(11)$ values < 1.77, all p values $> .11$). Therefore these five displays performed as we predicted.

When using the Sharp display participants were significantly biased towards selecting the right square when no image disparity was introduced ($M = 63\%$, $t(11) = 3.04$, $p = .01$). However, a more detailed analysis showed the performance for the Sharp display was not significantly different to the performance of the first five displays (all $t(11)$ values < 2.18, all p values $> .05$).

The Iris-3D also showed a performance significantly higher than chance ($M = 72\%$, $t(11) = 5.29$, $p < .001$) and also reliably higher than that for the first five displays (all $t(11)$ values > 1.99, all p values $= .07$ or lower). Therefore when using the Iris3D display, candidates were not performing by chance, but were systematically perceiving a difference in depth between the two squares (i.e. right square closer than left square).

5.2. 2-pixel Disparity

The second aspect of the data that we considered in detail was performance at 2-pixel disparity. Mean scores and standard deviation for this condition are shown in Figure 9. According to our predictions, all the displays should have the capability to reproduce an image disparity of 2 pixels. In order to investigate this, we performed a series of paired t-tests and all the displays performed as predicted, or only marginally poorer than our prediction, with the exception of the SeeReal display.

The t-tests revealed that the mean score for the SeeReal display ($M = 53\%$) was significantly lower than the mean score for all the other displays (all M values $> 88\%$; all $t(11)$ values > 4.28, all ps $= .001$ or lower). Specifically, when using the SeeReal display, participants were performing no differently than chance (SeeReal vs chance: $t(11) = 1.45$, $p > .1$), which suggests that they were unable to detect any difference in depth between the two squares. By contrast, when using any of the other displays, candidates were clearly able to detect depth and were performing significantly better than chance (all $t(11)$ values > 4.71, all p values $= .001$ or lower).

With respect to the SeeReal display, pairwise comparisons across disparity levels also showed that at 2-pixel disparity participants performed significantly worse than at 4-pixel disparity ($t(11) = 5.16$ and $p < .001$) but not reliably differently than at 0-pixel disparity ($t(11) = 0.51$ and $p > .5$). This suggests that with the experimental conditions adopted in our trials the SeeReal display does not have the predicted capability to reproduce 2-pixel image disparity.

5.3. 4- and 6-pixel Disparity

We predicted that all the displays in this study should be able to reproduce image disparities of 2 or more pixels. However clearly at least one display has problems reproducing 2-pixel disparity and here we investigate if high values of disparity are reliably presented on the displays. Mean task scores at 4- and 6-pixel image disparity levels are shown in Figure 10 and Figure 11.

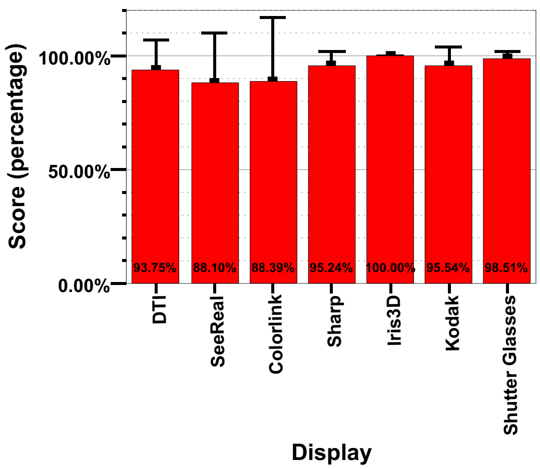

Figure 10. Mean score and standard deviation for 4-pixel image disparity.

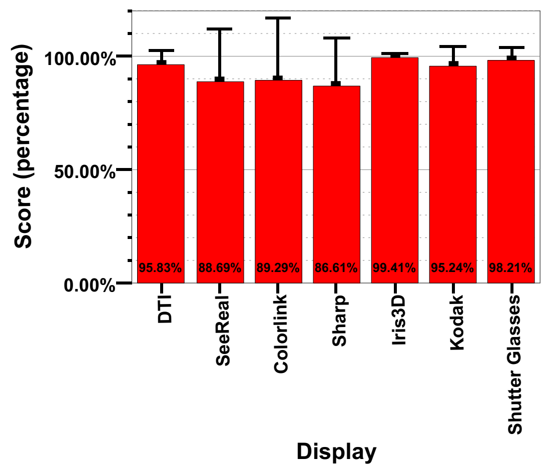

Figure 11. Mean score and standard deviation for 6-pixel image disparity.

Where a 4-pixel disparity level was applied, performance for all tested displays was high (all Mean values = 88% or higher). There were a some marginal effects however the only notable statistically reliable difference was the better performance observed for the Iris3D than for the Sharp ($t(11) = 2.46$ and $p < .05$).

