The Computation of Multiple Matching in Doubly Ambiguous Stereograms with Transparent Planes

Daphna Weinshall*
Center for Biological Information Processing, E25-201
MIT, Cambridge MA, 02139

Abstract

I have previously described psychophysical experiments that involved the perception of many transparent layers, corresponding to multiple matching, in doubly ambiguous random dot stereograms. Additional experiments are described in the first part of this paper. In one experiment, subjects were required to report the density of dots on each transparent layer. In another experiment, the minimal density of dots on each layer, which is required for the subjects to perceive it as a distinct transparent layer, was measured. The difficulties encountered by stereo matching algorithms, when applied to doubly ambiguous stereograms, are described in the second part of this paper. Algorithms that can be modified to perform consistently with human perception, and the constraints imposed on their parameters by human perception, are discussed.

1 Introduction

The depth of 3D objects is lost in the optical projection process. Stereo vision, in which two simultaneous images of the same scene are recorded in the two eyes, can be used to recover the lost depth. In computational stereo 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 order to solve the matching ambiguity at each feature in the image, neighboring features can be used. It is generally assumed that many neighboring features should have a match at about the same disparity for a matching to be plausible. Different stereo matching algorithms differ in how they implement this neighborhood interaction (or smoothness constraint), among other things.

I have previously described [9, 10] psychophysical experiments whose results could not be readily explained by existing stereo matching algorithms. In these experiments, subjects

[0]Present address: Department of Computer science, Hebrew University of Jerusalem, 91904 Jerusalem, ISRAEL.

were presented with doubly ambiguous stereograms (defined in Section 2.1). In some cases a few transparent surfaces were perceived corresponding to multiple matches, in other cases transparent surfaces corresponding to unique matches were perceived. Some stereograms were constructed to have the same cross-correlation between the left and right images, yet different numbers of transparent layers were perceived. The results of these experiments are briefly summarized in Section 2.

In Section 3, additional experiments are described. First, subjects were required to report the density of dots on each transparent layer of a doubly ambiguous stereogram by adjusting the density of dots on three simple transparent layers. In another experiment, the minimal density of dots on each layer, which is required for the subjects to perceive it as a distinct transparent layer, was measured. These experiments were designed to clarify which algorithmic principle can be used to explain the results in the experiments with doubly ambiguous stereograms.

In Section 4, the difficulties encountered by stereo matching algorithms, when applied to doubly ambiguous stereograms, are discussed. Three simple matching algorithms, representing two different simple matching principles, are discussed in detail: a patch-wise correlation algorithm (e.g. [5, 2]), Prazdny's matching algorithm [8], and PMF [7]. For comparison with human data, an additional stage was added to each algorithm, where the matching results were used to determine how many transparent layers exist in the image. The range of parameters for which the performance of these algorithms was consistent with that of humans, and the sensitivity of their tuning, is discussed.

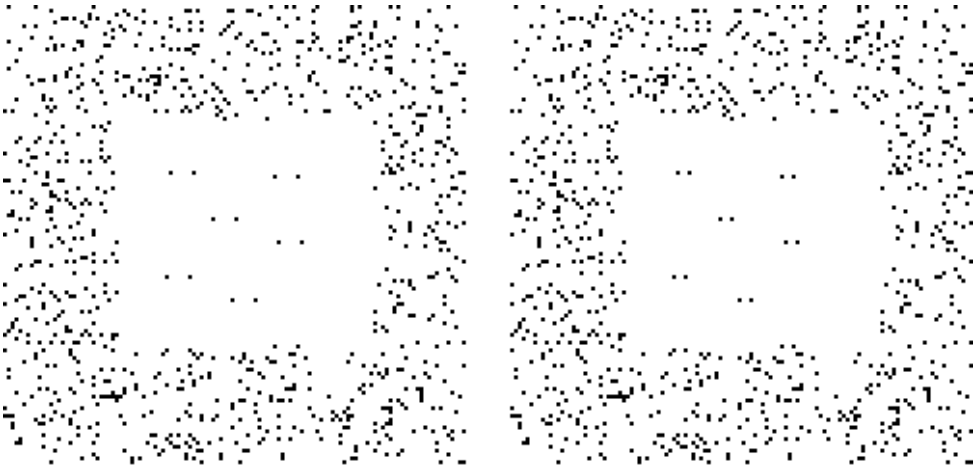
2 Multiple matching in ambiguous stereograms:

2.1 Doubly ambiguous stereograms

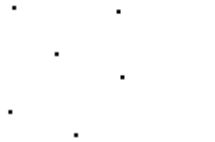
In a doubly ambiguous Random Dot Stereogram, a sparse random pattern (Fig. 1b) is copied twice in each image (Fig. 1a). The horizontal gap between the two copies is G_r pixels in the right image and G_l in the left image. Each dot of the original sparse random pattern (Fig. 1b) has two copies in each image. All these pairs, which are the **micropattern** of the doubly ambiguous RDS (Fig. 1c), are the same instance of the double nail illusion stimulus [4]. There are four possible matchings of the elements of the micropattern that are equally plausible, two mutually exclusive pairs if matching is unique (namely, a point can only be matched to a single point in the other image, as shown by full and hollow circles in Fig. 1c).

