Seeing "Ghost" Planes in Stereo Vision

Daphna Weinshall
Center for Biological Information Processing, E25-201
Department of Brain and Cognitive Sciences
MIT, Cambridge MA 02138, USA

Abstract

I have studied particular ambiguous random dot stereograms where multiple matches (that are equally possible) are available at each point. The human visual system resolves these ambiguities in two qualitatively different ways. In some cases a few transparent surfaces are perceived corresponding to all the ambiguous matches. In other cases a single dominant opaque surface is perceived. The conditions under which each behavior occurs are described. Additional experiments, designed to explore whether a number of modified stereo matching algorithms can predict human perception, are described, and their theoretical implications are discussed.

keywords: binocular vision, matching, uniqueness, smoothness, multiple matching, double nail illusion, disparity gradient.

1 Introduction

Since the image on the retina of an eye, or in a camera, is obtained by projecting a three-dimensional world onto a two-dimensional surface, the information on the third dimension, depth, is lost. However, a perception of depth can still be achieved by using cues such as motion parallax or stereopsis. In stereopsis, information is combined from two images, taken simultaneously from slightly different positions by the two eyes. In computer vision algorithms, the extraction of depth from binocular stereo begins with the formation of a disparity map by matching the two images (the disparity of an object is defined as the difference between its positions in the two images). Thus, a disparity value is assigned to every location in the image.

In a random dot stereogram, e.g. in figure 3a, (and in many natural scenes) matching objects in the two images to compute disparity is difficult. For every object there are many potential target matches, typically all but one of which are false. Two constraints on matching were proposed to solve the false targets problem (e.g., Marr and Poggio (1976)).

The first constraint, uniqueness, forces each object in one image to be matched to at most one object in the other image. The second constraint, smoothness (often called continuity or minimal variation in depth), requires that nearby objects in one image be matched to nearby objects in the other image. Most of the existing stereo algorithms, as well as most of the theories of human stereo vision, incorporate these two constraints, e.g. Marr and Poggio (1979), Mayhew and Frisby (1981), Nishihara (1984). [7, 8, 9, 6]

The double-nail illusion (Krol and van de Grind (1980))[5] illustrates the use of these constraints by human stereo vision. In figure 1, the two bars in the left image can be matched to the two bars in the right image in four different ways. Two of them correspond to matching the left bars in both images (L_1; R_1) with each other, and likewise the right bars (L_2; R_2). These two order-preserving matches are unique and smooth, henceforth they will be called the ordered matches (see also Yuille and Poggio (1984)[14]). The other two matches are the "ghost" matches[1]. Humans seem to prefer the unique and smooth ordered matches, seeing an illusory percept of two bars lying one beside the other in the same depth even when the stimulus is of two bars lying one behind the other, corresponding to the two "ghost" matches.

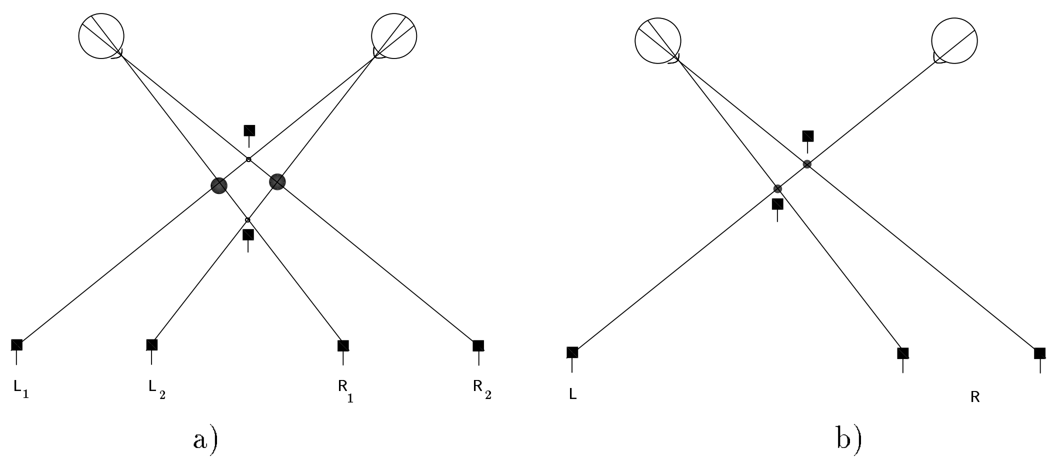


Figure 1: a) Ambiguous matching or the double nail illusion: there is one "natural" order preserving matching of L_1 to R_1 and L_2 to R_2. This solution is marked in full circles, and is always perceived, even when the correct matching is L_1 to R_2 and L_2 to R_1 (the "ghost" solution, marked with two "nails"). b) Panum's limiting case: here one eye sees two lines and the other sees only one. Most people perceive the two lines differing in depth, as if they match the single line in one image simultaneously with both lines in the other (if the distance between them is sufficiently small).

Problems arise in some situations when imposing the constraints of uniqueness and

[1]Note that ghost is used here as an adjective rather than in the sense of a physical state of affairs as in Krol and van de Grind's experiments with real nails.

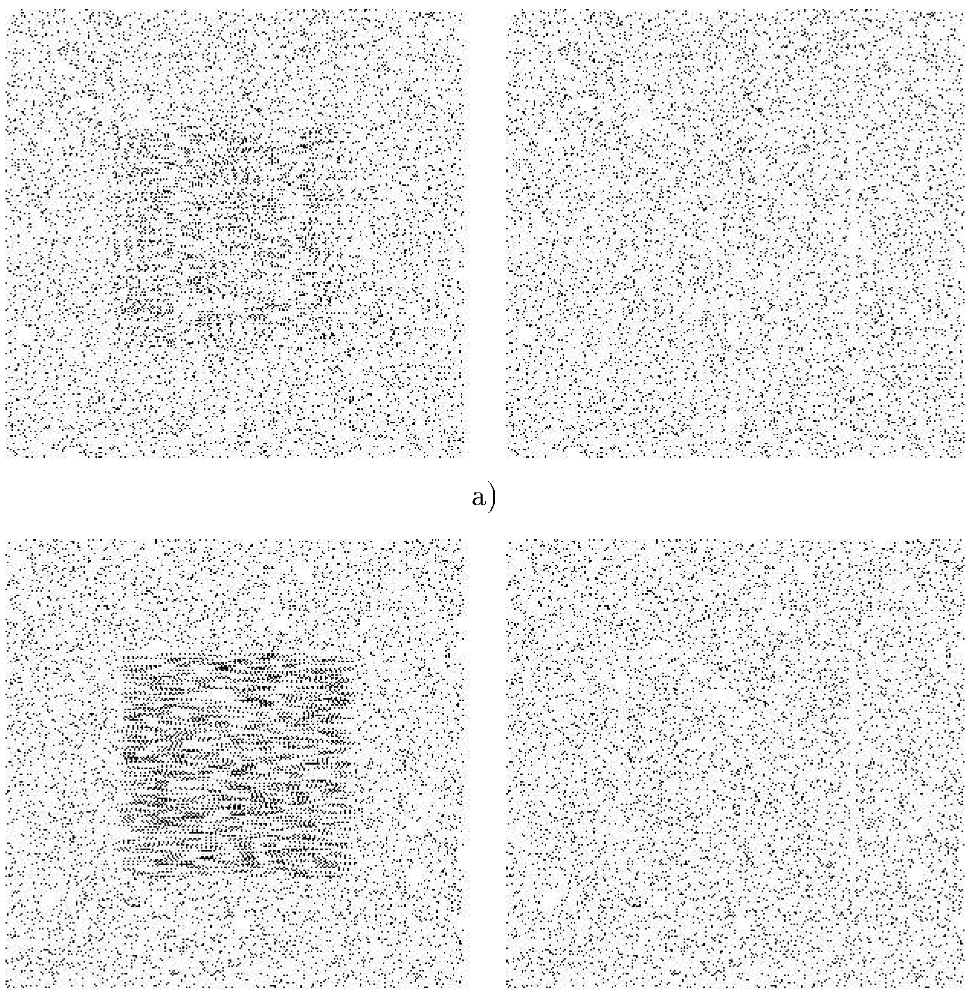
smoothness on the matching process is detrimental. The smoothness assumption breaks down for scenes that include transparent objects, such as fences. Looking through a fence, one may see a tree to the left of a given bar in the left eye and the same tree to the right of the same bar in the right eye. (It should be noted that simple modifications of the smoothness constraint can overcome this problem as in Prazdny (1985)[11]). A more serious problem with smoothness arises at those locations in the images where there are depth discontinuities. Smoothing across discontinuities is bound to produce errors in stereo, as discussed in Blake and Zisserman (1987)[2].

Panum's limiting case is an example where uniqueness does not seem to be fully obtained. Consider a stimulus where in one image there is one vertical bar and in the other image there are two (figure 1b). Here the perception corresponds to a multiple matching, namely, the two bars are seen at different depths corresponding to matching the single bar in one image to the two bars in the other. Braddick (unpublished results) has generalized this result by constructing a stereogram in which a random pattern is replicated once in one eye and twice in the other with a horizontal gap of a few pixels. As with single bars, humans perceive two planes at different depths, the upper one transparent. (Transparency is usually not perceived in line-stereograms of Panum's limiting case, but diplopia is often experienced.) I have examined this case further and observed that the disparities of the replicated pattern need not be constant, they can vary in any continuous way (see figure 2a). Also, the same pattern can be replicated more than twice (e.g., three copies at three different disparities), in which case more than two planes are perceived (e.g. three, see figure 2b), though it becomes more difficult to fuse the stereogram. The effect is visible with a very low density of points (0.001). Still, these results do not necessarily imply non-unique matching (following Grimson (1981)[4]). If matching is done simultaneously from each image to the other, the matching from the double image to the other one is indeed unique.

