The Uniqueness Constraint Revisited:

A symmetric Near-Far inhibitory mechanism producing ordered binocular matches

Jens K. M. Månsson

Lund University Cognitive Science

Kungshuset, Lundagård

S-222 22 Lund

Sweden

jens.mansson@lucs.lu.se

Abstract: Psychophysical studies have shown that image features, under certain conditions, can give rise to multiple visible binocular matches. These findings are difficult to reconcile with the traditional interpretation of the uniqueness constraint. A new interpretation, a *conditional uniqueness constraint*, is proposed that allows multiple matching of similar primitives when a one-to-one correspondence does *not* exist locally within corresponding image regions, but prohibits it when a one-to-one correspondence does exist. A cooperative network model and an implementation are also described, where this constraint is enforced at each network node by a simple inhibitory (dual) AND-gate mechanism. The model performs with high accuracy for a wide range of stimuli, including multiple transparent surfaces, and seems able to account for several aspects of human binocular matching that previous models have not been able to account for.

1 Introduction

Due to a frontally directed binocular visual system, humans and many other animals are able to perceive the 3-D structure of the environment, even when no monocular cues (e.g. motion-parallax, perspective, size etc) are available. Because our eyes are horizontally separated, and hence view the world from slightly different perspectives, the two different images that fall onto the left and right retinas will, in general, not be identical. With increasing distance from the plane of fixation, the relative binocular disparity between corresponding visual features in the left and right images will increase. If this binocular disparity, and the degree of convergence of the eyes, can be measured or somehow estimated, so can the distance to a visible feature. The problem of finding corresponding image features in the left and right images is basically a matching problem, and is referred to as the *correspondence problem*. A fundamental difficulty with the correspondence problem is that it is ill-posed; i.e. the image arrays do not in general contain sufficient information to determine with certainty which solution (of multiple possible) is the correct. Thus, given the total solution space, or the set of all possible

matches, the best any visual system can do is to choose the solution that seems most reasonable in the context encountered. What is reasonable in any particular context should preferably be determined with some kind of heuristic that is based on experience (learned or "built-in") of how the physical world behaves. With such knowledge it is possible to *constrain* the matching process, so that solutions that are in agreement with this knowledge are favored to solutions that are unrealistic, or "off the wall". However, if the constraints chosen are not flexible enough, or too strictly applied, there is a risk, when faced with unusual scenes, that a visual system will not deliver accurate descriptions of the environment because it will be too hard "locked into" interpreting all scenes in a certain way. Thus, an important factor in the design of both artificial and natural stereo systems should be to find an acceptable balance between the effectiveness and the flexibility of any particular constraint. A balance that may change from species to species depending on the accuracy needed, or that may change from situation to situation depending on the visual and physical context.

Over the years, a variety of different matching constraints have been suggested: The uniqueness constraint (Marr & Poggio, 1976) states that any given image primitive should be matched with one, and only one, other primitive in the opposite image, because (surfaces in general are opaque) any given point on a surface must have a unique position in space.

The cohesitivity (or continuity) constraint (Marr & Poggio, 1976) states that, because surfaces in general changes smoothly in depth, nearby image points should have similar binocular disparities.

The constraint of figural continuity (Mayhew & Frisby, 1981), or edge connectivity (Baker & Binford, 1981), is based on a similar motivation, but is concerned with edge information. This constraint basically says that an edge segment that is part of a longer continuous edge in one image, should be matched with an edge segment in the opposite image that is a part of a similar continuous edge, so that the figural continuity is preserved binocularly.

The ordering constraint (Baker & Binford, 1981) is motivated by the fact that under normal viewing conditions the relative ordering of image features, in the left and right images, is rarely broken; hence the ordering constraint imposes that binocular matches that preserve the relative ordering should be favoured to unordered ones.

Partly motivated by the same observation as that underlying the ordering constraint, and partly motivated by psychophysical observations (Burt & Julesz, 1980), Pollard, Mayhew and Frisby (1985) have suggested that a disparity gradient limit of 1 could be used to select correct matches. Given two binocular matches, the disparity gradient is defined as the difference in disparity, between the matches, divided by their cyclopean separation (Burt & Julesz, 1980). Simply put, given a number of neighbouring image features, this constraint can be said to favour (groups of) matches that have relatively similar disparity values , i.e. lie on a smoothly changing (but not necessarily fronto-parallel) surface, over (isolated) matches that have deviating disparity values.

Except for the disparity gradient limit, it seems as if the motivation and justification for these constraints have mainly come from computational considerations, and not so much from psychophysical observation of how the human visual system actually behaves. Naturally, from this it does not necessarily follow that these constraints are wrong, or that they may not account for how the human visual system operates. The ordering constraint, for example, seems rarely - perhaps never - to be broken by the human visual system (see below). Nor does it necessarily follow that a disparity gradient limit is actually used in human vision. In fact, it is unclear both from a computational and a psychophysical perspective why this limit should be fixed to 1, since (as shown below) this does not always seem to hold in human vision. However, the main point here is that although computational considerations can be a highly valuable source of inspiration, and suggest simple and elegant solutions, one must not be blind to when these solutions fail to correspond with the performance of the human visual system.

Of all the constraints above the perhaps most influential and most often encountered in other models of human stereopsis, and other algorithms proposed for solving the correspondence problem, is the uniqueness constraint proposed by Marr and Poggio (1976). It seems as if their particular interpretation has been the prevailing, and only accepted, one. This is surprising considering its obvious shortcomings, when it comes to explaining transparency, and in accounting for known instances of multiple matching in human vision; e.g. Panum's limiting case, the "double-nail" illusion (Krol & van de Grind, 1980), Weinshall (1991, 1993).

This article will start out by taking a closer look at the computational and ecological justification for the uniqueness constraint, and show that the particular interpretation, and implementation, proposed by Marr and Poggio (1976) may be unnecessarily restrictive in which matches it allows, and therefore unable to account for

certain aspects of human perception. An alternative interpretation, a *conditional uniqueness constraint*, is then proposed. In brief, this constraint operates in a similar manner (to the one proposed by Marr and Poggio) when there locally is an even number of similar matching primitives within corresponding binocular regions, but allows multiple matches when there is an uneven number of similar matching primitives within the same regions; when a one-to-one correspondence does not hold. A simple computational mechanism that enforces this constraint is then presented along with a cooperative network implementation, and some simulation results. Finally, the model's relationship to previous models, and its plausibility as a model of human depth perception, is discussed.

2 The uniqueness constraint revisited

The motivation for the uniqueness constraint (Marr & Poggio, 1976) is based on the observation that any given point on a surface must have a unique position in space. In their model, this observation was translated into a matching rule that requires any given image feature to be matched with only one feature in the opposite image. Although this particular interpretation has proved to be a highly effective constraint for a wide range of binocular stimuli - and seemingly undisputed - there are several reasons to question its general validity. The motivation for this comes from both computational and ecological considerations, but the main reason - which from the perspective of wanting to understand human stereopsis is more important - comes from the inability of this constraint to account for transparency and known instances of multiple matching in human vision; e.g. Panum's limiting case, the “double-nail” illusion (Krol & van de Grind), and Weinshall's random-dot stereograms (Weinshall, 1991,1993) discussed below.

First of all, it should be pointed out that although multiple matches violate the uniqueness constraint, it does not strictly speaking violate the principle on which the uniqueness constraint is based. That is, allowing multiple matches may not always be sensible, or facilitate the correspondence problem, but it is not in itself contradictory to the fact that any given surface point has a unique position in space; since even if a feature in one image were matched to two or more features in the opposite image, each individual match would still correspond to a unique position in space.