2.2 Summary of previous results

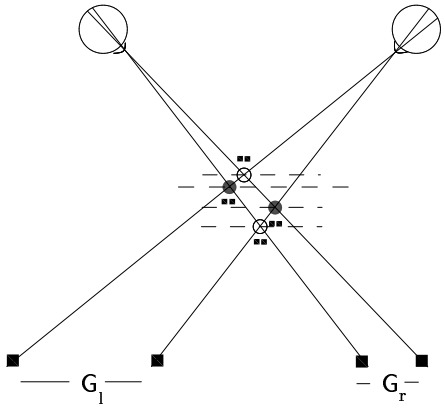
In an unpublished study, Braddick presented subjects with ambiguous stereograms that were, in effect, a special case of the doubly ambiguous stereograms described in Section 2.1. In these stereograms, the generating pattern was copied twice only in one image, equivalent to choosing $G_r = 0$ or $G_l = 0$. The micropattern of such a stereogram, one dot in one image and



a)



b)



c)

Figure 1: *a)* A doubly ambiguous random dot stereogram. *b)* The sparse random pattern that is used to generate the doubly ambiguous stereogram in *a*. For illustration purposes, the density of the sparse pattern is reduced; in the actual experiment it was equal to the density of the background. *c)* The enlarged micropattern of the RDS in *a*, where the two pairs of matches that are mutually exclusive if matching is unique are separately marked by filled and hollow circles.

two dots in the other image, is also known as Panum's limiting case. When presented with Panum's limiting case, subjects' perception corresponds to matching the single dot in one image to both dots in the other image[1] (if the disparity difference between the dots is within Panum's limiting area, which is the range of disparities that can be fused simultaneously). When viewing a stereogram composed of such micropatterns, the perception was similar: subjects reported seeing two transparent surfaces, corresponding to a multiple matching of the generating pattern.

2.3 Multiple matching

In the first experiment, an ambiguous stereogram of the type described in Section 2.1, with $G_r \neq G_l$ and dot density of 9% (of the generating pattern), was used. Subjects identified up to four transparent layers, corresponding to all four possible matches of the micropattern dots (Fig. 1c). The disparity differences between the transparent layers were approximately 6 minutes of arc. The smaller the differences were, the easier it was to see the layers simultaneously, but the harder it was to distinguish them in depth.

In the following discussion, the correlation function between the left and right images is computed at the center of the right image for a varying horizontal disparity D (vertical disparities are neglected). A window of width W around the center of the right image, covering all the generating pattern, and a similar window around the center of the left image, translated horizontally by D, were correlated according to the following equation:

$$\text{correlation} \;=\; \sum_{i}^{W}\sum_{j}^{W} I(x_i^R, y_j^R) I(x_i^L + D, y_j^L).$$

where $I(x_i, y_j)$ is the image intensity at location (x_i, y_j) of the right (superscript R) or left (superscript L) image.

The correlation function between the left and right images of a stereogram of the type used in this experiment is given in Fig. 2a. There are four peaks in this function, corresponding to the disparities of the four transparent layers that were seen. This result seems to suggest that all peaks in the correlation function give rise to the perception of distinct transparent layers.

2.4 Unique matching

In the second experiment, an ambiguous stereogram of the type described in Section 2.1, with $G_r = G_l$, was used. Two of the four possible disparities of the micropattern (Fig. 1c) are identical and therefore the correlation between the left and right images (Fig. 2b) has only three peaks. The conclusion given in the previous section predicts that three transparent layers will be identified in this case, corresponding to the three peaks. However, subjects

[1] This is different from their perception when presented with the micropattern of a doubly ambiguous stereogram, as will be discussed in Section 3.3

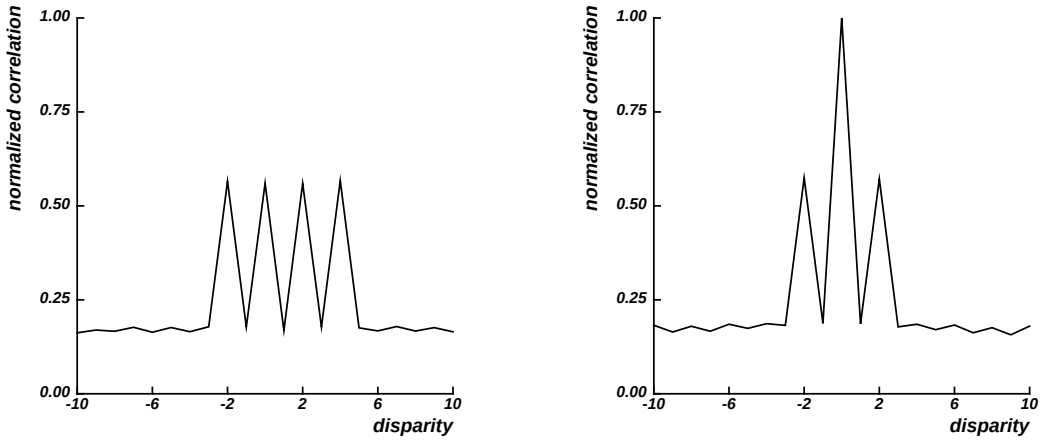


Figure 2: The correlation (as a function of horizontal disparity) between the left and the right images of doubly ambiguous RDS's. The correlation window was equal in size to the generating pattern.

identified only one opaque surface, whose disparity corresponded to the maximum correlation in Fig. 2b. Thus not all the local maxima in the correlation function give rise to the perception of distinct transparent layers.

2.5 The information in the correlation function

There is one difference between the correlation functions plotted in figures 2a,b that may explain the difference in human perception. The disparity of the single opaque plane perceived in the second experiment corresponds to the global maximum of the correlation function in Fig. 2b, whereas the correlation function in Fig. 2a has four identical maxima. Additional experiments showed, however, that this difference cannot explain human perception.