In this paper I discuss difficulties with the uniqueness constraint. To investigate whether human stereo vision always obtains a unique matching, I use a particular family of ambiguous random dot stereograms in which no single matching is in any way "better" than all the others. These stereograms are described in the methods section below. (A preliminary report of the first two experiments has appeared in Weinshall (1989)[13].)

2 General methods

The basic stimulus is an ambiguous random dot stereogram. This stereogram is an extension of the double nail illusion stimulus to a Random Dot Stereogram (RDS), similar to Braddick's extension of Panum's limiting case. Thus an ambiguous configuration of dots similar to the configuration of the double nail illusion stimulus is the **micropattern** of the RDS. The RDS is generated in the following way: first a sparse random pattern is computed, in which about 9% of the points are black and the rest white. The pattern of black points is replicated twice in each image with a different horizontal spacing between the two copies – G_r pixels in the right image and G_l pixels in the left image (e.g. figure 3a and figure 4a). The background is



a)

b)

Figure 2: Variations on the extended Panum's limiting case: a) the disparity of the pattern in the left image changes in a periodic way, cos and -cos, with period equals to the width of the image. Two transparent surfaces are perceived. b) A random pattern is replicated three times in the left image and once in the right image. Three transparent planes are perceived.

a random pattern, identical in both images.

Each black point from the original pattern generates a pair of points in each image. Each point in such a pair can be correctly matched to both points in the corresponding pair in the other image. These two pairs are the **micropattern** of the stereogram: the stereogram consists of randomly distributed identical pairs like that. The matching of each **micropattern** is ambiguous since it is an instance of the double nail illusion stimulus (figure 3b). There are four possible matchings that are equally plausible, two mutually exclusive pairs if a dot can only be matched to a single dot in the other image (full and hollow circles in figure 3b). The ambiguity in the matching of each micropattern leads to the ambiguity of the entire RDS (in addition to the "usual" ambiguity in RDS due to the existence of many false matches).

The stereograms were presented on a CRT monitor (Mitsubishi Diamond Scan high resolution color-display monitor, resolution 1280x1024 pixels). The disparate images were displayed using a flicker-free method (StereoGraphics^{TM} system) with an interlaced frame rate of 60 Hz. The images were viewed through shutter glasses (StereoGraphics^{TM} system) that were synchronized to present the appropriate images to the left and right eyes. In all the experiments, the viewing time was unlimited and vergence movements unconstrained. The subjects sat in a dark room, about 90 centimeters away from the stereo monitor. Each pixel on the monitor corresponded to roughly 1.2 minutes of arc.

In a set of experiments subjects viewed a series of 23 stereograms and were asked to report all the surfaces they could see in each one, from one opaque layer to any number of transparent layers. They were first asked to estimate the number of layers they could see simultaneously and then use a probe (a "3D cursor", 4 pixels, or approximately 5 minutes, wide) whose depth they could control to identify the depth of each of these layers by moving the probe to lie on each of them. At this second stage subjects were asked to report the depth of any layer they subjectively perceived as a plane, ignoring isolated points in depth. The order of the stereograms was randomized for each subject. Ten subjects participated in these experiments, five of them naive. Only three subjects have had any previous experience with stereograms of this type. All subjects had normal or corrected to normal vision and were not selected for high stereo acuity. The author and one additional subject were unable to perceive distinct layers in these stereograms, and are not included in this sample.

The full resolution of these stereograms usually required some time (typically a few seconds) and the use of vergence and memory. It has previously been shown in Akerstrom and Todd (1988)[1] that vergence is needed in the perception of transparent surfaces in stereo. The depths of surfaces reported by some subjects were inaccurate by up to 2.5 minutes of arc away from one of the possible disparities (this agrees with the reported gap resolution of 2-3 min of arc in Stevenson, Cormack and Schor (1989)[12]) and sometimes a single layer was reported twice. In the analysis of the results each reported disparity was identified with its nearest neighbor in the set of possible disparities and ignored (namely, treated as a false identification) if the difference was equal or larger than three minutes of arc. Repetitions were ignored.

3 Results and discussion

3.1 Experiment 1: multiple matching

Results:

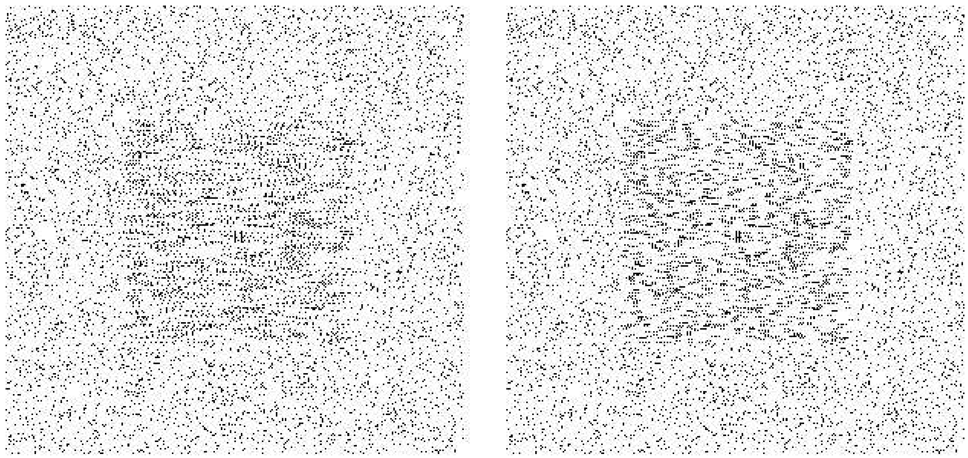
The first stimulus was an ambiguous random dot stereogram, described in the general methods, with $G_r \neq G_l$ (figure 3). There were three such stereograms in the experiment, with an interlayer disparity difference of 4, 5, and 6 pixels (4.8, 6 and 7.2 minutes of arc respectively). The disparities of the four possible layers in pixels were (0 4 8 12), (-5 0 5 10) and (-12 -6 0 6) respectively. Positive disparities indicate layers lying in front of the background whose disparity was 0. When only single features (bars) arranged in the same way are observed, only the two ordered matches marked in figure 3b with full circles are seen as reported in Krol and van de Grind (1980))[5].

	2 layers	3 layers	4 layers	ghost plane
first stereogram (0 4 8 12)	3	4	3	8
second stereogram (-5 0 5 10)	1	7	2	10
third stereogram (-12 -6 0 6)	5	3	2	5

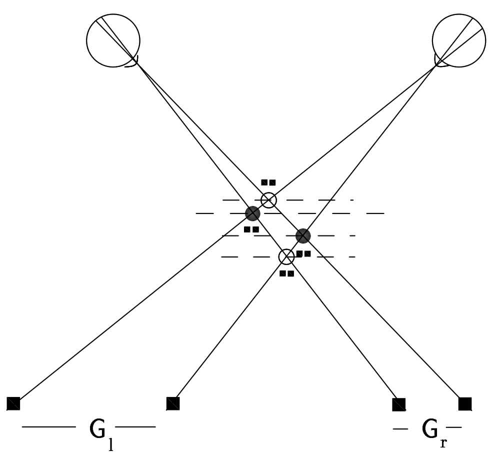
Table 1: Results of the first experiment. The first three columns of the table list how many observers identified two, three or four layers respectively with each stereogram. The last column lists how many subjects identified at least one "ghost" layer.

Table 1 summarizes the results of the first experiment. Note that at least one subject identified one "ghost" plane while not identifying both of the ordered planes. The degraded performance in the third stereogram can probably be explained by the fact that two layers were behind the fixation plane and divergence of the eyes (instead of convergence) was required to perceive them. The perception of a few transparent surfaces took time to build. Almost always the surfaces were reported to be seen simultaneously, although vergence or attention shifts had been probably used to determine the depth of each layer (in most cases only three layers lied within Panum's fusion area for the average person). In figure 4a there are smaller disparity differences between neighboring layers than in figure 3a. It is therefore easier to see all four transparent surfaces simultaneously, but it is harder to separate them in depth.