More important in this context is the fact that the uniqueness constraint is not always justified (a fact that Marr and Poggio, of course, were fully aware of). Although it is true that any given point on a surface must have a unique position in space, it is *not* always true that a given image point will correspond to a unique position space, nor that a given surface point will always be represented in both images. For example, when two overlapping transparent surfaces are seen, there is not one, but two “true” disparity values associated with each image point, one for each surface. Further, due to surface slant (and the difference in perspective between the eyes) some features, or surface regions, may appear differently sized, or shaped, on the two retinas. Finally, because the images in our eyes are decompositions of a 3-D world it is unavoidable that some surface points become occluded to one eye but not the other (half occlusion).

In order for the uniqueness constraint to be meaningful there need to exists a one-to-one correspondence between matching primitives. When surfaces are opaque and smoothly curved in depth, and all surface points are visible from both eyes, such a relationship does exists, and the uniqueness constraint can then be an important key to solving the correspondence problem. However, when surfaces do not have these “nice” properties, but for instance there are many occurrences of half-occlusion, there is no guarantee that a one-to-one relationship exists between the left and right image features. A given image feature may have one, none or many potential matches in the opposite image half. In situations like these, when there locally is an uneven number of identical, or similar, matching primitives in the two images halves, there simply is no way of uniquely matching the primitives without leaving some primitives out. Why some image primitives should be ignored altogether is not easily motivated. One could argue that, while it is true (and indisputable) that any given surface point must have a *unique* position in space, it is equally true that any given surface point must have *some* position in space. And therefore, in situations when the one-to-one correspondence condition does not hold, the price to pay for ignoring some features might sometimes be higher than the price for allowing certain multiple matches (see discussion below).

In previous models of stereopsis (Marr & Poggio, 1976; Marr & Poggio, 1979; Pollard, Mayhew & Frisby, 1985) this problem seems to have been basically ignored, or avoided by assuming that surfaces in general are

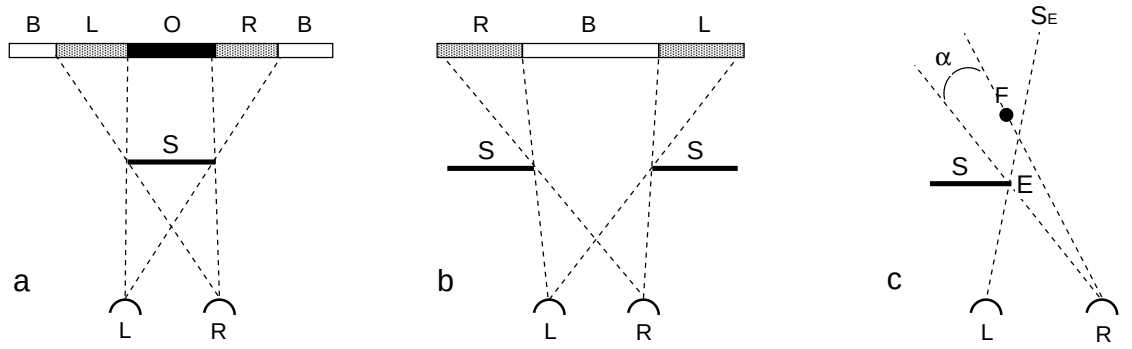

Figure 1. Schematic (top) view of the 3-D layout in cases where unpaired monocular regions may be visible. (a and b) S=occluding foreground surface, B=binocularly visible regions, O=occluded region, L=visible to left eye only, R=visible to right eye only. Left eye monocular regions can only be seen to left of an occluding surface; and right eye monocular regions can only be seen to the right of an occluding surface.(c) Given a binocular fusible surface edge, E, the (minimum) depth of the monocular (right eye) feature, F, depends on the monocular separation, α, between the edge, E, and the feature, F. (Modified after Nakayama & Shimojo, 1990).

opaque and smooth. However, if 3-D surface reconstruction is one of the important goals of early vision (Marr, 1982), this is a strange approach to take since any cues of transparency, and particularly half-occlusion would be highly valuable information to such a process.

Because half-occlusion, in one way or another, is the only fundamental reason why a one-to-one correspondence may *not* exist between image features, it is interesting to consider in some detail the possible variations of 3-D surface layouts in these cases. Regarding an unpaired monocular feature (or region), Nakayama and Shimojo (1990) have pointed out that there are only two ecologically valid situations where such stimuli can arise (figure 1a,b). If a monocular feature is seen by the left eye only, the feature must be behind an occluding surface that is located to the right of the feature; or vice versa, if seen by the right eye only, the feature must be behind a surface that is located to the left of it. They further pointed out that the depth ambiguity a monocular feature presents, to some extent is constrained if a nearby binocularly fusible edge is present. In figure 1c for example, the region of space from where the monocular feature F could have arisen is delimited by the line RF and the line LS_E (which tangents the fusible right edge, E, of the surface, S). That is, the feature F must have arisen from some point along the line RF that is located to the left of the line LS_E, otherwise it would be visible to both eyes. Thus, the minimum depth corresponding with F is were LS_E and RF intersect. Moreover, this minimum depth depends on the angular separation (α) between the binocularly fusible edge (E) and the monocular feature (F). Hence, the greater the angular separation (α), the greater the minimum depth. Interestingly, Nakayama and Shimojo showed in one of their experiments that subjects did seem to use this information in their depth judgements. When subjects were asked to estimate the depth of an unpaired monocular stimuli their estimates varied quantitatively with the angle of monocular separation from a nearby fusible edge. The larger the angle of separation, the further away the monocular target was perceived.

The case when there is an uneven number of identical, or similar, features in the two images halves (and multiple matching is an option) is highly similar to the case with a single monocular feature described by Nakayama and Shimojo, and it is interesting to consider the problems posed to a binocular matching system when faced with such stimuli.

Consider for example the simple and well known Panum's limiting case (figure 2a) where the left eye sees only one vertical bar, but the right eye sees two. Basically, there are three possible 3-D layouts (figure 2b,c,d) that could give rise to such an image pair: In the simplest case, 2b, the two right eye features (R_1 and R_2) are exactly aligned along the left eye's line-of-sight from L_1. In 2c the single visible left eye feature (L_1) corresponds to the leftmost right eye feature (R_1), and is located on (or in front of) an opaque surface (S) that terminates somewhere between the two features R_1 and R_2. The rightmost feature (R_2) is here occluded to the left eye by the surface (S). Note that this surface is not directly visible in itself, but only indirectly due to the half-occlusion. Finally, in 2d the single left eye feature (L_1) corresponds to the rightmost, right eye, feature (R_2), and is located further away than an opaque (invisible) surface that terminates somewhere to the left of L_1 and R_1. The