In one experiment, two stereograms, whose correlation functions are given in figures 3a,b respectively, were presented to subjects. The stereogram corresponding to Fig. 3a was similar to the stereogram corresponding to Fig. 2a, with additional dots that could all be matched at disparity -2. In this case subjects identified up to four transparent layers. The stereogram corresponding to Fig. 3b was similar to the stereogram corresponding to Fig. 2b, with additional dots that could all be matched at disparity 4. In this case subjects identified only two transparent layers.

These experiments show that the correlation between the left and right images cannot account for the subtleties of human perception. This conclusion is valid for the "average" correlation, where the correlation is computed over a large region around a point (of an order of magnitude of the whole image). As the window of the correlation function decreases, the variability in the correlation function computed at each dot increases when compared with

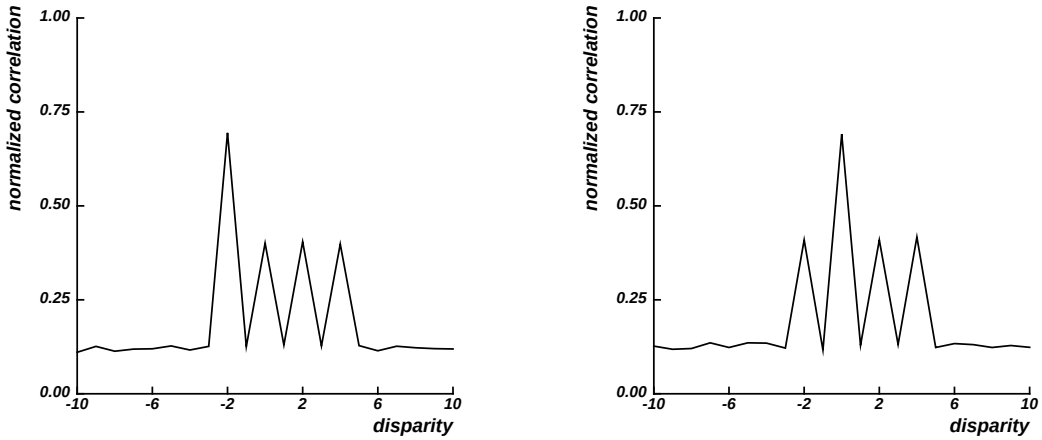


Figure 3: The correlation (as a function of horizontal disparity) between the left and the right images of doubly ambiguous RDS's.

the average correlation in figure 3, due to local perturbations and short-range correlations. As we shall see later, when the window is sufficiently small, this variability may be sufficient to account for human perception.

3 Additional experiments

The following experiments were designed to help clarify which kinds of stereo matching algorithmic principles can more readily explain the results in the experiments described in Section 2.

3.1 Experiment 1: the density of the transparent layers

In the experiment described in Section 2.3, most subjects identified three to four transparent layers. This perception corresponds to multiple matching of the generating sparse pattern of the stereogram, which can be matched as a whole to one of its copies in the second image with a single disparity, leading to the simultaneous perception of only two transparent layers. This multiple matching of the generating pattern can be implemented by the assignment of a unique disparity to each dot in the generating pattern, where some of the dots are assigned one disparity and the rest are assigned another disparity. Alternatively, it may be that multiple matches are assigned to each dot in the generating pattern, since there are at each dot two disparities that are equally supported by neighboring dots. Both solutions of the matching problem lead to the identification of four transparent layers.

The present experiment was designed to choose one of these two explanations for the multiple matching results.

Methods:

Five subjects participated in this experiment. They were presented with one of the doubly ambiguous stereograms described in Section 2.3. Adjacent to this stereogram, another stereogram of three transparent layers was presented[2], where the height of the three transparent layers matched the height of the top three ambiguous layers of the ambiguous RDS. The subjects were asked to modify the density of dots on each layer of the second (adjacent) stereogram, in steps of 1% or 4%, until the density of each transparent layer matched the density of the corresponding ambiguous transparent layer.

Density matching of transparent layers was initially quite difficult. The subjects started with two training sessions. In the first session they were asked to match the density of a single opaque layer to another single opaque layer. They received feedback, and ended this session when their matching was perfect. This session proved to be quite easy for everyone. In the second session of the training, the subjects had to match the densities of three transparent layers to the densities of three other transparent layers. This task was initially quite hard, but after a few trials and feedback, the subjects had learned to do this task and felt quite confident at being able to do it well. They stopped the second training session when the error in density matching per layer was less than or equal to 1%. One subject could not obtain this level of performance.

After finishing the two training sessions, subjects were presented with two doubly ambiguous RDS's of the type described in Section 2.3, where the densities of dots on the generating sparse pattern were 9% and 11% respectively. They had to match the densities of the top three ambiguous layers. Two of the four subjects were presented with a third stereogram, an ambiguous RDS's of the type described in Section 2.2, where the density of dots on the generating sparse pattern was 9%. In this case the subjects were asked to match the densities of two ambiguous layers. All the subjects said, after the experiment was over, that the density matching of the ambiguous layers was more difficult than the density matching of the three transparent layers in the training session.

Results:

Table 1 gives the data of the four subjects that were able to learn to do the density matching accurately enough, to within 1% error.

[2]In this stereogram each dot could be matched to any other dot in the image, the "usual" ambiguity, but the additional ambiguity created by doubling a certain generating pattern as described in Section 2.1 was eliminated.

3 layers (density 9%)			3 layers (density 11%)			2 layers (density 9%)	
Top	Middle	Bottom	Top	Middle	Bottom	Top	Bottom
5%	6%	5%	5%	6%	9%	6%	6%
2%	5%	8%	6%	6%	7%	4%	9%
5%	4%	6%	6%	3%	8%	–	–
5%	0%	7%	0%	5%	6%	–	–

Table 1: First two major columns give the density of dots on each of the top three ambiguous layers: top, middle and bottom, for the two stereograms described in the text, reported by four subjects. The last major column gives the density of dots on each of two ambiguous layers: top and bottom, for the third case described in the text, reported by two subjects.