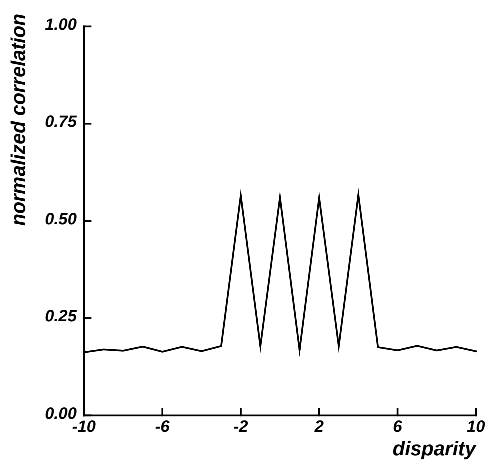
In a control experiment, four subjects were presented with a stereogram in which four nonambiguous planes existed at disparities (-5 0 5 10). The number and depth of the layers they identified was very similar to the number and depth of layers they identified under the equivalent condition with ambiguous stereograms described above. Thus the difficulty in the perception of more than three layers seems unrelated to the ambiguity in these stereograms.



a)

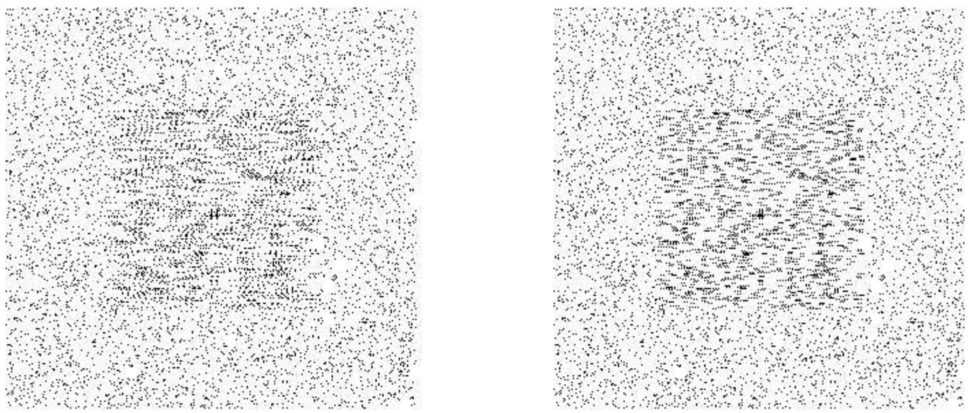


b)

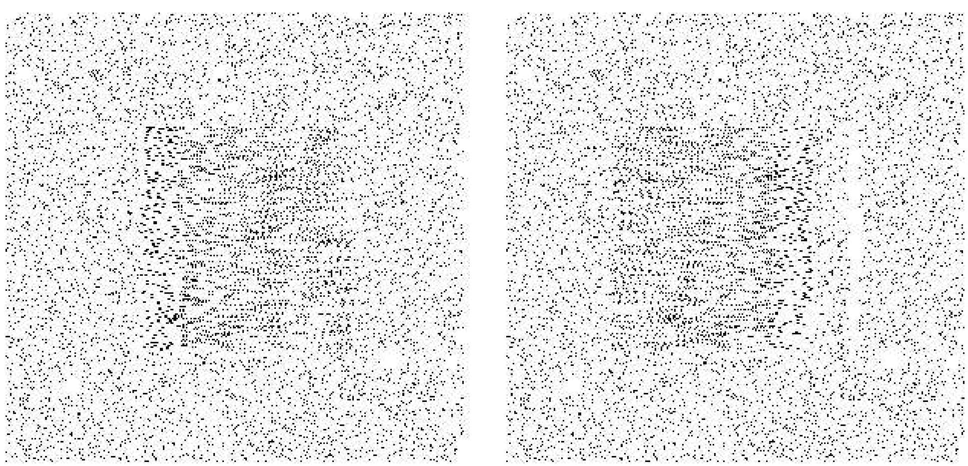


c)

Figure 3: a) An ambiguous stereogram with $G_r = 2$ pixels and $G_l = 4$ pixels. After some time four planes emerge: two in front of the background, the background (all three transparent), and one behind the background. The deepest plane is often difficult to see. It helps to diverge the eyes slightly to capture it. Note the two bars ("nails") in the middle of the stereogram, which are usually seen on the two middle surfaces only, and which are hard to flip to the other surfaces. b) A graphic illustration of the projections of two corresponding pairs of points from the stereogram in (a) (the micropattern of the stereogram). The two pairs of matches that are mutually exclusive if matching is unique are separately marked by filled and hollow circles. c) The normalized correlation between the left and the right images at their center (1). The X-axis is the disparity D. The size of the window W is the width of the ambiguous square at the middle of each image. The correlation value is normalized by the autocorrelation of the right image, i.e. the number of features in the window, so that it equals 1 if there is perfect matching of all the points at some disparity value.



a)



b)

Figure 4: Variations on figure 3a: a) smaller disparities of 5'-6' are separating two nearby layers, so that it is easier to see four layers simultaneously but somewhat more difficult to separate them in depth; b) two fronto-parallel planes and two tilted planes that create an "X" shape between them in depth can be seen.

Discussion:

A few remarks seem appropriate at this point:

- Subjects were very confident they could see distinct layers and not a single fuzzy slab. At the beginning of each session they were asked to report the number of distinct layers they perceived. The number reported generally agreed with the number of layers they went on to record. Subjects were also confident they saw complete surfaces and not isolated patches that lie on different depth in different locations.

- The surfaces perceived after the fusion of these stereograms are not those predicted by the resolution of ambiguity based on single features. More specifically, features almost never produce ghost matches whereas random dot stereograms almost always do.

- If bars were added to the stereogram (see figure 3), usually only two of them were seen in the two middle (ordered) planes. (Note that Krol and van de Grind (1980)[5] also found that one observer occasionally saw a third match in the "double-nail" experiment.)

- Multiple matching of ambiguous stereograms is not restricted to fronto-parallel planes: one can construct tilted planes like in figure 4b, and other surfaces.

- The density of points affects the number of transparent layers that are perceived: one subject that had identified four transparent layers in the experiment described above could see only two layers when the density of points was as low as 0.2%, for example.

In these ambiguous stereograms, each dot has a few possible disparities and the support from neighboring points can be used to select the best disparity. One way to measure the support for each disparity from neighboring points is the use of the correlation between patches around the point. More specifically, let me define the support of a given disparity at a given dot in one of the images to be the value of the correlation function between the two images. The correlation for disparity D is computed between a window of width W around the dot in one image and a window around the same location in the other image, translated horizontally by D:

$$\text{correlation} = \sum_i^W \sum_j^W I(x_i^R, y_j^R) I(x_i^L + D, y_j^L). \tag{1}$$

Among all the possible disparities, few are supported by the neighbors and they will be called "solutions". A "solution" is a disparity whose correlation value is sufficiently above noise level, or, equivalently, a disparity that has a local peak in the correlation function (figure 3c). The results reported above suggest that all the solutions, or disparities with sufficient support, give rise to the perception of distinct transparent layers. In other words, when the different matchings are equally likely, the resolution of matching ambiguity gives all those disparities that are locally supported by neighboring points and not a unique disparity. Thus surfaces corresponding to multiple matching are perceived.

The next experiment tests the hypothesis that all the solutions give rise to the perception of distinct transparent surfaces.

3.2 Experiment 2: unique matching

Results:

The second stimulus was an ambiguous random dot stereogram, described in the general methods, with $G_r = G_l$, see figure 5. There are three possible solutions (figure 5c) in this case: one “strong” solution (the highest peak in figure 5c) where all the points in one image are matched to all the points in the other image, and two “weak” solutions (the lower peaks in figure 5c), in which half the points in one image are matched to half the points in the other image (the “ghost” disparities). In analogy to the first experiment, three transparent layers should be identified.

Three such stereograms were used in the experiment, with interlayer disparity difference of 4, 5, and 6 pixels as before. The disparities of the three possible layers were (-3 1 5), (-10 -5 0) and (0 6 12) respectively, where a positive disparity indicates layers lying in front of the background whose disparity was 0. In contrast to the first experiment, all ten subjects identified only one opaque layer corresponding to the “strong” coherent solution. They never identified the “ghost” solutions even when instructed to look for them while changing vergence and using the probe tuned to the depth of one of the ghost planes. Many subjects felt that the single layer they had identified had some volume and sometimes reported two transparent surfaces lying very close to each other (one disparity unit, or 1.2 minutes of arc, apart). Other subjects estimated the width of this layer to be of an order of magnitude smaller than a pixel.

Discussion:

It is still possible that the initial hypothesis is correct. Perhaps only solutions (locally supported disparities) with approximately equal support, i.e. comparable maxima in the correlation function, are detected, whereas weaker solutions are suppressed. The next experiment tests the hypothesis that only disparities for which the correlation function between the two images attains some global maximum give rise to the perception of distinct transparent surfaces.

3.3 Experiment 3: multiple and unique matching

Results:

The third stimulus was an ambiguous random dot stereogram similar to that used in the first experiment, but the number of points in one or more planes was doubled by adding points that could be matched unambiguously at the disparity of that plane(s). Thus the support value at the disparity of this plane(s) was twice the support of the other planes. Two