A similar situation arose in the 6-pixel disparity presentation conditions. Again there were some marginal effects but we only observed reliably better performance for the Iris3D than for the DTI display ($t(11) = 2.25$ and $p < .05$).

On the basis of the numerical trends that we observe in the data, the Sharp display deserved particular attention. As the graph of Figure 6 illustrates, it is the only display that shows an apparent decrease in performance with increased disparity. To be specific, the mean score for the Sharp display drops from 95% at 2- and 4-pixel disparity to 87% at 6-pixel disparity. Nevertheless, pairwise comparisons showed that this decrease in performance was not reliable (all t values < 1.70, all p values $> .1$).

Overall, the data for 4-pixel and 6-pixel disparity conditions show that all the displays are performing as we predicted. Notably the difference in variances between scores on different displays suggests that there is scope to further investigate the possible sources and effects of this variance. For example the Iris-3D results demonstrate much lower variance than the ColorLink display.

6. SUBJECTIVE RESULTS

At the end of each experimental session participants were asked to complete a detailed questionnaire relating to the display they had just used (level of ease seeing 3D, disturbing factors, level of discomfort, general comments about the display). They were also asked to fill in a more general survey at completion of the whole experiment. A cumulative comparison of participants subjective display rankings from the questionnaire is shown in Figure 7.

7. GENERAL DISCUSSION

Overall, the fact that participants generally had high performance is evidence that the task was appropriate and clearly defined. We therefore believe that the data obtained are meaningful and provide insight into the characteristics of depth perception on the different 3D displays we tested.

The general ANOVA reveals a strong effect of the independent variables, Display and Disparity, on participant's score and a strong interaction between the two. In regard to disparity we have demonstrated a clear threshold effect, that is we can identify at what level of image disparity perceived depth becomes visible on a display. We also find that display has a direct influence on participant's score and additionally that within one class of 3D display there are significant variations between displays.

7.1. 0-pixel Input Image Disparity

The analysis of 0-pixel disparity data using t-tests revealed that performance for the DTI, SeeReal, ColorLink, Kodak and Shutter Glasses displays was not significantly different than chance and therefore in accordance with our predictions. That is for these displays, when the input image disparity is zero for both squares there is no perceived depth difference between the left and right squares.

With respect to the Iris3D display task performance was significantly different from chance and in this case also significantly different to any other display. We therefore need to consider why participants might be seeing perceived depth on this display when there is no input image disparity. We identified three possible sources of modification to the signal that might affect perceived depth. The first is optical, a number of participants reported observing a secondary peripheral reflection of the stimulus when viewing the monitor. The second is electronic, one channel from the driving PC was fed to the display via a video splitter in order to simultaneously drive an external 2D monitor. this could result in a delay to the signal to one eye. The third possibility is a mechanical component misalignment in the display itself.

When we ran a similar version of this experiment using the Iris3D display in a later trial the first two possibilities were removed and it was found that there had been a small mechanical alignment error in the prototype display used in the current trial which when corrected resulted in performance as predicted for the Iris3D. It is encouraging that our evaluative methodology is sensitive enough to detect this low level of display misalignment.

7.2. 2-pixels and above Input Image Disparity

From our geometric predictions we expected all the displays in the study to show perceived depth for an input image disparity of 2-pixels or above. That is we would expect all participants on all displays to achieve a score significantly higher than chance. Our analysis showed that this was the case for the ColorLink, the Iris3D, the Kodak, the DTI, the Sharp and the Shutter Glasses displays. Participants were able perceive depth on all these displays when the input image disparity was 2-pixels or higher.

The results for the SeeReal display were however not consistent with our prediction. Participants were unable to perceive depth for 2-pixels image disparity and it was not until 4-pixels disparity and above that the score was reliably better than chance. Investigating this effect in detail led us to conclude that this was an effect directly due to aliasing of the input image disparity. In adjusting the input image to show 2-pixels disparity we shifted one square to the left one image pixel and the other square to the right one image pixel. The outcome was that both adjustments were masked by the interleaving process shown in Figure 2 and the result was no visible view disparity. Had we chosen to shift one view image by 2-pixels our investigation showed that participants would have seen some view disparity. Further investigation on the other column-interleaved display in this study, the DTI, showed the same aliasing effect at 2-pixels disparity could be generated for the DTI simply by altering the starting image position of the stimulus from even to odd pixel columns. In both cases the source of this aliasing is the software drivers and/or the electronic interface rather than the display itself.

We would anticipate a similar aliasing artefact will be generated by the colour-column interleaving process used on the Sharp displays but our experiment would need to be repeated at image disparity intervals of 1-pixel to investigate this.