Conclusions:

The hypothesis that each dot in the generating pattern is assigned a unique disparity predicts that the density matching results approach 4.5% per layer in the first stereogram, and 5.5% per layer in the second stereogram. This is based on the assumption that each dot is equally likely to be assigned one of its two ambiguous matches, but not both. The density of each layer should therefore approach half the density of the generating pattern. The hypothesis that each dot in the generating pattern is assigned multiple disparities predicts that the density matching results approach 9% per layer in the first experiment, and 11% per layer in the second experiment (identical to the density of the generating pattern).

The average density matching by the more accurate subjects (the first three rows in Table 1) was 5.1% per layer in the first experiment and 6.2% per layer in the second experiment. These reported densities are somewhat larger than the density predicted by the unique matching hypothesis, but much smaller than the density predicted by the multiple matching hypothesis[3]. Note that the subjects did not have to report the density of the fourth layer, which most subjects found difficult to see simultaneously with the other three layers. Consequently, the reported density can be expected to be higher than predicted.

The results with the third stereogram are interesting since when presented with the micropattern of this stereogram, people's perception corresponds to multiple matching of the dots (Section 2.2). The average density reported by subjects in this case, 6.25%, is larger than before (5.1%), but still intermediate between the prediction of the hypothesis of multiple matching at each dot (9%) and the prediction of the hypothesis of unique matching (4.5%), closer to the latter. This suggests that either this experiment does not measure

[3]I should note, however, that the results in a somewhat different experiment, where the subjects were asked to match the density on each transparent layer separately, were more ambiguous; in this experiment, the densities assigned to each layer were close to the average between the prediction of the multiple matching hypothesis and the prediction of the unique matching hypothesis. These results are not included in this paper since the task was harder for the subjects and the results were less reliable.

correctly the number of dots that are matched at each disparity, or that there is a difference between the matching of isolated features and the matching of images with texture.

The results of the density matching experiment, if indeed this experiment correctly measures the number of dots matched at each disparity, seem to support the hypothesis that a unique disparity is assigned to each dot of the generating sparse pattern of a doubly ambiguous stereogram, where some dots are assigned one disparity and some another disparity.

3.2 Experiment 2: the lowest density of the transparent layers

When looking at stereograms with transparent layers, ambiguous or not ambiguous, subjects reported seeing points floating in a range of depth values. Subjects were asked to report a layer when they subjectively perceived a layer. This could be a difficult decision for them in some cases. The present experiment was designed to identify the lowest density of dots at a given disparity, above which these subjects subjectively decide that there exists a transparent layer at this disparity.

Methods:

Four subjects participated in this experiment. They were presented with eight stereograms of three transparent layers. The density of two main layers was always 9% (namely, 9% of the pixels were black). The density of the third layer, either the top or the bottom layer, was lower and variable. The number of dots in this layer, measured as a fraction of the number of dots in the stereogram altogether, was 4%, 6%, 8% or 10% (where 33% means that all layers are of the same density). The subjects were asked to report how many transparent layers they subjectively perceived as layers, and put a cursor, whose depth they could change, on each of the layers they identified. The last procedure was required to verify which layers they had actually seen and how accurate their judgement was.

Results:

Four subjects participated in this experiment. The subjects always identified the main dense layers. Table 2 shows for which conditions each subject also identified the sparse layer. Two subjects (the first two rows in Table 2) were fairly accurate in their depth judgement of the two main layers, whereas the other two subjects were less accurate (deviating by more than a pixel from the actual depth of a dense layer).

All four subjects judged the depth of the main layers more accurately than the sparse layer. One subject identified a layer at an intermediate depth value between the two dense layers when the density of the sparse layer was 6% (namely, the density was too low for this subject to perceive the sparse layer as a distinct layer, but large enough to indicate that there were more than two layers in the image).

Sparse top layer				Sparse bottom layer			
4%	6%	8%	10%	4%	6%	8%	10%
no	no	yes	no	no	no	yes	yes
no	no	yes	yes	no	no	yes	yes
no	no	yes	yes	no	no	no	no
no	no	no	yes	no	no	no	no

Table 2: Answers whether a subject identified the sparse layer for a given condition. The first two subjects (whose data is shown in the first two rows) were more accurate than the other two subjects.

Conclusions:

The results of the four subjects participating in this experiment, especially the two more accurate subjects, were fairly consistent. They subjectively perceived a distinct layer at a given disparity when more than 7% of the dots in the image could be matched at that disparity.

3.3 Experiment 3: the micropattern of a doubly ambiguous RDS

In the double nail illusion experiment [4], in which a configuration similar to the micropattern of a doubly ambiguous stereogram (Section 2.3) was used, Krol & van de Grind reported that subjects selected only those disparities corresponding to the full circles in Fig. 1c. However, in all their experiments G_r was almost identical to G_l. In the present experiment, subjects were asked to match two-dot patterns, as in Fig. 1c, but with $G_r \neq G_l$. The conditions of the experiment described in Section 2.3 were repeated, where the isolated micropatterns were presented to the subjects instead of the ambiguous stereograms. Two subjects participated in this experiment. In agreement with [4], both subjects selected only those disparities corresponding to the full circles in Fig. 1c in all cases.