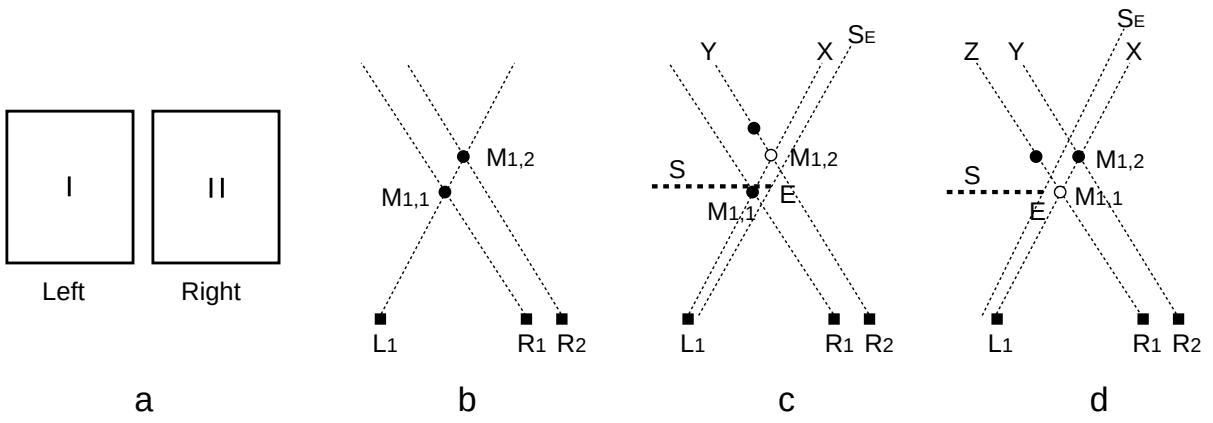

Figure 2. Panum's limiting case (a), and the only three plausible (ecologically valid) 3-D layouts that could have produced it (b, c and d). The filled circles mark the actual positions of the visible features L_1, R_1 and R_2 (filled squares). The thick dashed line is an occluding surface, S, that is not in itself visible. (b) Both match $M_{1,1}$, and $M_{1,2}$, correspond to the actual position of R_1, and R_2, respectively. (c) The true position of R_1 is on, or in front of, the surface S, and R_2 is behind S. Match $M_{1,1}$ is correct, but $M_{1,2}$ is false. (d) Both R_1 and R_2 correspond to points/regions behind the "invisible" occluding surface, S. Match $M_{1,1}$ is false, but $M_{1,2}$ is correct.

leftmost right eye feature (R_1) is also further away than the occluding surface, but is only visible to the right eye.

Now, from a computational perspective, given no direct visual evidence of the occluding surface (S) in case 2c and d, all three of these 3-D layouts are equally likely to have produced the visible stimuli. According to the uniqueness constraint, however, only one of the matches $M_{1,1}$ and $M_{1,2}$ can be allowed. In case 2b, whichever match was chosen, this would never result in a true representation of the situation. And in case 2c and 2d there would be a 50% chance of getting one of the matches right. Worse yet, which is true in all three cases, is that the gist of the scene, the important fact that there exists a disparity jump, would be lost. If on the other hand, both matches were allowed to co-exist in this type of situation, this important fact would not be lost. And not only would it correctly represent the situation in case 2b, but it would also correspond to a fairly good approximation of the true layout in case 2c and 2d. Moreover, as Nakayama and Shimojo (1990) pointed out in their analysis of the strictly monocular case, this approximation would not only be qualitative, but to some extent also quantitative since the correct match (whichever it is) will delimit the possible locations of the other match.

In case 2b it is obvious that matching L_1 with both R_1 and R_2 would correspond to the actual positions ($M_{1,1}$ and $M_{1,2}$) of the two features. In case 2c, match $M_{1,1}$ would be a correct match and although $M_{1,2}$ may not correspond to the actual position of R_2 we know (from the analysis of Nakayama and Shimojo) that it must be somewhere along the line-of-sight from R_2, but also to the left of the line L_1X. Therefore match $M_{1,2}$ at least conveys an approximation of the actual layout, to the degree that it corresponds closely to the minimum depth of R_2. Strictly speaking the minimum depth would correspond to where line ES_E intersect with R_2Y, but since neither the surface (S), nor the edge (E), are visible this information is not available. However, if the edge where too far to the right of L_1 it would occlude feature R_2 as well, and therefore the difference should not be very large. Case 2d is perhaps the most interesting because here it is obvious that $M_{1,1}$ is neither the correct position of L_1 and R_1, nor a very good approximation of it. However, using the reverse argument from Nakayama and Shimojo (regarding the location of an unpaired monocular feature), one can say that if $M_{1,2}$ is a correct match (true in this case) and there exists an occluding surface that is not directly visible (as is the case in 2b); the position of this surface edge will be delimited by the angular separation between match $M_{1,2}$ and the monocular feature R_1. That is, this surface (edge) must lie to the left of both line L_1X (otherwise none of the features would be visible to the left eye), and line R_1Z (otherwise R_1 would be occluded to the right eye); but we also know that it can not lie too far to the left of L_1 and R_1, since otherwise R_1 would be visible to the left eye as well. In case 2d, one can therefore say that although $M_{1,1}$ is an incorrect match of the visible *features*, it does represent the 3-D layout in the sense that it approximates the position of the *edge* in the same manner as $M_{1,2}$, in case 2c approximates the position of feature R_2.

To rephrase this in different words, because the only three ecologically valid surface layouts that could generate the stimuli in Panums's limiting case, all involve an occluding surface that terminates in the near vicinity

to the left of the occluded feature (and also immediately to the right in the border case 2b); it does not really matter which match is actually the correct when it comes to capturing the essence of the scene; i.e. that there exists a disparity discontinuity near L. Thus the best any visual system could do in such cases seems to be to accept both matches and, given the available (lack of) evidence, or cues, treat them as the best possible interpretation of the 3-D layout. Contrasting this approach with the traditional version of the uniqueness constraint will at any rate result in a less grave error in the interpretation of the scene, and a smaller loss of potentially valuable information.

Given the strong similarity between Panum's limiting case, discussed above, and the strictly monocular case discussed by Nakayama and Shimojo (1990), two things are worth pointing out about their study. First, in the experiment where subjects experienced the depth of the unpaired stimulus to depend on the monocular separation from the matchable surface edge, it is quite possible that the binocularly visible edge was matched twice. The stimuli used for both the binocularly fusible surface, and the monocular target, were white rectangular regions. If the surface edge were matched twice, this would explain why the unpaired bar appeared further away with increasing separation from the binocularly fusible edge. Second, in another of their experiments where unpaired vertical bars had been inserted (in an ecologically valid manner) to simulate an occluding (but in itself invisible) surface, subjects did perceive illusory, or subjective, contours at the appropriate side of the unpaired (half-occluded) features.

3 Psychophysical evidence for multiple matching

There are several examples where the human visual system, when faced with an uneven number of similar features, seems to make multiple matches. In the Panum's limiting case (figure 3a) it has been known since long that, given a small separation between the bars, the fused percept is not that of a single bar, but instead one sees two bars (one in front of the other). In the trivial case where both eyes sees two bars (figure 3b), it is interesting to note that the fused result is not that of two bars superimposed on two bars that lie further away (schematically depicted to the right in figure 3b) even though this is a possible interpretation, but instead that of two stable bars in the same depth plane. In this trivial case it seems as if the two redundant (unordered) matches indeed are suppressed. This shows that the visual system clearly separates between situations, and treats stimuli different, depending on if there is an even or uneven number of matching primitives; and it seems to indicate that the visual system chooses the least complicated interpretation, given the available information.