8. CONCLUSION

We have investigated the quality of reproduction of image disparity as perceived depth on a range of desktop 3D displays. Our methodology has generated statistically robust results and demonstrates significant differences between the human perception of depth on different 3D displays. It has also proved to be sensitive enough to detect small display misalignments.

For application users the results suggest that care should be taken when selecting 3D displays for tasks where critical depth judgements are made. Not all displays are capable of reproducing the same image disparity and there are significant differences between displays, even between those that belong to the same class of 3D display. The differences between displays belonging to the same class appear to be due to aliasing introduced by software drivers and electronic interfaces rather than the display's optical design. The fact that some participants who passed a stereo vision test were unable to generate reliable scores in this study also suggests a need to carefully screen operators using 3D displays for critical tasks.

Finally we conclude there is a need to run this task using single pixel image disparity increments. This will identify performance differences between displays in representing disparity for fine depth judgements. In addition it should provide a more detailed understanding of the aliasing artefacts for each class of display.

9. ACKNOWLEDGEMENTS

The authors would like to thank all those who supported this work. In particular ColorLink Corp., Iris3D Ltd., and Kodak Corp. for the loan of their respective display equipment and technical discussions regarding these systems. The authors also thank Prof. John Findlay for his advice on the experimental design and Dr. Gustav Kuhn for his assistance with the statistical analysis of the data. Additionally we thank the Faculty of Science at Durham University for support of the Durham Visualization Laboratory.

REFERENCES

1. M. S. Habib, J. Lowell, D. Vaideanu, N. Holliman, A.Hunter, and D. Steel, "Assessment of qualitative stereo viewing and quantitative mapping of optic disc using polarized goggles verses autostereoscopic screen," *Proceeedings of the Annual Conference of the American Academy of Ophthalmology*, October 2005.
2. K.J.W.McCaffrey, R.R.Jones, R. Holdsworth, R. Wilson, P. Clegg, J. Imber, N.S.Holliman, and I. Trinks, "Unlocking the spatial dimension: digital technologies and the future of geoscience fieldwork," *J. Geological Society* **162**(6), pp. 927–938, 2005. ISSN 0016-7649.
3. Y. Yeh and L. Silverstein, "Limits of fusion and depth judgements in stereoscopic color displays," *Human Factors* **1**(32), 1990.
4. A. Woods, T. Docherty, and R. Koch, "Image distortions in stereoscopic video systems," *Proceedings of SPIE* **1915**, 1993.
5. G. Jones, D. Lee, N. Holliman, and D. Ezra, "Controlling perceived depth in stereoscopic images," in *Stereoscopic Displays and Virtual Reality Systems VIII*, *Proceedings of SPIE* **4297A**, 2001.
6. N. Langlands, "Experiments on binocular vision," *Trans. Optical Soc.* **XXVII**(2), pp. 4–82, 1926.
7. B. Julesz, *Foundations of cyclopean perception*, The University of Chicago Press, 1971.
8. D. Diner and D. Fender, *Human engineering in stereoscopic viewing devices*, Plenum Press, 1993. ISBN 0-306-44667-7.
9. Z. Y. Alpaslan, S. Yeh, A. A. R. III, and A. A. Sawchuk, "Effects of gender, application, experience, and constraints on interaction performance using autostereoscopic displays," in *Stereoscopic Displays and Applications XVII*, *Proceedings of SPIE* **6055A**, 2006.
10. L. Lipton, "Liquid crystal shutter system for stereoscopic and other applications.." United States Patent, Patent No. 4967268, 1983.
11. J. Cobb, "Autostereoscopic desktop display: an evolution of technology.," in *Stereoscopic Displays and Applications XVI*, *Proceedings of SPIE* **5664**, pp. 139–149, 2005.
12. S. McKay, S. Mason, L. Mair, P. Waddell, and S. Fraser, "Stereoscopic display using a 1.2-m diameter stretchable membrane mirror," *Proceedings of SPIE* **3639**, pp. 122–131, 1999.
13. J. Eichenlaub, "Developments in autostereoscopic technology at dimension technologies inc.," in *Stereoscopic Displays and Applications IV*, *Proceedings of SPIE* **1915**, pp. 177–186, 1993.
14. A. Schwerdtner and H. Heidrich, "Optical system for the two and three dimensional representation of information." US Pat. No. 5,774,262, June 1998 (filed Germany 1993).
15. A. Jacobs, J. Mather, R. Winlow, D. Montgomery, G. Jones, M. Willis, M. Tillin, L. Hill, M. Khazova, H. Stevenson, and G. Bourhill, "2D/3D switchable displays," *Sharp Technical Journal* (4), 2003.
16. N. Holliman, *Handbook of Opto-electronics*, ch. Three-Dimensional Display Systems. Taylor and Francis, May 2006. ISBN 0 7503 0646 7.