4 Computational discussion

The first experiment described in Section 3.1 suggests that the multiple matching effect, discussed in Section 2.3, can possibly be explained by an algorithm that selects at each feature a unique disparity. In this section, three such algorithms are discussed. These algorithms were selected for their simplicity and as representatives of two different matching principles; it is not suggested here that they are biologically plausible and it is not assumed that they can deal with noisy real images.

The first is a patch-wise correlation algorithm (e.g. [5, 2]), in which the disparity selected at each feature is the disparity that maximizes the correlation between a patch around the

feature in one image and a corresponding displaced patch in the second image.

The second algorithm is Prazdny's stereo matching algorithm [8]. This algorithm identifies and matches features in both images using a measure closely related to the disparity gradient defined in [1] in order to enforce smooth matching. Disparity gradient is defined for two features in one image, each assigned a specific disparity: it is the disparity difference between the two features divided by the distance (averaged over both images) between the features. In Prazdny's algorithm, at a given feature in the image, each disparity that corresponds to a feasible match receives decreasing support from its neighbors (within a certain neighborhood) with increasing disparity gradient. The most supported disparity is selected at each feature. Thus, since the algorithm uses the disparity gradient to evaluate the quality of a particular match, larger differences in disparity are tolerated for features that are further apart.

The third algorithm is PMF [7]. Like Prazdny's matching algorithm, the PMF algorithm uses the disparity gradient between two features to enforce smoothness of the disparity field. In this algorithm, each candidate match (disparity) at a given feature accumulates support from neighboring features before the selection of the best (or most supported) match. This support is given only if the disparity gradient between the two neighbors is smaller than a certain threshold, the disparity gradient limit. The use of this particular smoothing method was based on psychophysical evidence [1] that simultaneous stereo fusion of two features is possible only if the disparity gradient between them is smaller than 1.

In the rest of this section, using the three algorithms mentioned above, the questions of whether these algorithms can be modified to explain human perception, and if they can how narrow the tuning range of their parameters is, are studied.

4.1 First difficulty: different perception for micropatterns and RDS's

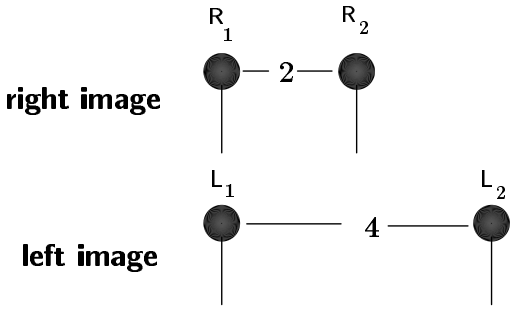


Figure 4: A stereogram of two nails as in the double nail illusion experiment [4]. Two nails are seen in both images. The right image is shown above the left image for purpose of illustration. The separation between the nails is 2 pixels in the right image and 4 in the left, where each pixel corresponds to roughly 1.2 minutes of arc as in [9].

The example shown in Fig. 4 is a simple ambiguous configuration, similar to the one used in the double nail illusion experiment [4]. This example is the micropattern of the experiment described in Section 2.3 (namely, the stereogram in that experiment is made of a random distribution of such patterns). There are four possible matchings of the two nails in the left image (L_1, L_2) to the nails in the right image (R_1, R_2). Table 3 gives the disparity gradient between L_1 and L_2 for each of these matchings. The disparity gradient is defined to be [1] the difference between the disparities assigned to the two features divided by the average distance (in the two images) between the two features.

L_1		L_2		Disparity
Match	Disparity	Match	Disparity	gradient
R_1	0	R_1	-4	2
R_1	0	R_2	-2	$\frac{2}{3}$
R_2	2	R_1	-4	2
R_2	2	R_2	-2	2

Table 3: Four possible pairings of nails L_1 and L_2 in the left image to nails R_1 and R_2 in the right image are listed. The disparity gradient is calculated for each. The complete derivations of the disparity gradient, for the four possible matchings respectively, are $\frac{0-(-4)}{(4+0)/2} = 2$, $\frac{0-(-2)}{(4+2)/2} = \frac{2}{3}$, $\frac{2-(-4)}{(4+2)/2} = 2$, $\frac{2-(-2)}{(4+0)/2} = 2$.

It is clear from Table 3 that a disparity gradient limit of 0.5 is smaller than the disparity gradient between any possible pairing of L_1 and L_2. Therefore no pairing can support the other. Thus the PMF algorithm with disparity gradient limit of 0.5, when matching this stereogram, is equally likely to detect any pairing without any preference. However, the results of experiment 3 reported in Section 3.3 show that humans, when presented with this stereogram, **always** see a single matching, L_1 with R_1 and L_2 with R_2.

The PMF algorithm fails because it uses a fixed threshold (the disparity gradient limit). Favoring low disparity gradient in a gradual way leads to better results. Prazdny's stereo matching algorithm, which uses the disparity gradient to give support with a continuous penalty function rather than by comparing to a threshold, can explain simultaneously the response to the isolated micropattern and to the stereogram for the same range of parameters.

The correlation based algorithm performs similarly to PMF with disparity gradient limit of 0.5, namely, it fails. However, since a correlation-based algorithm is not designed for the matching of isolated features, this failure is not surprising.

Discussion

Initially it was suggested in [7] that a disparity gradient limit of 1 should be the threshold in the PMF algorithm when used to model human stereo vision. In order to explain human perception in the experiments discussed in Section 2.3, the threshold had to be reset to 0.5

(as noted in the caption of Fig. 1 in [6]). Unfortunately, this change of the threshold value resulted in a failure of the PMF algorithm to account for some other known psychophysical results, as discussed above.