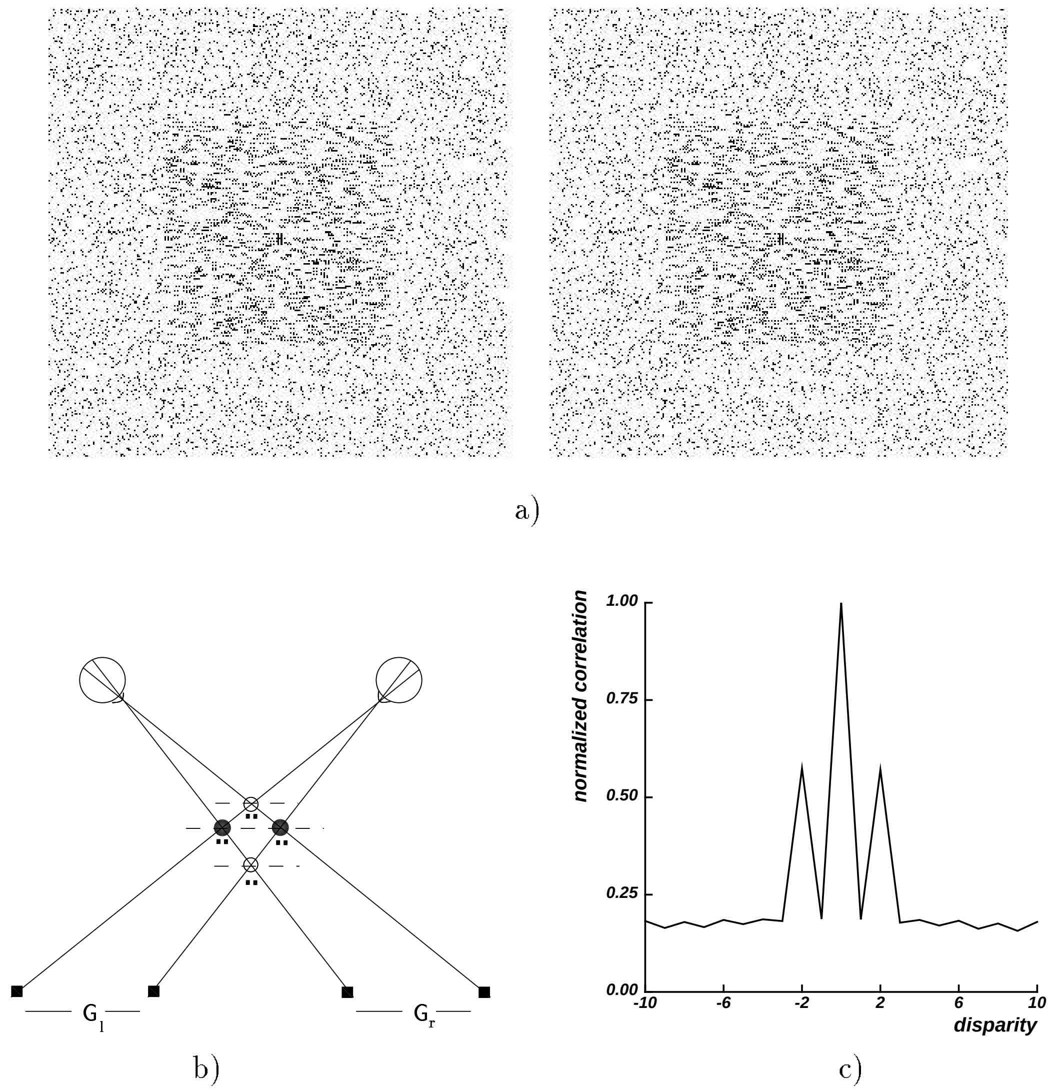


Figure 5: a) An ambiguous stereogram with $G_r = G_l = 2$ pixels. Here only one opaque surface is seen, the background, and no vergence can help detect the other two surfaces (one above the background and one below it). b) A graphic illustration of the projections of two points from the stereogram in (a) (the micropattern of the stereogram). c) The normalized correlation between the left and the right images at their center (1).

such stereograms were used, similar in structure to the second stereogram used in the first experiment (with solutions at disparities (-5 0 5 10)). In the first stereogram unambiguous points were added to the deepest plane (at disparity -5), one of the ghost planes (see plot of correlation in figure 6b). In the second stereogram unambiguous points were added to the two ordered planes at disparities 0 and 5 (see figure 6a and plot of correlation in figure 6c). The last case is particularly interesting since points were added to the two ordered surfaces, whose selection would satisfy smoothness and uniqueness, analogously to the dominant surface in figure 5a. The support of these two solutions was doubled.

	2 layers	3 layers	4 layers	ghost plane
first stereogram (-$\hat{5}$ 0 5 10)	1	5	4	10
second stereogram (-5 $\hat{0}$ $\hat{5}$ 10)	2	7	1	8

Table 2: Results of the third experiment. The first three columns of the table list how many observers identified two, three or four layers respectively with each stereogram. The last column lists how many subjects identified at least one "ghost" layer.

Table 2 summarizes the results for this experiment. In the second stereogram, 8 subjects identified at least one of the ghost planes even though their support values were half the support of the ordered planes. It turned out that even if the number of points in one surface was quadrupled, the other surfaces were visible.

Discussion:

The conclusion from this experiment is that the correlation between images at each disparity is not sufficient to predict when the disparity will give rise to the perception of a transparent layer and when it will not. This conclusion is expanded on in the next experiment.

3.4 Experiment 4: intermediate cases

Results:

We have seen that a unique matching was perceived when the two ordered matches were at the same disparity (experiment 2) and multiple matchings were perceived when the disparity difference between the two ordered surfaces was larger than 4 minutes of arc (experiment 1). How does the transition occur? To clarify the transition between these two qualitatively different perceptions, three additional stereograms of the type used in the first experiment were used, with disparity difference between the two ordered matches of 1, 2 and 3 pixels. The disparity of the four possible layers relative to the background were (-4 2 3 9), (-3 2 4 9) and (-6 -2 1 5) respectively. In the first stereogram all subjects identified the ordered

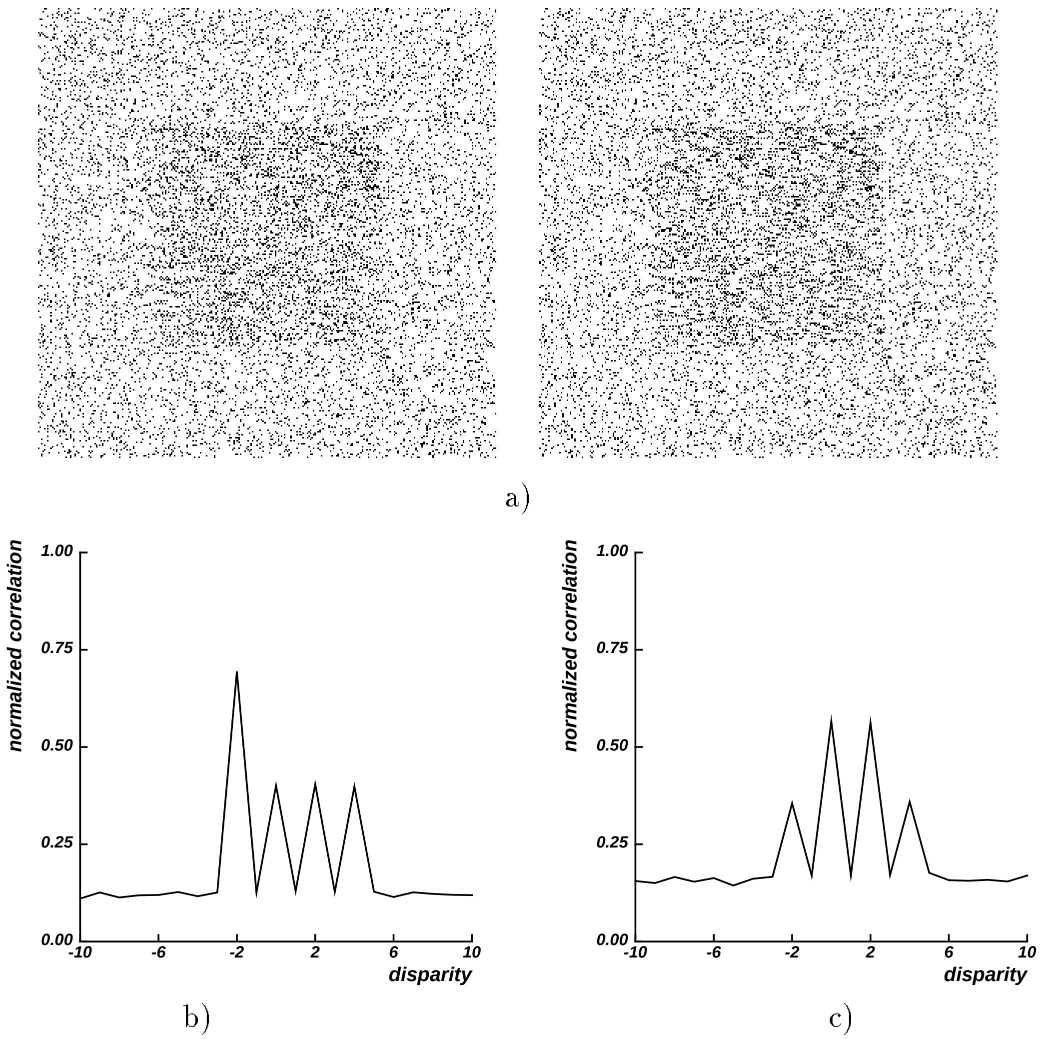


Figure 6: a) An ambiguous stereogram with $G_r = 2$ pixels and $G_l = 4$ pixels (as in figure 3a), where the number of points in the two internal planes have been doubled with new unambiguous points. b) The normalized correlation between the left and the right images at their center for a stereogram similar to the one used in experiment 3 (first type). c) The normalized correlation for the stereogram in (a), similar to the one used in experiment 3 (second type).

planes only, and three of them identified a single layer at the depth of one of the two. This agrees with the gap resolution of more than 2 min of arc required to distinguish between two planes in a RDS reported in Stevenson et al. (1989)[12]. In the second stereogram none of the subjects identified four layers but four subjects have identified three layers. Figure 7 summarizes the results for all the different conditions. One plot depicts how often 4 layers were identified, the second plot depicts how often three layers or more were identified, and the third plot depicts how often at least one of the ghost layers was identified.