Another example where multiple stable matches have been demonstrated is in the “double-nail” illusion (Krol & van de Grind, 1980). In the basic version of this illusion (figure 4a) two nails of the same length are placed, with their heads vertically aligned, on a lath, which is aligned with the midsaggital plane. When the setup is viewed with both eyes, one does not see the nails at their actual positions, but instead at the positions corresponding to the false, or ghost, matches (figure 4b). In a variation of this experiment (which also can be seen as a variation of Panum's limiting case), Krol and van de Grind added a third nail that was placed behind the other two and aligned with the left eye's line-of-sight through the middle nail so that it was only visible to

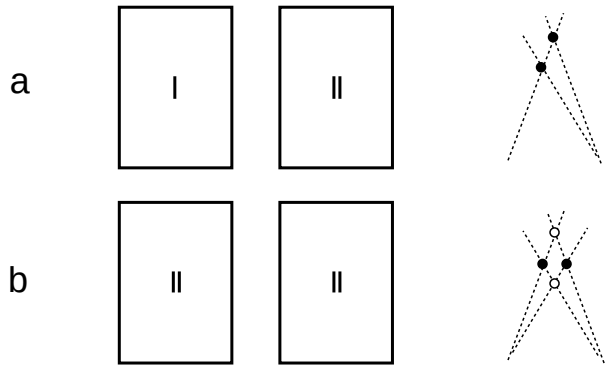

Figure 3. (a) Panum's limiting case, and (b) the “trivial” case where only the two ordered matches (filled circles) are seen.

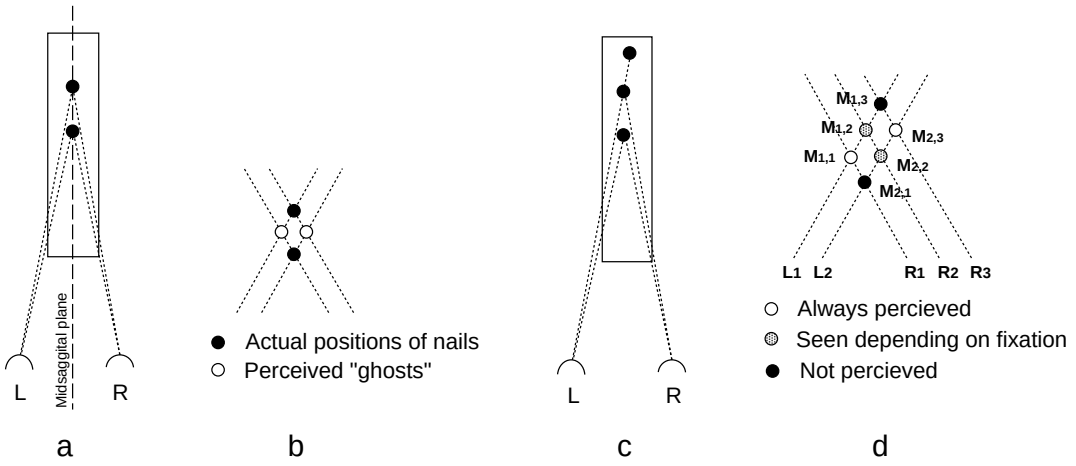

Figure 4. (a) Basic setup of the double-nail illusion. (b) Schematic view of the possible matches. (c) The variation of the double-nail illusion, where a third (most distant) nail is placed so it is occluded to the left eye only. (d) Possible matches in the latter setup. Modifed after Krol & van de Grind (1980).

the right eye (figure 4c). With this setup their subjects reported that there were two stable vergence conditions (4d). In both cases the two matches $M_{1,1}$ and $M_{2,3}$ were seen, but depending on if fixation was on either $M_{1,1}$, $M_{2,2}$ was also seen, or if on $M_{2,3}$, $M_{1,2}$ was seen.

A more massive form of multiple matching has been demonstrated by Weinshall (1991, 1993). In a series of experiments, using ambiguous random-dot stereograms (RDS), which consisted of multiple copies of the (basic) double-nail illusion stimuli, Weinshall reported that subjects saw up to four different transparent depth planes. These ambiguous RDS were constructed by making a single random-dot image which was copied twice into each half of the stereogram with an inter-dot separation that was G_L in the left image, and G_R in the right image. When the dot separation was the same ($G_L = G_R$), subjects saw only one opaque surface, the one corresponding to the ordered matches - as would be expected from the single pair case. In contrast to the single pair case, however, when the inter-dot separation was different ($G_L \neq G_R$) their subjects often reported of seeing, not only the two planes corresponding to the ordered matches, but also one or both of the ('ghost') planes corresponding to the unordered matches. Weinshall also reported of an unpublished study by Braddick (see Weinshall, 1993) in which similar RDS stimuli where used, with the difference that the RDS pattern was only copied once in one of the stereo pair images. Like in the single pair case (Panum's), the dots in the single-copy image were matched twice and two depth planes was perceived.

As a final example, MaKee, Bravo, Smallman and Legge (1995) have found, measuring the contrast-increment threshold for high-frequency targets, that a single target in the left eye can mask both of two identical targets in the right eye, when arranged in a configuration like the Panum's limiting case; and that this binocular masking was nearly the same as when only one target was visible to the right eye. To confirm that the left target actually had been matched twice, they further conducted a depth-judgement test where subjects (over 200 trials) had to determine if a test target in the right eye where in front of, or behind, a reference target. The stimuli presentation time (200ms) was too short to allow for voluntary vergence movements. The results of these depth judgements where essentially the same as when only a single target was present in the right eye, and hence supported the idea that the left target was matched with both targets in the right eye. MaKee et al. (1995) concluded that “uniqueness is not an absolute constraint on human matching”.

Taken together the above examples strongly suggest that when there locally exists an uneven number of similar features in the left and right retinal images, and there consequently does not exist a one-to-one correspondence between matching primitives, the human visual system does allow multiple matches to co-exist. If so, this clearly speaks against a strict enforcement of the uniqueness constraint; i.e. one that ignores or suppresses other potential matches under such conditions, and consequently it also speaks against models of human stereopsis that blindly rely on this constraint, since they are not flexible enough to account for the above results. What instead seems to be called for is a more sensitive usage, or enforcement, of the uniqueness constraint that, given a feature for which a match is sought, not only takes into consideration the similarity to features in the opposite eye, but also to the presence (and/or absence) of similar features in the eye-of-origin.

Returning briefly to the double-nail illusion, this is interesting not only because it demonstrates multiple matching, but also because it shows how strongly the human visual system seems to prefer matches that preserve the relative ordering of image features. In the basic version of this illusion, subjects consistently perceived the targets at their ghost positions instead of at their actual position. Perhaps counter intuitively this percept survived despite remarkably large changes in the nails dimension/appearance such as when the distal nail was longer than the other and both heads could be seen at different heights; one nail was twice the diameter of the other; or the proximal nail was rotated 5 deg around an horizontal axis through its centre. Despite the many, both monocular and binocular, cues that could have been used by the visual system to choose the correct matches the ghosts were consistently preferred. To explain this phenomena, Krol and van de Grind suggested a (somewhat vague) rule which stated that the perceived pair was always the pair of matches closest to the fixation point. A simpler (and in their case equivalent) explanation is that our visual system always chooses matches that preserve the relative (left-right) ordering of features in the two retinal images. It should be noted that Panum's limiting case does not, strictly speaking, break this principle but can in fact be considered as a border case.

4 A revised uniqueness constraint

The alternative interpretation of the uniqueness constraint, proposed below, arose from an attempt to reconcile and explain the three following properties of, or rather assumptions about, human binocular matching: i) When a one-to-one correspondence exists between matching primitives, any given primitive is matched only once. ii) When a one-to-one correspondence *does not* exists locally between *similar* matching primitives, any given primitive is matched at least once. iii) Binocular matches that preserve the relative ordering of image features are always preferred to unordered matches.