The example above is not an accident. In fact, no disparity gradient limit exists with which PMF can explain all the experiments with doubly ambiguous RDS's. The range of disparity gradient limits for which the PMF algorithm can explain the experiments described in Section 2.3 is identical to the range of disparity gradient limits for which it fails to explain the experiments described in Section 3.3. The difficulty follows from one of the points discussed in [9], namely, humans' response to isolated micropatterns (such as Fig. 4) appears to be different from their response to the stereograms discussed in Section 2.3. The PMF algorithm, on the other hand, responds to both type of stimuli in a similar manner.

Thus Pollard & Frisby's letter [6] demonstrates the difficulty encountered by stereo matching algorithms when dealing with doubly ambiguous random dot stereograms. A particular selection of parameters can make it possible for an algorithm to explain human perception in some cases, but the same parameters are unsuitable to explain human perception in other cases.

4.2 Second difficulty: different stereograms with the same correlation

Another problem arising from the experiments discussed in Section 2 concerns RDS's with identical correlation functions. In such stereograms (described in Section 2.5), when large regions (the regions including the doubled generating pattern) in the two images are correlated with each other, the resulting graphs look very similar (figures 3a,b), yet subjects perceive a different number of transparent layers in each. Another example (with different parameters) is given in figures 5a,b, where subjects perceive up to four layers and only two layers respectively, when presented with these stereograms that have exactly the same correlation function.

This problem does not concern only the patch-wise correlation algorithm. The correlation between the left and right images is a good measure of the kind of interaction and disparity support neighboring features provide to the matching at a certain feature. All successful stereo matching algorithms require such interactions and use support from neighboring features in one form or another to select a disparity at a given point. Thus the fact that humans perceive a different number of layers in stereograms where the neighborhood interactions seem similar poses a difficulty to any stereo matching algorithm.

In order to compare stereo matching algorithms to humans' stereo matching, a post-processing stage was added to all of them, in which the number of transparent layers was decided. In the following, the number of dots assigned (uniquely) to each disparity, summed over the whole image, was compared to a threshold to determine whether a transparent layer should be reported at that disparity[4]. The value of the threshold parameter, along with the

[4]This postprocessing stage mimics humans' subjective decision of whether they see a transparent layer

values of other parameters of each algorithm, were varied to determine whether there exists a tuning of the algorithm, a particular set of parameters, for which the algorithm can explain human perception. The sensitivity of the algorithms to any particular tuning was also studied.

The correlation with variable window sizes

The correlation function at a point in the right image (x_i^R, y_j^R), computed for a horizontal disparity D (vertical disparities are neglected in the present discussion) and using a correlation window of size W, is defined as follows:

$$\text{correlation} \;=\; \sum_{i}^{W}\sum_{j}^{W} I(x_i^R, y_j^R) I(x_i^L + D, y_j^L)$$

(where $I(x, y)$ is the image intensity at point (x, y).)

To study the interactions between neighboring points in the stereograms corresponding to figures 5a,b, the correlation function was recomputed using different correlation window sizes W. When small windows were used, in particular when the window was smaller than 5x5 pixels and G_r and G_l ranged between 2 and 4 pixels, the correlation function for stereograms corresponding to Fig. 5a showed a different distribution when compared to the correlation function for stereograms corresponding to Fig. 5b. Two examples of the correlation function for the stereogram corresponding to Fig. 5a are shown in figures 5c,e. Two examples of the correlation function for the stereogram corresponding to Fig. 5b are shown in figures 5d,f.

This variability suggests that the algorithms discussed here, Prazdny's matching algorithm and a patch-wise correlation maximization algorithm, should be restricted to small regions of interaction with neighboring dots in order to replicate human perception. In the next section, where these algorithms are studied in detail, the size of the interaction neighborhood is one of the parameters studied.

Simulations

In the following simulations, a simple implementation of Prazdny's stereo matching algorithm and a patch-wise correlation maximization matching algorithm were used, with an added postprocessing stage as discussed above. The algorithms were tested on the following ten cases:

1. A doubly ambiguous RDS, with $G_r = 4, \quad G_l = 2$. The algorithm was expected to report at least three layers at disparities $-2, 0, 2, 4$.

2. A doubly ambiguous RDS, with $G_r = 2, \quad G_l = 2$. The algorithm was expected to report a single layer at disparity 2.

or only isolated points at a particular disparity. I should note here that people seem to have difficulty with identifying more than three layers in stereograms that have four "simple" transparent layers. No attempt is made here to mimic this constraint.