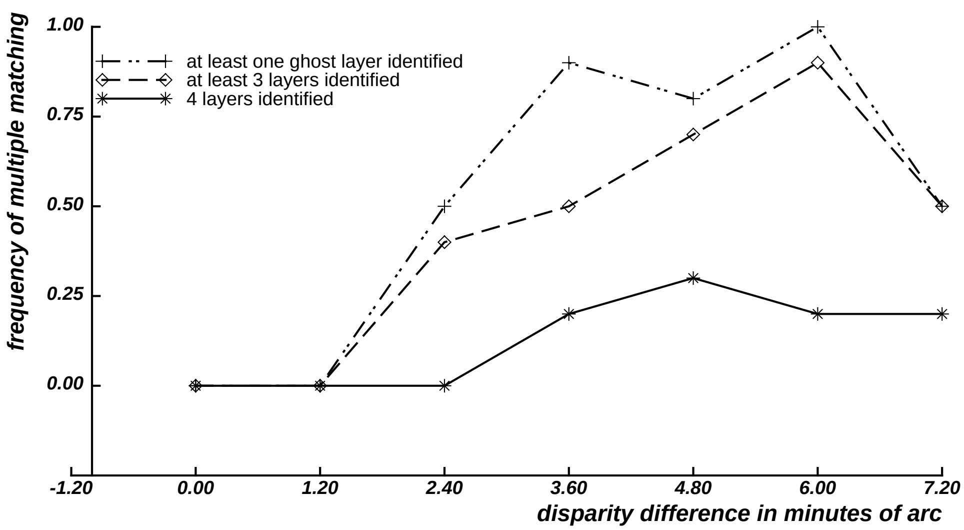


Figure 7: Three plots illustrating different aspects of the multiple matching effect as a function of the disparity difference between the two ordered matches. Multiple matching starts to take place when the difference is larger than 1 pixel (larger than 1.2 min of arc). The lower plot depicts how often four layers were identified, the intermediate plot depicts how often three layers or more were identified, and the upper plot depicts how often at least one of the ghost layers was identified. The results are pooled over all the subjects and all repetitive experiments (there were three stereograms with disparity difference 0).

Discussion:

Experiments 1-4 show that the shape of the correlation function between the left and right images (the number of peaks and their heights) does not determine alone what and how many disparities will give rise to the perception of distinct transparent surfaces. Can we conclude that it is not possible that a hypothesized human stereo vision algorithm uses only the correlation between images to select disparity? The last two experiments were designed to check this point, and also to check the effect of accidental matches on human perception.

3.5 Experiment 5: the reappearance of ghost planes

Results:

Two ambiguous stereograms, described in the general methods, were used in this experiment. The first was similar to the stereogram used in experiment 2 (figure 5), with the number of points in each image doubled by adding unmatched random dots (figures 8a and 8c). The second was similar to the stereogram used in experiment 2 (figure 5) with correlated points that could be matched at a new (fourth) disparity added to both images (figures 8b and 8d). There were two stereograms of each type, with interlayer disparity difference of 4 and 5 pixels. The disparities of the three possible layers relative to the background were (-4 0 4) and (0 5 10) respectively.

	1 layers	2 layers	3 layers	4 layers	5 layers	ghost plane
first stereogram (-4 0 4)	2	4	3	1	0	3
first control (0)	3	6	1	0	0	1
second stereogram (0 5 10)	0	6	1	2	1	9
second control (5)	2	7	1	0	0	7

Table 3: Results of the fifth experiment, first type. The first five columns of the table list how many observers identified one, two, three, four, or five layers respectively with each stereogram and its control. The last column lists how many subjects identified at least one "ghost" layer.

The two stereograms of the first type, where half of the points had no "good" match, were hard to fuse and very annoying to the subjects. The number and depth of layers identified by subjects therefore varied with time due to the rivalry in these stereograms. Another outcome of this rivalry was that subjects often did not follow the instructions and identified layers in depths where they have actually perceived only isolated points. Two stereograms, where half the points could be fused unambiguously to give a single plane (at disparities 0 and 5 respectively) and half the points could not be fused, were used as control. Table 3 summarizes the results for this experiment. The average number of layers identified in the test stereograms (2.31) was twice the average number of layers identified in the control (1.85). However, the disparity of most of these layers did not match the disparity of any of the ghost surfaces, as can be seen in figure 9. It seems that the reappearance of the suppressed layers happened almost as often in the control stereogram (where there were no suppressed layers) as in the test stereogram. (It is possible, though, that the rivalry and the difficulty of fusing these stereograms decreased the accuracy of depth judgement.)

In the stereograms of the second type unambiguous points were added at disparity 8 to the first stereogram (with solutions at disparities (-4 0 4)) and disparity 13 in the second stereogram (with solutions at disparities (0 5 10)). Table 4 summarizes the results for this

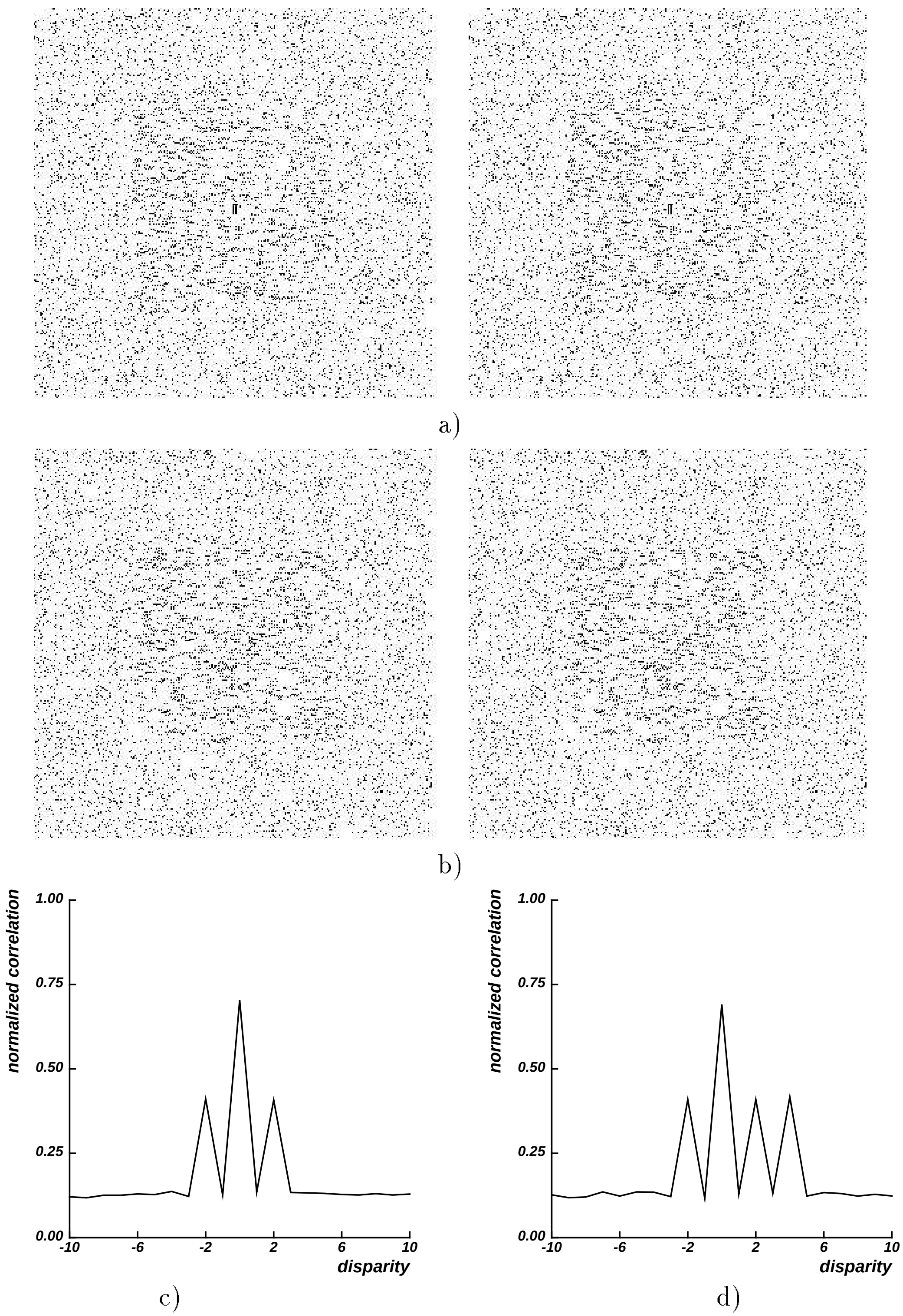


Figure 8: a) An ambiguous stereogram as in figure 5a, where the number of points in each image is doubled with additional random unmatched points (noise); b) An ambiguous stereogram as in figure 5a, with additional points that lie on a different plane at disparity 4, double the disparity of the suppressed planes (2 and -2). c) The normalized correlation for the stereogram in (a). d) The normalized correlation for the stereogram in (b).