At a first glance, these three conditions on the matching mechanism may seem unrelated and therefore not easily integrated, but in fact they can be implemented by a quite simple (dual) mechanism. The basic architecture of the model consist of a network of interconnected nodes that each represent a particular point in disparity space. Figure 5 shows a horizontal layer of this network. Initially the activity at each node $\mathbf{M}(x_L,x_R,y)$ in the network is proportional to the similarity, between the stimulus at position (x_L,y) in the left image and position (x_R,y) in the right image. In this default mode, or rather without any additional modifications, no particular constraint would be enforced (except that matches are sought only along corresponding epipolar lines), and given, for example, the stimuli in Panum's limiting case, both of the nodes that represent the two possible matches (figure 3a) would remain active. Although such a simple model could account for the percept in Panums's limiting case, it obviously can not account for why we only see the two ordered matches in the trivial case (figure 3b); or why we, for example, only see the five ordered matches with stimuli like that depicted in figure 6, al-

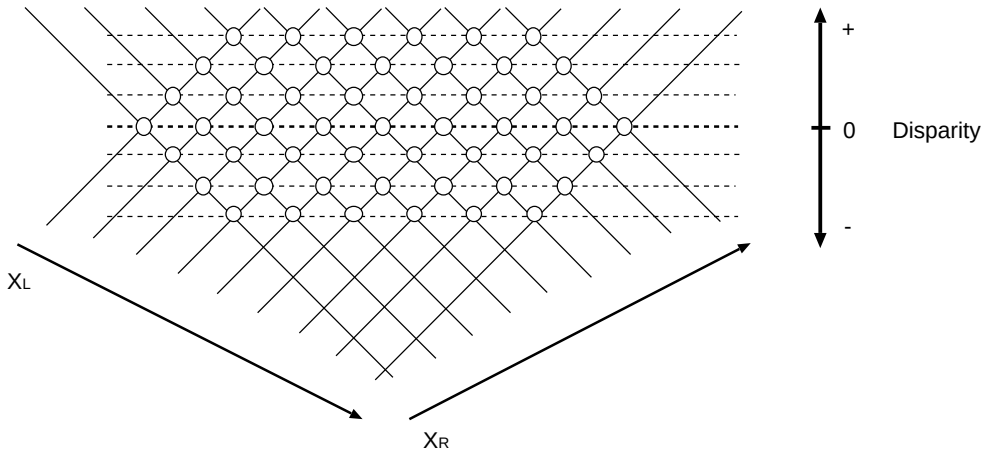

Figure 5. Schematic view of a horizontal layer in the network, showing the mapping of the left and right x-axis. Positive disparities correspond to further depths, and negative disparities correspond to positions closer to the plane of fixation.

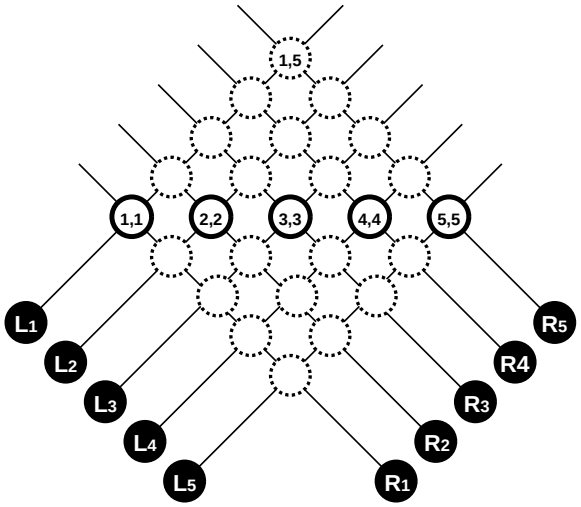

Figure 6. Example of the basic difficulty with the correspondence problem. Five identical stimuli (L_1...L_5 and R_1...R_5) are visible to both the left and right eye, and 25 different matches ($M_{1,1}$...$M_{5,5}$) are possible. In general however, only the five ordered matches are perceived (filled circles). The match M1,1 (of L1 with R1) has no competing matches to the left of it. Neither its Near, nor Far, AND-gate will therefore be activated, and the node M1,1 will not be suppressed. Match M1,5, on the other hand, will be strongly suppressed, because its Near AND-gate have input from both sides (multiple competing matches along both the left, and right, line-of-sight).

though there are 25 possible combinations between the left (L_1...L_5) and right (R_1...R_5) features. For the model to handle such stimuli, some kind of inhibition must be introduced. Hence, to prevent multiple matching from occurring in the model when a one-to-one correspondence does exists between the left and right image features, a *conditional uniqueness constraint* is introduced. Like in the model proposed by Marr & Poggio (1976), the uniqueness constraint is enforced by mutual suppression between matches that lie along the same lines-of-sight, but unlike their model this suppression is dependent upon that two conditions must hold for the binocular stimuli. The first of these two conditions is required assure property (i) above, and the second condition is needed to assure property (iii). However, although the two conditions have different motivations, they are not independent of each other but on the contrary tightly coupled. In fact, without the second condition the first would be meaningless.

Given a target match $\mathbf{M}(x_L,x_R,y)$ the first condition is that other matches, lying along one of the two lines-of-sight (e.g. $\mathbf{M}([0....n],x_R,y)$), may suppress $\mathbf{M}(x_L,x_R,y)$ only if there also exists potential matches along the opposite line-of-sight (e.g. $\mathbf{M}(x_L,[0....n],y)$). Or differently put, if a feature in, for example, the left eye can be matched with more than one feature in the right eye, the corresponding matches compete for dominance only if there also exists other (similar) features in the left eye that also can be matched with the same set of features in the right eye.

The second condition on the suppression, given a target match $\mathbf{M}(x_L,x_R,y)$, is that only combinations of matches that have the same sign of the disparity relative to $\mathbf{M}(x_L,x_R,y)$ contribute to the suppression (e.g. $\mathbf{M}([(x_L+1)...n],x_R,y)$ in combination with $\mathbf{M}(x_L,[(x_R+1)...n],y)$, and/or $\mathbf{M}([0...(x_L-1)],x_R,y)$ in combination with $\mathbf{M}(x_L,[0...(x_R-1)],y)$). Or in other words, matches along one line-of-sight that lie further away than the target **M**, are allowed to suppress **M** only if there are also matches along the opposite line-of-sight (through **M**) that lie further away than **M**; And, vice versa, matches along one line-of-sight that are closer than **M**, are allowed to suppress **M** only if there are also matches along the opposite line-of-sight that are closer than **M**.