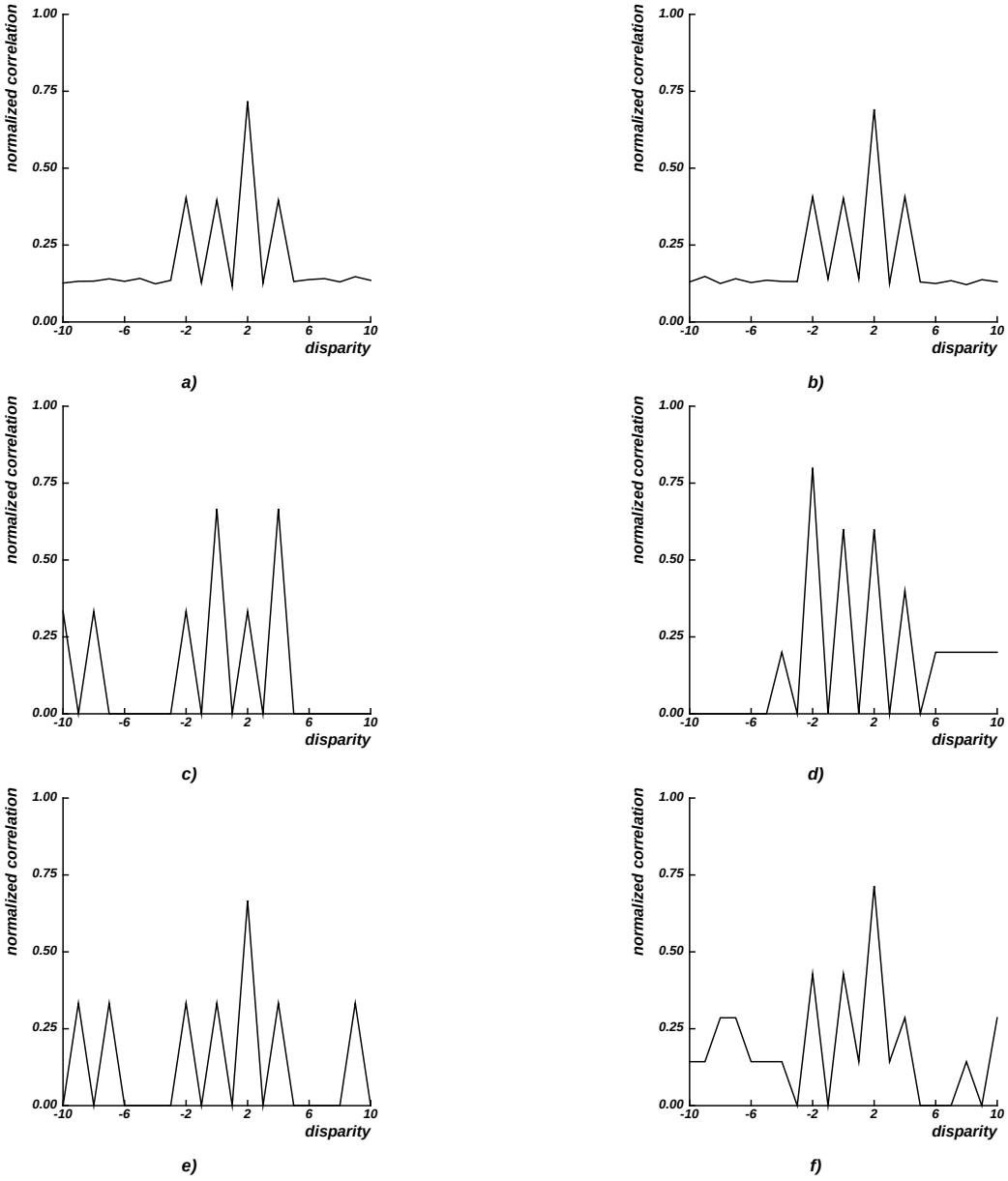


Figure 5: The correlation (as a function of disparity) between the left and the right images of doubly ambiguous RDS's. In the first row the correlation window size is 120x120, in the second and third rows the window size is 5x5. The left column gives the correlation for one stereogram, the right column gives the correlation for a different stereogram.

3. A doubly ambiguous RDS, with $G_r = 2, \quad G_l = 0$ (Section 2.2). The algorithm was expected to report two layers at disparities $-2, 0$.

4. A doubly ambiguous RDS, with $G_r = 4, \quad G_l = 2$, and with additional points at disparity 2. The algorithm was expected to report at least three layers at disparities $-2, 0, 2, 4$.

5. A doubly ambiguous RDS, with $G_r = 2, \quad G_l = 2$, and with additional points at disparity -2. The algorithm was expected to report two layers at disparities $-2, 2$.

6. A doubly ambiguous RDS, with $G_r = 4, \quad G_l = 2$, and with additional points at disparities $0, 2$. The algorithm was expected to report at least three layers at disparities $-2, 0, 2, 4$.

7. An RDS as described in Section 3.2, where the sparse layer includes 4% of the image points. The algorithm was expected to report two layers at disparities $2, 4$.

8. An RDS as described in Section 3.2, where the sparse layer includes 6% of the image points. The algorithm was expected to report two layers at disparities $2, 4$.

9. An RDS as described in Section 3.2, where the sparse layer includes 8% of the image points. The algorithm was expected to report three layers at disparities $0, 2, 4$.

10. An RDS as described in Section 3.2, where the sparse layer includes 10% of the image points. The algorithm was expected to report three layers at disparities $0, 2, 4$.

These cases were chosen as a representative subset of the stereograms used in the psychophysical experiments described in sections 2-3, including all the stereograms that may pose difficulty to a matching algorithm for one of the reasons described above. In particular, the stereograms in cases 4 and 5 have the same correlation function, given in figures 5a,b.

The simulations were repeated for 5 different data sets produced randomly, and a few times for each data set to determine consistency[5]. The performance of both algorithms was fairly consistent, though Prazdny's algorithm showed slightly higher variability. The parameters that were varied for both algorithms were the window of interaction, from 5x5 pixels to 15x15, and the threshold on the minimal density of points that elicit the impression of a distinct layer, from 2% to 10%. A normalization parameter in Prazdny's matching algorithm, which is a free parameter in the algorithm that multiplies the disparity-gradient between any two matches to compute the support function, was also varied.

[5]When more than one disparity was given the maximal support at a point, a single disparity was selected at random in the implementation of the two algorithms, and therefore a consistency check was necessary.

Results:

Case 9 seemed to be a limit case, in which subjects could identify the sparse layer less reliably. This case was therefore discarded from the initial performance evaluation of the algorithms. It was considered in a subsequent analysis of the patch-wise correlation algorithm as discussed below.