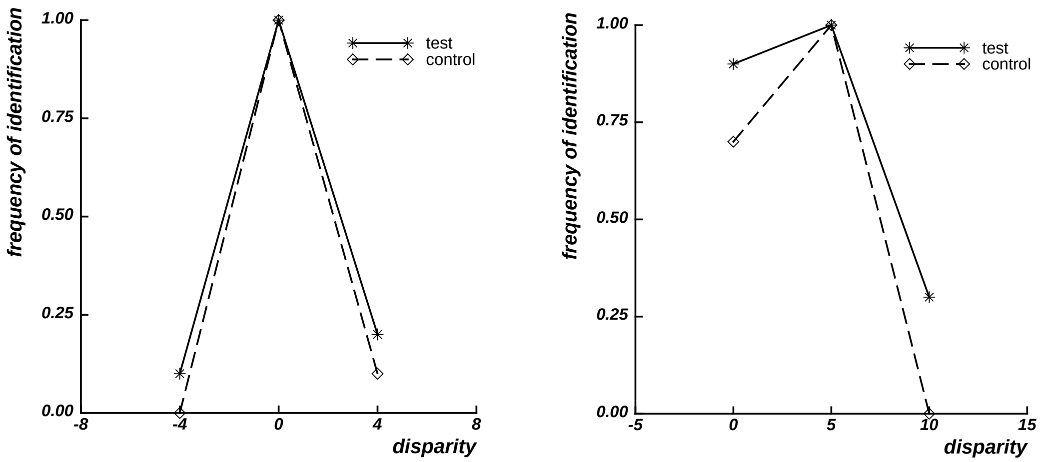


Figure 9: The frequency of identification of the three possible solutions in stereograms like in experiment 2 with 50% random matches. The results for two different stereograms are plotted. One with solutions at disparities (-4 0 4) (left), and one with solutions at disparities (0 5 10) (right). The identification of layers at the same disparities is plotted in dashed lines for the control stereograms that had solutions at disparities 0 (left) and 5 (right) only.

	1st layer -4/0	2nd layer 0/5	3rd layer 4/10	4th layer 8/13
first stereogram (-4 0 4 8)	0	10	0	10
first control (0 8)	0	10	0	10
second stereogram (0 5 10 13)	6	10	0	10
second control (5 13)	3	10	0	10

Table 4: Results of the fifth experiment, second type. The four columns of the table list how many observers identified each of the four possible layers, where first layer refers to the lowermost layer in the particular stereogram (-4 in the first and 0 in second), and fourth refers to the uppermost layer in the particular stereogram (8 in the first and 13 in the second).

experiment. None of the ghost planes reappeared in the first stereogram. In the second stereogram, the ghost layer at disparity 0, the disparity of the background, was identified by six subjects. However, three of these subject identified this layer also in the control, where it did not exist. Thus it seems that a layer at disparity 0 was interpolated from the background due to the existence of accidental matches at disparity 0. (Figure 9–right shows that the background had a tendency to reappear also in the previous condition, where random points were added to the stereogram, probably for the same reason.)

Discussion:

In the two conditions described above, where accidental matches were artificially added to the solutions of the second experiment, additional layers could be identified, some of them at the depth of the suppressed (ghost) solutions. However, in comparison with the perception in the control stereograms, it is not clear that the perception of additional layers corresponded to the reappearance of the suppressed surfaces.

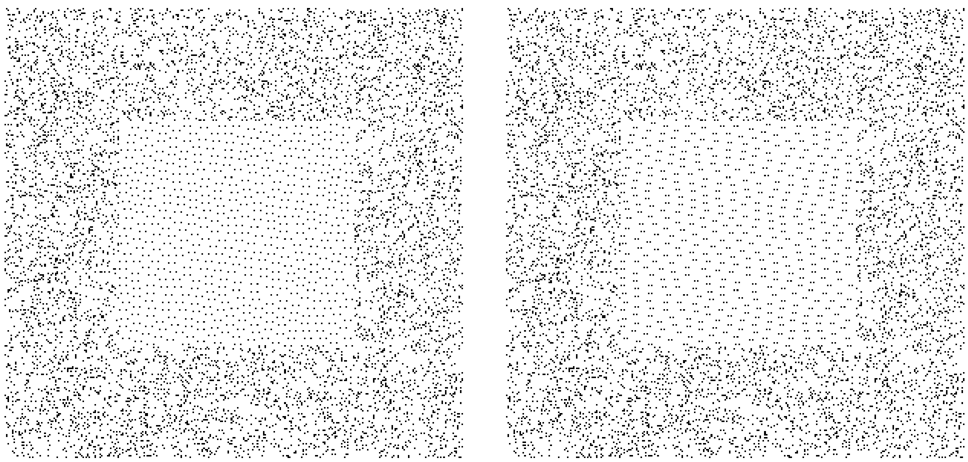
3.6 Experiment 6: surface competition

Results

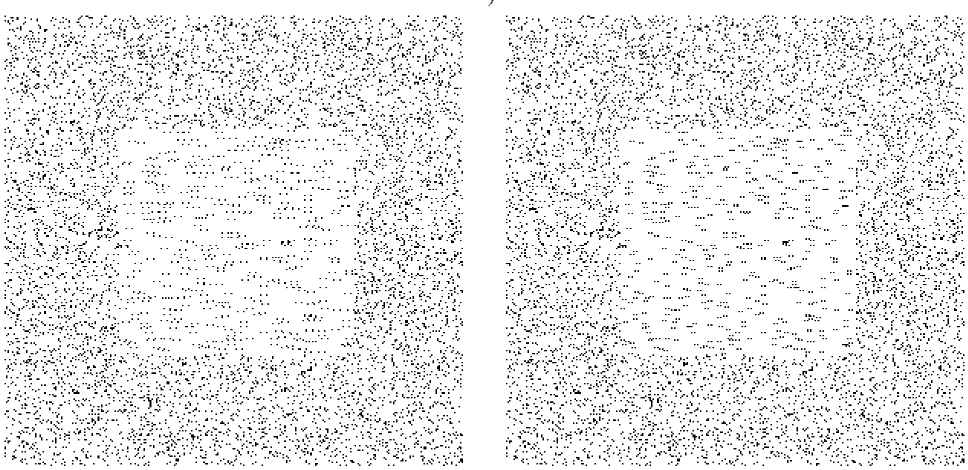
This experiment was designed to check the effect of accidental matches on human perception in these ambiguous stereograms. Accidental matches are matches due to the "usual" ambiguity in a RDS: matches between a dot from one micropattern in one image to a dot from another micropattern in the other image. Stereograms of the type used in the first experiment (figure 3a) were used in this experiment, with accidental matches eliminated as much as possible. This was achieved by spacing the points horizontally and vertically so that the disparity of possible accidental matches were larger in absolute value than the range of possible ambiguous disparities (figure 10a). Stereograms with the same relatively low density of points, where accidental matches had not been specifically eliminated, served as control (figure 10b).

Two such stereograms were used in this experiment in addition to two controls. The disparities of the four possible layers were (-4 0 4 8) in the first stereogram (figure 10a), and (-9 -3 3 9) in the second stereogram. The density of points in the control was identical to the density of points in the test stereogram, about a quarter of the original density (figure 10b). Table 5 summarizes the results. Note that most of the subjects who had identified three layers in the test stereograms did not report seeing them simultaneously. It is possible that a third layer could be seen only with the help of the probe and eye vergence to flip between competing surfaces.

Figure 11 shows the effect of density and accidental matches on the perception of the ghost layers. We see a deterioration in the ability of subjects to perceive multiple matchings under all three criteria used before, namely, how often four layers are perceived, how often more than three layers are perceived, and how often a ghost layer is perceived. This deterioration is partially explained by the decrease in point density, but density does not seem to explain



a)



b)

Figure 10: a) Stereograms of the type used in experiment 1, with solutions at disparities (-4 0 4 8), and where accidental matches were eliminated as much as possible; b) Control: stereograms of the type used in experiment 1, with solutions at disparities (-4 0 4 8), and where accidental matches have not been especially eliminated. The overall density of points in (a) and (b) is identical.

	1 layers	2 layers	3 layers	4 layers	ghost plane
first stereogram (-4 0 4 8)	3	5	2 (0)	0	4
first control (-4 0 4 8)	0	6	3 (1)	1	5
second stereogram (-9 -3 3 9)	1	6	2 (0)	1	4
second control (-9 -3 3 9)	0	2	7 (6)	1	10

Table 5: Results of the sixth experiment. The first four columns of the table list how many observers identified one, two, three or four layers respectively with each stereogram and its control. The last column lists how many subjects identified at least one "ghost" layer. The number in parentheses in the three layers column indicates how many subjects reported seeing three layers simultaneously, in addition to correctly identifying the three layers.

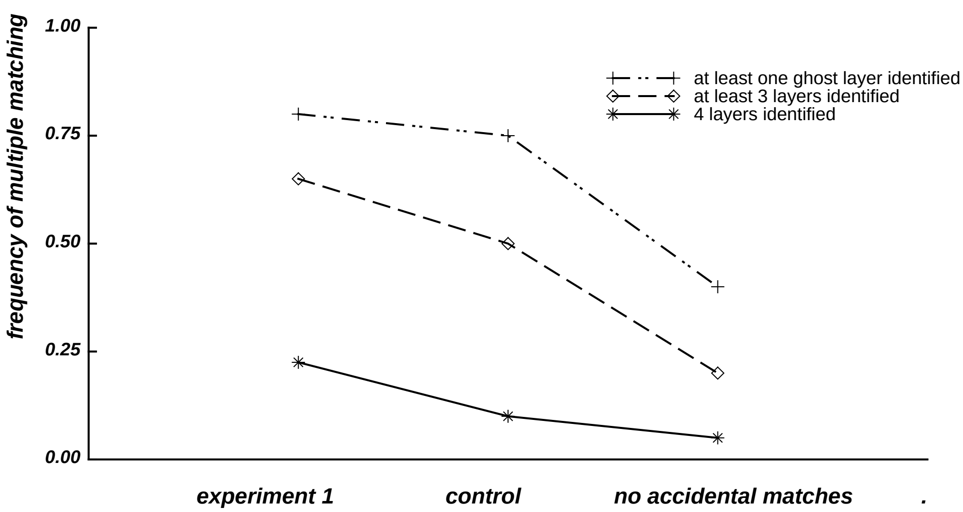


Figure 11: Three plots illustrating different aspects of the multiple matching effect for three conditions (listed on the abscissa). The data for the first condition (experiment 1) is the average result pooled over all the stereograms used in the first experiment, with 9% points density. The data for the second condition were taken from the control of experiment 6, where stereograms with roughly 2% dot density were used. The data for the last condition were taken from the results of experiment 6, using stereograms with roughly 2% dot density and almost no accidental matches at the four solutions. The lower plot depicts how often four layers were identified, the intermediate plot depicts how often three layers or more were identified, and the upper plot depicts how often at least one of the ghost layers was identified. The results are pooled over all the subjects and all the repetitions of the same conditions.

all the deterioration in the no accidental matches condition (compare the results in the control to the results in the test stereogram). Only one subject could see the ghost layers simultaneously with the other layers in one of the stereograms lacking accidental matches.