One simple mechanism that enforces both of these conditions is depicted in figure 7, and consists of a pair of inhibitory AND-gates: a Near AND-gate (or Near-gate) and a Far AND-gate (Far-gate). Consider that each node, in the basic network described above, has such a dual mechanism connected to it (as shown for only one node in figure 7). Now, the Near-gate separately sums the activity, along the left and right line-of-sight, of all nodes that lie in front of $\mathbf{M}_T$. Only if both the (left and right) sums are larger than zero does the Near-gate pro-

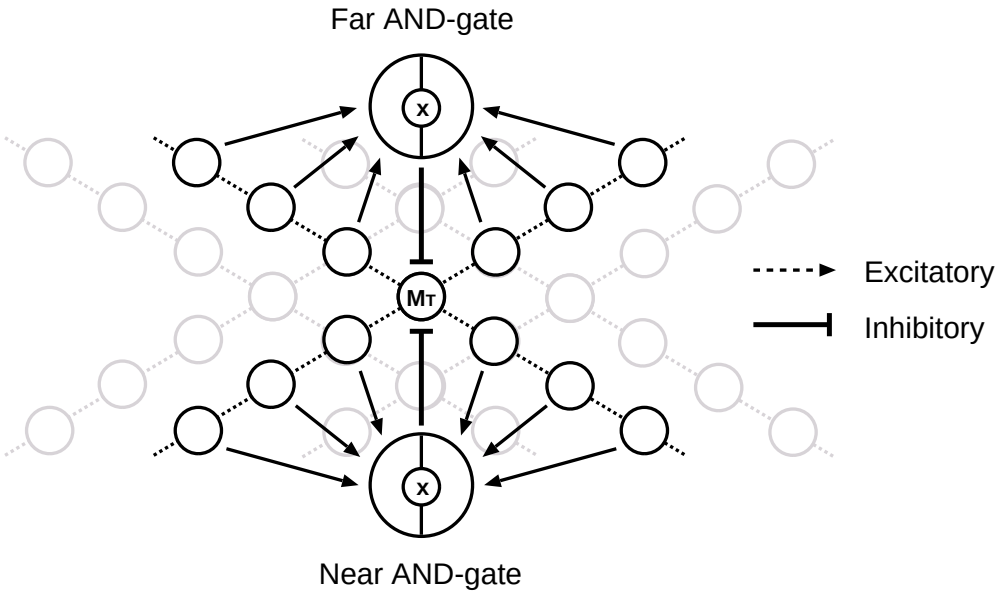

Figure 7. Depiction of the proposed dual mechanism consisting of a Near, and a Far, inhibitory AND-gate. Each AND-gate separately sums the activity along the left, and right, lines-of-sight (through the target node M_T), and multiplies the two sums. If both the left, and right, side input to an AND-gate is greater than zero, the target node, M_T, will be suppressed.

duce an output that suppresses node $\mathbf{M_T}$. The Far-gate is identical in operation to the Near-gate, but receives its input from matches that correspond to depths further away than the target node $\mathbf{M_T}$.

If one considers the dynamic interactions between the nodes in the network and their associated inhibitory mechanisms, it should become clear why this model will produce a unique and ordered set of matches for the stimuli in figure 6, i.e. the matches in the middle horizontal row (this solution set would be predicted by any known stereo constraint).

Consider for example the two features L_1 and R_1 (figure 6), from either view these two are the leftmost feature, and should therefore be matched with each other. Any other solution would be incongruent with the ordering constraint. Because both the Near-gate and the Far-gate associated with node $\mathbf{M}_{1,1}$ only receives input from matches along one of the line-of-sights, the net output of both of these AND-gates will be zero, and consequently node $\mathbf{M}_{1,1}$ will not be suppressed. For any other node along the line-of-sights from L_1 or R_1 this is not the case. At all these other positions there is some input (either in front of, or behind, the node) from both the left and right line-of-sight, and they will consequently all be inhibited to some degree. Consider particularly the node $\mathbf{M}_{1,5}$, corresponding to the match of feature L_1 with R_5. From the left view, L_1 is the leftmost feature, but from the right view R_5 is the rightmost. Clearly, matching these two features would severely violate the ordering principle, and it is therefore an unacceptable solution. In this case the Near-gate connected with $\mathbf{M}_{1,5}$ will receive a strong input from both the left and right line-of-sight, and the node will therefore be strongly suppressed. This latter situation quite nicely illustrates the motivation for the second condition above. That is, except for cases when the binocular stimuli consists of highly repetitive patterns (such as in the Wallpaper-illusion), a high number of competing matches on both the left and right side of the target match ($\mathbf{M_T}$) indicates that the target is not an ordered match. The output of the (multiplicative) AND-gate can therefore been seen as a measure (although not the only) of how well the match is ordered in relation to its neighbours. However, even if the stimuli is (finitely) repetitive as in figure 6, the proposed mechanism will produce an ordered set of matches. Consider, for example, the "correct" match $\mathbf{M}_{3,3}$ in figure 6, which has competing matches with crossed and uncrossed disparities (relative to it) along both the left and right line-of-sight. Initially, this node will be suppressed. However, because node $\mathbf{M}_{1,1}$ is not suppressed but all other nodes along its lines-of-sight are, both of node $\mathbf{M}_{2,2}$'s AND-gates will eventually loose their input from one side and consequently node $\mathbf{M}_{2,2}$ will be "disinhibited". Once $\mathbf{M}_{2,2}$ has been disinhibited, it in turn will suppress the nodes that input to the AND-gates connected to node $M_{3,3}$, leading to its revival. As this disinhibitory process propagates further in to the network eventually only nodes that preserve the ordering will remain.

5 An Implementation

A computer implementation of the proposed model have been realised and tested on a variety of different stimuli. The model consists of two major levels: a preliminary matching stage (I), and a secondary relaxation stage (II) where the conditional uniqueness constraints is enforced. At the preliminary stage, suitable matching primitives are identified independently in each image array, and subsequently binocularly matched for similarity. The result (i.e. the set of all potential matches for the stereogram in question) is then fed forward to the secondary stage. At the secondary stage, all potential matches are allowed to inhibit (disinhibit) each other, over a number of iterations (according to the principles described above), until a stable state is reached. To balance the inhibition in the network and avoid that a "dead" state is reached, there is a small continuous excitatory feed from the preliminary matching stage into the secondary stage.

Because the purpose of the current implementation is mainly to show that the principles enforced in the secondary stage are sound (i.e. that the correspondence problem can be effectively solved for a wide range of stimuli), the preliminary stage has been kept as simple as possible. In the current implementation, the preliminary matching stage simply uses the intensity values at each image point as "matching primitives", and a binary function for the similarity evaluation. That is, if the image intensity at a given point in the left image array (x_L,y), and at some point (x_R,y) in the right image array, are both greater than zero, then the node $\mathbf{M^I}(x_L,x_R,y)$ in the preliminary (and initially also the secondary) network will represent a potential match and be assigned a value of 1. If there is no image intensity at one, or both, of these points, the node $\mathbf{M^I}(x_L,x_R,y)$ will not represent a match and will hold a value of 0.

Because the ordering constraint is a key principle in the model, it should be obvious that the current (preliminary) matching strategy is not well suited for natural images that contain large surfaces regions with homogenous intensity values; since surface points within such regions can not be meaningfully ordered. However, for sake of demonstrating the proposed mechanism the current preliminary stage is sufficient, since all examples below are either random-dot, or simple (vertical) line, stereograms with binary intensity values (black/white). The advantages with using artificial, over natural, stereograms as performance probes are that the stimuli can be precisely controlled and arranged as desired, and more importantly, an accurate disparity map is available for validation, which is usually difficult to obtain for natural image stereograms.