Table 4 summarizes the results for the two algorithms, five test cases, and two to three repetitions of each case. The result in each case is the set of pairs, the window size in pixels and the threshold value in percents, for which the algorithm succeeded. Prazdny's algorithm was considered successful for a particular window size and threshold value if there existed a normalization coefficient with which these two parameters produced a successful result.

Prazdny's algorithm			Patch-wise correlation		
1st test	2nd test	3rd test	1st test	2nd test	3rd test
f	*f*	*f*	(5,5%) (7,2%) (7,3%)	(5,5%) (7,2%)	(5,4%) (5,5%) (7,2%)
(5,4%)	*f*	(5,4%)	*f*	*f*	*f*
(5,5%)	*f*	–	(5,4%)	(5,4%)	–
f	*f*	–	*f*	(7,2%)	–
f	*f*	–	*f*	*f*	–

Table 4: Summary of the results of the simulations discussed in the text. Each row summarizes a different data set. Each algorithm is assigned three columns, for three separate tests of the algorithm on the same data. The algorithm was tested only twice on the last three data sets. *f* indicates a failure of the algorithm, otherwise the list of parameters for which the algorithm succeeded is given.

When an algorithm was successful for a particular set of parameters, typically the density of dots assigned to each disparity matched human performance in experiment 1 (Section 3.1).

The correlation algorithm selects at each feature the disparity d that maximizes the correlation of a patch around the feature and a patch displaced by d in the second image. However, the disparity selected in this way may not correspond to a feasible match; there need not be a corresponding feature displaced by d in the second image. An improved version of the correlation algorithm was also simulated, where the disparity with the highest correlation value, among the disparities corresponding to feasible matches, was selected at each feature. This algorithm was more successful, in particular when dealing with low density transparent layers (cases 7-10). It was tested twice on each of four data sets (corresponding to the last four rows in Table 4).

The results of the improved patch-wise correlation algorithm are given in Table 5. The performance of this algorithm was robust enough to handle the limit case 9. These results show one set of parameters, (7,4%), for which the improved correlation algorithm succeeded in every trial, for all test cases. It almost always succeeded for the sets of parameters (9,3%), (7,4%) and (5,5%).

Improved patch-wise correlation	
1st test	2nd test
–	–
(9,3%) (7,4%) (7,3%) (5,5%)	(11,3%) (11,2%) (9,3%) (7,4%) (5,5%)
(9,4%) (9,3%) (7,4%) (7,3%) (5,5%)	(9,3%) (7,4%) (5,7%) (5,5%)
(7,5%) (7,4%)	(9,3%) (7,5%) (7,4%) (5,5%)
(9,3%) (7,4%) (5,5%)	(9,4%) (9,3%) (7,4%)

Table 5: Summary of the results of the simulations for the improved patch-wise correlation algorithm tested on four data sets, corresponding to the last four rows in Table 4.

Discussion

The results of these simulations show that neither algorithm performs consistently with human perception all the time. This should not be considered as a major problem since the results some subjects reported also varied in time. Both algorithms agreed with humans for a rather small window of interaction, 5x5 pixels for Prazdny's algorithm and from 5x5 to 7x7 for the correlation algorithm. The results of the patch-wise correlation algorithm seem to be consistent with humans' more often, and for a wider range of parameters. Moreover, an improved version of the correlation algorithm proved to be consistent with human behavior all the time, for a wide range of parameters. It should also be noted that this algorithm is much faster and simpler to implement. However, it is not appropriate for the matching of single features, as discussed in Section 4.1.

5 Summary

I have discussed a number of experiments with doubly ambiguous random dot stereograms. In these stereograms there is often no single "correct" matching of the left and right images; a few different solutions to the matching problem are conceivable. Humans select a particular solution. Their performance in these tasks, which has been described in this paper, can be used to evaluate stereo matching algorithms, identifying those that are more appropriate as models of human stereo vision.

Three simple stereo matching algorithms, representing two different matching approaches, were discussed in Section 4. One algorithm, PMF, failed to explain the difference in ambiguity resolution between the random dot stereograms and the micropatterns of the stereograms presented in isolation. Prazdny's stereo matching algorithm could explain this difficulty, but its tuning to explain the other experimental results proved to be hard. The patch-wise correlation maximization algorithm could be easily tuned to agree with human perception, requiring a small correlation window, though it was not suitable for the matching of isolated features.

These results, and the conclusions of experiment 1 (Section 3.1), support the idea that the matching of isolated features may involve different processes than the matching of random dot stereograms (cf. [3]).

Acknowledgements: I thank P. Cavanagh who suggested experiment 3.1 to me, T. Poggio who suggested looking at smaller correlation windows, and both E. Hildreth and T. Poggio for helpful comments regarding the manuscript. I also thank D. Bar-Natan, H. H. Bülthoff, F. Girosi, Y. Karshon, S. Kirkpatrick, and J. McFarland. This research was done partly in the MIT AI Laboratory. It was supported by an MIT postdoctoral fellowship. It was also supported 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 E. Hildreth.

References

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

[2] M. Drumheller and T. Poggio. On parallel stereo. In *Proceedings of IEEE Conference on Robotics and Automation*, 1986.

[3] B. Julesz. *Foundations of Cyclopean Perception.* University of Chicago Press, Chicago, IL, 1971.

[4] 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.

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

[6] S. B. Pollard and J. P. Frisby. Transparency and the uniqueness constraint in human and computer stereo vision. *Nature*, 347:553–556, 1990.

[7] 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.

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

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

[10] D. Weinshall. Seeing 'ghost' planes in stereo vision. *Vision Research*, 31(10):1731–1748, 1991.