4 General discussion

What do the results reported here have to say about theories of human stereo matching? Very crudely, a stereo matching algorithm can use one of two different strategies: it can either look at a whole region around a dot and compare it to similar regions in the other image (e.g. Nishihara (1984) and Drumheller and Poggio (1986)), or it can look at localized features and match them to similar localized features in the other image. Most stereo algorithms use some intermediate strategy: they match localized features but let neighbors influence the selection of a disparity at each feature to enforce the smoothness constraint on the disparity field (e.g. Marr and Poggio (1979) and Mayhew and Frisby (1981)). In the following I use the correlation between image patches to study the first strategy, and a detailed analysis of the interactions within the micropattern to study the second strategy.

The first challenge to stereo matching algorithms is posed by experiment 1. In this experiment I used a stereogram composed of ambiguous micropatterns. When the dots of the micropattern are presented in isolation (such as in the double nail illusion stimulus), the ghost dots are never perceived. However, most subjects identify at least one ghost plane when a stereogram composed of these micropatterns is presented. Thus a stereo matching algorithm should resolve the ambiguity differently for the isolated micropattern and the RDS.

The second challenge to stereo matching algorithms concerns the interactions between dots of neighboring micropatterns in the RDS. Experiments 3 and 5 show that the information in the correlation between neighboring dots, computed over very large regions, is not sufficient to explain human perception. I have shown two stereograms with very similar correlation between the left and right images, yet humans perceived four transparent surfaces in one (experiment 3) and two in the other (experiment 5). Thus using neighboring features to enforce smoothness by a stereo matching algorithm may not be sufficient.

Whether correlation or weak local interactions are used, it seems that human perception in experiments 1-5 for both isolated features and RDS's cannot be explained simultaneously. This is the basic computational problem posed by the experiments described in this paper to any stereo matching algorithm that attempts to model human perception. In the rest of this section I will discuss in detail why some stereo matching algorithms fail, and how they can be modified and tuned to be able to explain human perception under the conditions of the experiments described here. A more quantitative discussion will be given in a paper in preparation.

4.1 Local interactions

Let me first consider the interactions among the pairings of dots in the micropattern in isolation. Take the micropattern of experiment 1 shown in figure 12a. L_1 can be matched to R_1 with disparity 0 and R_2 with disparity 2. Thus its list of possible disparities is {0,2}. L_2 can be matched to R_1 with disparity -4 and R_2 with disparity -2. Thus its list of possible disparities is {-4,-2}. The two closest (thus the smoothest) disparities are 0 for L_1 and -2 for L_2, and they correspond to a unique matching also. Thus an algorithm that looks for a unique and smooth matching chooses the ordered matches [L_1; R_1] and [L_2; R_2]. This is what happens when isolated features are presented to human subjects. However, this argument does not predict human perception of four planes at disparities {-4,-2,0,2} in the RDS of experiment 1.

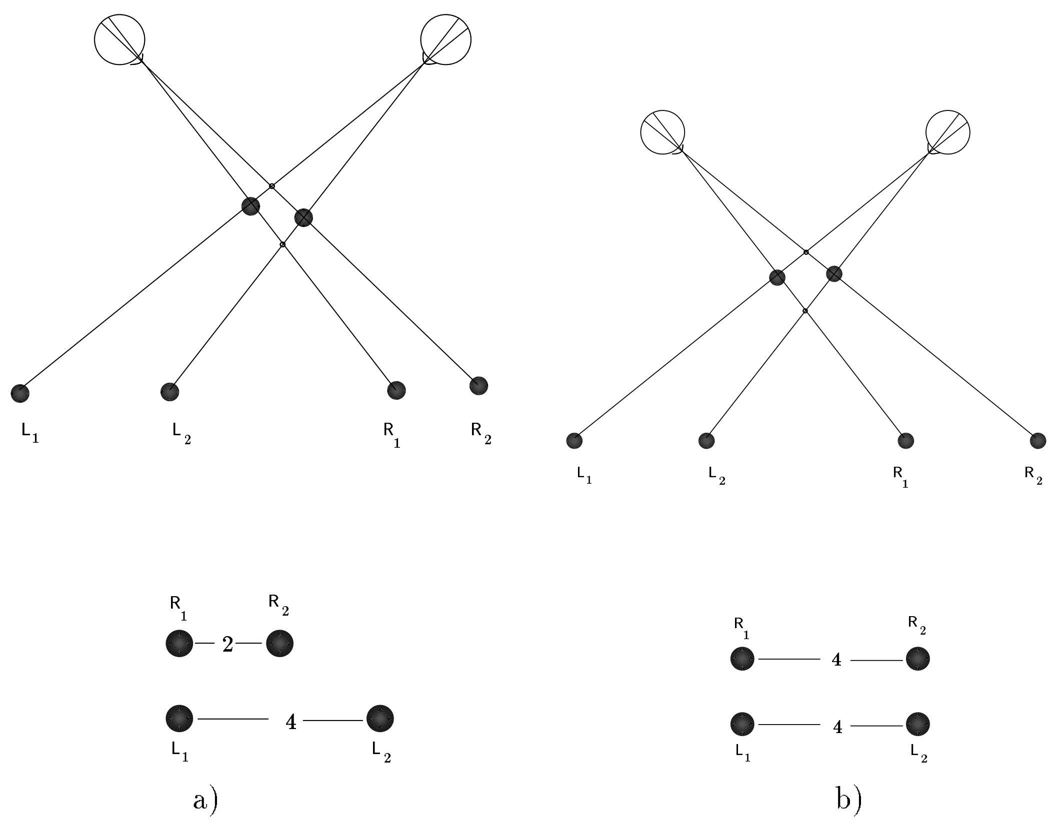


Figure 12: Above: The micropattern of experiment 1 (a) and experiment 2 (b). Below: the micropattern in the right image is shown above the micropattern in the left image. The horizontal coordinate agrees with the location of each dot in the image, the vertical coordinate is indented for the right image (which has been moved up) for illustration purposes, since the two pairs of the micropattern have the same vertical coordinate.

Take the micropattern of experiment 2 shown in figure 12b. L_1 can be matched to R_1 with disparity 0 and R_2 with disparity 4. Thus its list of possible disparities is {0,4}. L_2 can be matched to R_1 with disparity -4 and R_2 with disparity 0. Thus its list of possible disparities is {-4,0}. The two most similar (thus the smoothest) disparities are 0,0, and they correspond to a unique matching also. Once again, an algorithm that looks for a unique and smooth matching chooses the ordered matches [L_1; R_1] and [L_2; R_2]. This is what happens when isolated features are presented to human subjects. This argument now predicts human perception of only one plane at disparity 0 in the RDS of experiment 2.

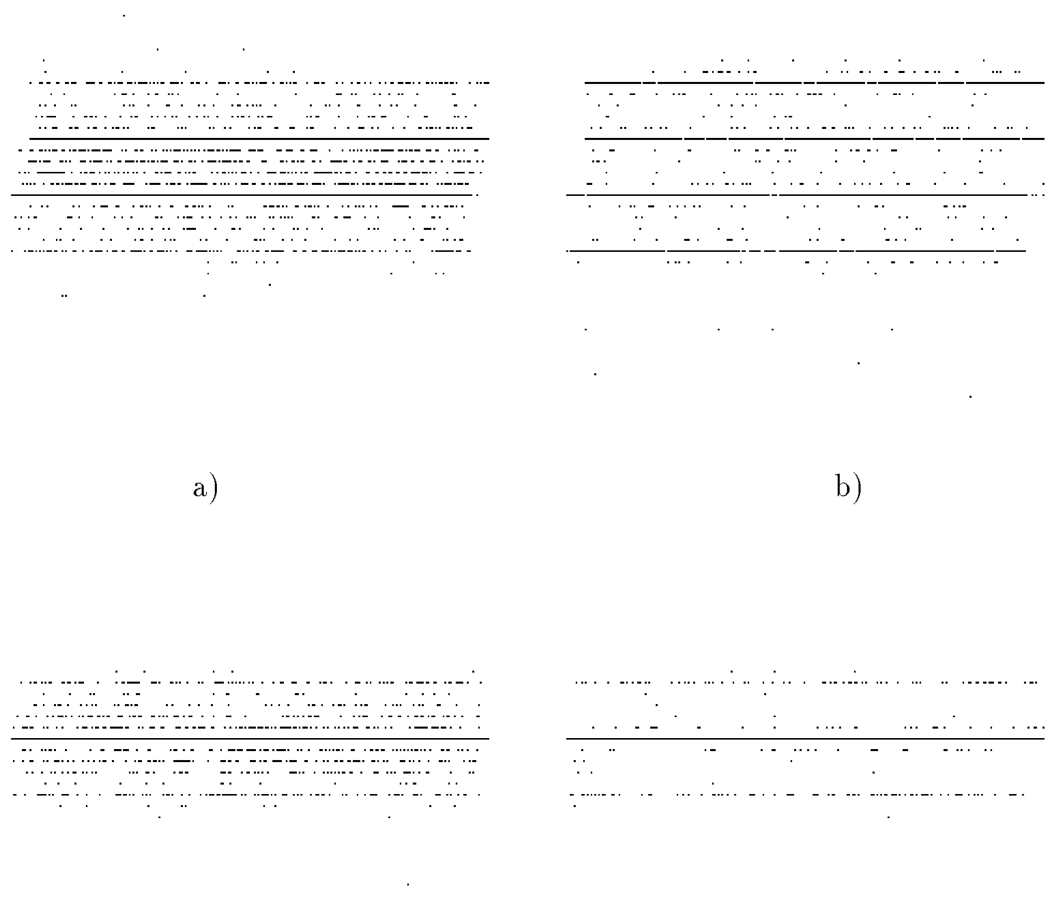
Another possible local interaction is no interaction at all, namely, the disparity at each dot is selected independently of its neighbors. Such an algorithm will select any of disparities {-4,-2,0,2} at random in figure 12a, predicting the perception of four surfaces at those disparities, in agreement with human perception in experiment 1. On the other hand, in figure 12b this algorithm will detect disparities {-4,0,4}, predicting the perception of three surfaces in disagreement with human perception. Such an algorithm fails to explain human perception when isolated features are presented to human subjects under the conditions of both experiments.