Assuming the preliminary stage has matched each image point, with all points in the opposite image that lie within a horizontal disparity range of +/-D pixels (in the examples below D=12); and the result have been loaded into the secondary network, the value of any node in the (secondary) network is given by the semi-saturation function:

$$f(x,\sigma) = K \cdot \frac{x^2}{x^2+\sigma^2}$$

which reaches half K when $x=\sigma$. Here, x and σ are compound terms representing the total excitatory respectively inhibitory input to the node. More specifically (with K=1), given a particular node $\mathbf{M^{II}}(x_L,x_R,y)$ in the secondary network, its value at iteration $t+1$ is given by:

$$M^{II}_{t+1}(x_L,x_R,y) = \frac{\left[M^{II}_t(x_L,x_R,y) + M^{I}(x_L,x_R,y)\cdot A\cdot e^{-B\cdot S_t(x_L,x_R,y)}\right]^2}{\left[M^{II}_t(x_L,x_R,y) + M^{I}(x_L,x_R,y)\cdot A\cdot e^{-B\cdot S_t(x_L,x_R,y)}\right]^2 + \left[\sigma + C\cdot S_t(x_L,x_R,y)\right]^2}$$

where $\mathbf{M^I}(x_L,x_R,y)$ is the node in the preliminary network holding the result of the initial matching. *A, B, C* are constant weight factors, and σ is the semi-saturation constant. Finally, the term $\boldsymbol{S_t}(x_L,x_R,y)$ is the sum of, the inhibitory input, from the Near AND-gate, $\mathbf{N_t}(x_L,x_R,y)$, and the Far AND-gate, $\mathbf{F_t}(x_L,x_R,y)$, connected to node $\mathbf{M^{II}}(x_L,x_R,y)$:

$$S_t(x_L,x_R,y) = N_t(x_L,x_R,y) + F_t(x_L,x_R,y)$$

$$N_t(x_L, x_R, y) = \sqrt{\sum_{d=1}^{D+(x_R-x_L)} M_t^{II}(x_L+d, x_R, y) \cdot \sum_{d=1}^{D+(x_R-x_L)} M_t^{II}(x_L, x_R-d, y)}$$

$$F_t(x_L, x_R, y) = \sqrt{\sum_{d=1}^{D-(x_R-x_L)} M_t^{II}(x_L, x_R+d, y) \cdot \sum_{d=1}^{D-(x_R-x_L)} M_t^{II}(x_L-d, x_R, y)}$$

6 Simulation results

For all the examples presented below the same parameter set was used (A=σ=0.5, B=8 and C=4). As of yet, no proper formal analysis of the parameter space has been carried out, but the above values were empirically found to be a good trade-off between speed-of-convergence and the number of correct matches. In each stereo triplet, the left and middle images can be free-fused with un-crossed eyes, and the middle and right images with crossed eyes.

6.1 Example 1: Vertical lines

The first example (figure 8) contains several of the simple cases described in previous sections, and illustrates the basic implications of the conditional uniqueness constraint both in cases when there is an even number of similar features in the two halves of the stereogram, and in cases where there is an uneven number of similar features. From top to bottom, the first and second row are examples of the Panum's limiting case, and the corresponding "trivial" case, respectively, from figure 3. The third row contains the variation of the "double-nail" illusion illustrated in figure 4c. In each of the last three rows, there is an even number of features in the two image-halves. In row four, all lines lie in a plane at zero disparity (the example from figure 6). In row five, they form a peak with the middle bar on top, and in row six, they form a peak and a trough (from left to right).

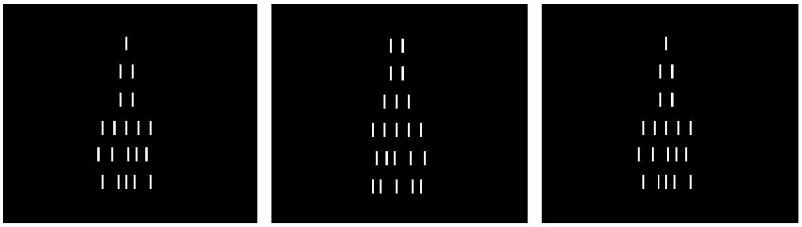

Figure 8. The left and middle images can be free fused with un-crossed eyes, and the middle and right images with crossed eyes. Row: 1. Panum's limiting case; 2. Its "trivial" case; 3. Variation of the double-nail illusion; 4. Plane at zero disparity; 5. Peak/Pyramid; 6. Peak and through.

The output is displayed in figure 9 as horizontal cross-sections where the rightmost horizontal line mark zero-disparity. As can be seen, the model allows both of the two matches (fig 9-1) in the Panum's limiting case, but only the two ordered matches (fig 9-2) in its corresponding "trivial" case. Case three corresponds to the variation of the double-nail illusion (figure 4c). As explained above (in the context of figure 4d), when subjects were presented with this stimuli they reported of seeing both of the matches $M_{1,1}$ and $M_{2,3}$ (the two visible ones in figure 9- 3), but also either $M_{2,2}$ if fixation was on $M_{1,1}$, or $M_{1,2}$ if the fixation was on $M_{2,2}$. This seem to suggests that human stereopsis to some degree is biased towards matches that lie in the plane of fixation. However, in the current implementation no such bias has been introduced, and therefore the two matches (in fig 9-3), that correspond to the matches $M_{1,2}$ and $M_{2,2}$ in figure 4d, are equally suppressed (not visible in fig 9-3). The final three cases are quite straight forward, and as the output shows (figure 9-4, 9-5 and 9-6) there are no instances of multiple matching since there exists a one-to-one correspondence between the features.

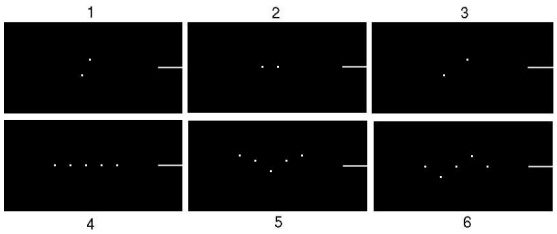

Figure 9. Horizontal cross-sections of the network, after it has converged (with the input in figure 8) into a stable state. The right horizontal line marks zero disparity.

6.2 Example 2: RDS - Occluding square

The random-dot stereogram in figure 10 contains two opaque surfaces: a background plane at zero disparity, and an occluding square in the foreground. The dot-density in this particular example is 10 %, and the disparity difference between the two planes is 4 pixels. Figure 11 shows a 3-D plot of the model output. The amount of correct, and false, matches were 93.3%, and 10.5%, respectively. 6.7% of the dots were unmatched (see table 1 for results with different dot-densities).

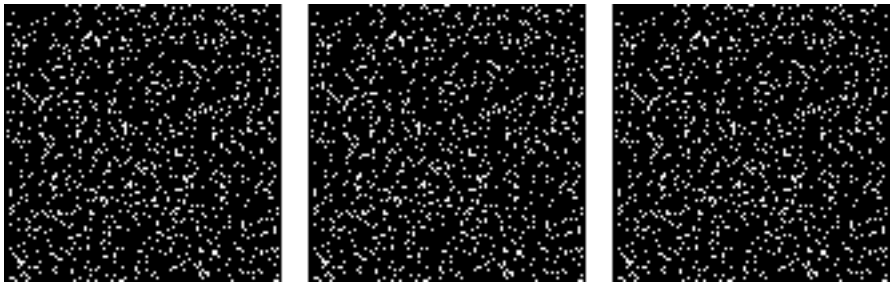

Figure 10: Random-dot stereogram displaying a smaller occluding square, in front of a background surface at zero disparity.