This discussion suggests that, when the interactions inside the ambiguous micropattern are considered in isolation, human perception in experiments 1 and 2 cannot be explained by either algorithmic principle.

Pollard and Frisby (1990) have shown one way to solve this problem using interactions within the micropattern only. Basically, they require that the algorithm would enforce smoothness in figure 12b, but allow for random matching in figure 12a. It is possible in a not entirely *ad hoc* way if we notice that in figure 12a the supporting disparities are -2,0 and in figure 12b they are 0,0. Obviously 0 should support 0, but 0 may not support -2 if the algorithm restricts support when the difference in disparities is large. To implement this idea, Pollard and Frisby (1990) used their PMF matching algorithm (Pollard et al., 1985[10]) with disparity-gradient limit 0.5. The performance of the modified algorithm agrees with human perception for both experiments 1 and 2 (see figure 13b,d). The original PMF algorithm, where the disparity-gradient limit was fixed to 1, predicts the perception of only two surfaces in experiment 1, in disagreement with human perception (see figure 13a,c). [2] Other stereo matching algorithms, such as Prazdny's (1985), can be tuned in a similar way.

The particular implementation proposed by Pollard and Frisby (1990) fails because of its reliance on local interactions and a fixed threshold. Consider two nails arranged as the micropattern in the first experiment (figure 12a). The original PMF, with disparity-gradient limit of 1, assigns the ordered disparities to both nails, in agreement with human perception. The modified PMF, with disparity-gradient limit of 0.5, cannot prefer any disparity for any of the two nails since the disparity gradient between any pair of possible solutions is larger than 0.5. Thus a ghost nail is as likely to be detected by this algorithm as an ordered nail, in

[2] The disparity gradient limit in the PMF algorithm was fixed to 1 when used by its authors as a model of human stereo matching algorithm. They based the use of a fixed threshold and its value on psychophysical data showing that two features can be fused simultaneously if the disparity gradient between them is smaller than 1 (Burt and Julesz (1980)[3]).



a) b)

c) d)

Figure 13: The output of the PMF algorithm operating on an ambiguous stereogram used in experiment 1 (a,b) and experiment 2 (c,d). The output of the algorithm is shown in the following way: for a given disparity, the output is a binary array where 1 indicates a match was found at the given disparity (0 otherwise). This array is condensed to a line in the following way: a point (*,j) is painted black if there exist any point (i,j) in the output array that has the value 1 (has a match). In the above plots, such condensed description are shown for different disparities one on top of the other. That is, the disparity varies along the vertical dimension. a) The result for a stereogram as in figure 3a obtained by using the original PMF (disparity gradient limit of 1). The number of dots in the ghost disparities is below noise threshold, less than 10% of the number of dots in the ordered disparities. b) The result for a stereogram as in figure 3a obtained with disparity gradient limit of 0.3. Here the number of dots in the ghost disparities is above noise threshold, about 80% of the number of dots in the ordered disparities. c) The result for a stereogram as in figure 5a obtained by using the original PMF (disparity gradient limit of 1). d) The result for a stereogram as in figure 5a obtained with disparity gradient limit of 0.3.

disagreement with human perception. (Although Pollard and Frisby argue that interactions between neighboring micropatterns are the key to the success of their algorithm, I believe the example I have discussed disproves this.) On the other hand, Prazdny's stereo matching algorithm (1985) can be tuned to agree with human perception in both experiments 1 and 2 for isolated features and a RDS. How these algorithms deal with experiments 3-5 is discussed next.

4.2 Long range interactions

I have shown two stereograms whose cross-correlation was almost identical, yet humans perceived two transparent planes in one case (figure 8d) and four transparent planes in the other (figure 6b). The correlation plotted in the figures was computed over very large regions. The question that should be asked before we conclude that a correlation based algorithm cannot explain human perception in these experiments is whether local effects, that can be measured by the correlation over small image patches, can distinguish the two cases. When local correlations in these stereograms are studied, differences are indeed observed.

I have used a simple area-based algorithm that selects at each feature the disparity that maximizes the cross-correlation between the 2 images over a certain window around that feature. As expected, the algorithm failed to agree with human perception when a large window was used, but its behaviors changed when the window size was decreased. When a small window (around 9x9 pixels) was used, and when a threshold on the minimal number of features that can define a distinct surface was set, human perception could be replicated for certain threshold values.

This analysis suggests that area-based stereo matching algorithms, such as Nishihara (1984) or Drumheller and Poggio (1986), can be tuned and adjusted to replicate human perception in all the experiments discussed in this work. These experiments strongly constrain the parameters of the algorithm (in particular the size of the neighborhood of interaction) and the way its results are interpreted to indicate distinct surfaces. Experimental predictions of these constraints can be checked. The same constraints should be enforced on any algorithm that is suggested as a model of human vision, such as PMF or Prazdny's. Both algorithms, in particular Prazdny's, require a rather specific tuning of the interaction parameter and interpretation threshold, since they distinguish the two stereograms less well than the local correlation.

5 Summary

I have shown ambiguous stereograms where the number of perceived surfaces corresponded in some cases to a unique matching, in others to nonunique matching. The global correlation between the left and right images could not explain these results. In the discussion, I have argued that existing stereo matching algorithms cannot explain all the results discussed in this and related works. I then showed how some algorithms can be modified, constrained and tuned to replicate human perception. The experiments reported in this work can help screen stereo algorithms that are inappropriate as models of human stereo matching, and strongly constrain those algorithms that can be modified to survive this screening.

Acknowledgements: I thank S. Kirkpatrick, G. Mitchison, T. Poggio and M. Fahle for very helpful comments and advice. I also thank S. Barash, H. Bülthoff, N. Cornelius, S. Edelman, J. McFarland, M. Ross, A. Shashua, B. Snowden and L. Topa for their help, and the anonymous referees for many useful comments. This research was done partly in the MIT AI Laboratory. It was supported by a Fairchild postdoctoral fellowship, and in part by grants from the office of Naval Research (N00014-88-k-0164), from the National Science Foundation (IRI-8719394 and IRI-8657824), and a gift from the James S. McDonnell Foundation to Professor Ellen Hildreth.

References

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

[2] A. Blake and A. Zisserman. *Visual reconstruction.* MIT Press, Cambridge, MA, 1987.

[3] P. Burt and B. Julesz. A disparity gradient limit for binocular fusion. *Science*, 208(9):615–617, 1980.

[4] W. E. L. Grimson. *From Images to Surfaces.* MIT Press, Cambridge, MA, 1981.

[5] J. D. Krol and W. A. van de Grind. The double nail illusion: experiments on binocular vision with nails, needles, and pins. *Perception*, 9:651–669, 1980.

[6] D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, 194:283–287, 1976.

[7] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proceedings of the Royal Society of London B*, 204:301–328, 1979.

[8] J. E. W. Mayhew and J. P. Frisby. Psychophysical and computational studies towards a theory of human stereopsis. *Artificial Intelligence*, 17:349–386, 1981.

[9] H. K. Nishihara. Practical real-time imaging stereo matcher. *Optical Engineering*, 23(5):536–545, 1984.

[10] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby. A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14:449–470, 1985.

[11] K. Prazdny. Detection of binocular disparities. *Biological Cybernetics*, 52:93–99, 1985.

[12] S. B. Stevenson, L. K. Cormack, and C. M. Schor. Hyperacuity, superresolution and gap resolution in human stereopsis. *Vision Research*, 29(11):1597–1605, 1989.

[13] D. Weinshall. Perception of multiple transparent planes in stereo vision. *Nature*, 341:737–739, 1989.

[14] A. L. Yuille and T. Poggio. A generalized ordering constraint for stereo correspondence. Technical Report AIM–777, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1984.