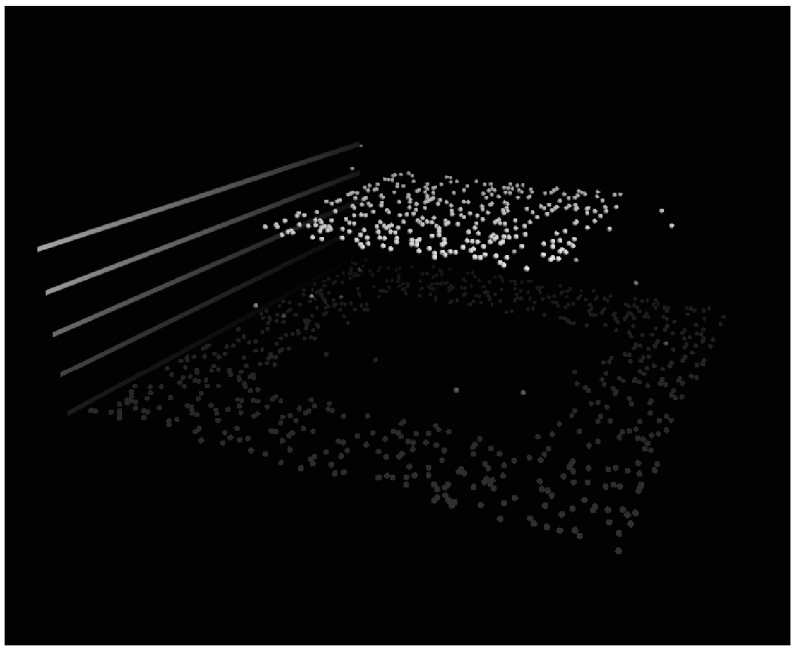

Figure 11: 3-D plot of the model output for the stereogram in figure 10. Each dot represents an active node in the network. The horizontal lines (to the left) mark the separation between different disparity layers.

1. 6.3 Example 3: RDS - Gaussian "needle"

In the third example (figure 12), the disparity (dX_m) in the middle image, relative to the left (right), image was generated by the following Gaussian-like distribution:

$$dX_m = -10 \cdot e^{-\frac{(I-x)^2+(I-y)^2}{\sigma^2}}$$

where I is half the image width (I=64 pixels) and σ=12. When fused, a sharp peak is perceived pointing out from a flat background. At the steepest part (between the background and the peak), the disparity-gradient is approximately 0.7. Figure 13 shows a 3-D plot of the model output (see also table 1). This example is an interesting test because it should pose a problem for models that assume smooth (particularly fronto-parallel) surfaces, and rely on some kind of spatial summation (i.e. mutual support between neighbouring matches with equal or similar disparities) to solve the correspondence problem. What should pose a problem for such models is that, at the top of the peak, there are only a very few dots. The support that the correct matches give each other could easily be "drowned" by the more massive support that the background could give to potential false matches with disparities closer to the background (see also discussion).

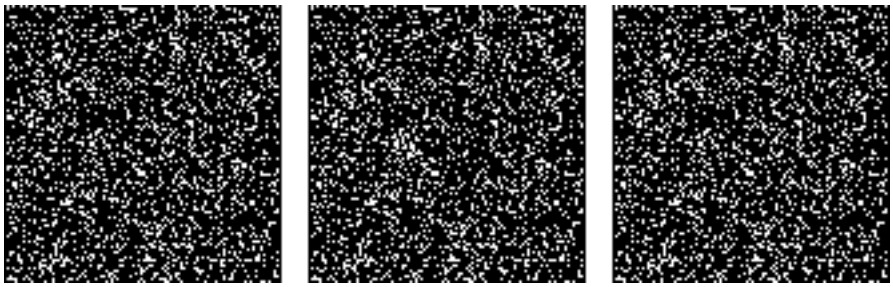

Figure 12: RDS of a flat opaque surface with a sharp narrow peak coming out in the middle. Dot-density 20%.

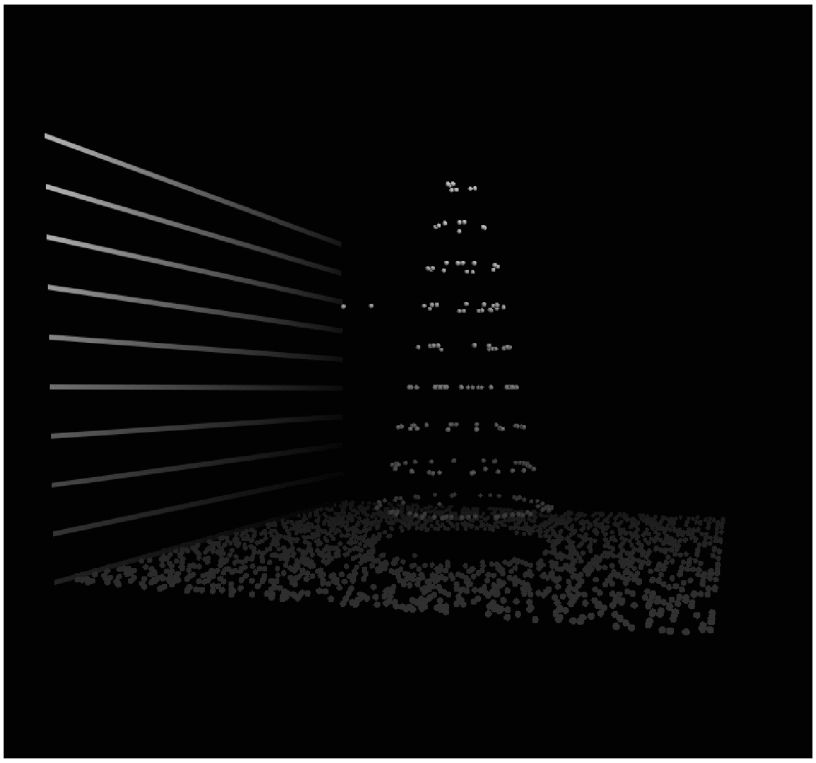

Figure 13: 3-D plot of the model output for the stereogram in figure 11.

Stimuli (dot-density %)	*Matches*			*Iterations*
	Correct (%)	*False* (%)	*Unmatched* (%)	
Square				
5%	98.0	4.6	2.0	24
10%	93.3	10.5	6.7	51
15%	91.6	12.3	8.4	72
20%	88.6	12.6	11.4	162
"Needle"				
5%	100.0	0.0	0.0	31
10%	99.3	0.8	0.7	65
15%	98.9	1.1	1.1	85
20%	98.0	1.3	2.0	126
Transparency				
5%	93.4	5.8	6.6	23
10%	82.3	16.0	17.7	42
15%	72.9	25.4	27.1	60
20%	65.5	33.0	34.5	121
"Needle" + Transparency				
5%	96.6	3.0	3.4	21
10%	89.7	9.5	10.3	44
15%	80.4	18.3	19.6	71
20%	72.8	24.6	27.2	116
RDRDS				
5%	95.3	3.8	4.7	16
10%	84.8	14.5	15.2	34
15%	80.0	19.6	20.0	72
20%	68.3	30.9	31.7	107

Table 1: Performance results of the implementation when tested on example 1-6 above, with different dot-densities. The rightmost column shows the number of iterations required for the network to settle down to a stable state; i.e. when there was no, or very small (<0.001%), change in the activity between successive iterations.

6.4 Example 4: RDS - Transparent layers

Because uniqueness is not an absolute constraint in the model, multiple transparent surfaces do not pose as serious a problem as it does to models that allow only a single disparity value at any image point. However, as the number of planes, or the dot-density, rises, naturally, the chance increases that the relative ordering of adjacent features within the two image halves will be jumbled. The fact that the model is cooperative does to some extent counteract such mismatching, if there are strong unambiguous matches in the near vicinity. In example 4 (figure 14) there are two transparent planes with a disparity separation of 4 pixels. Note in table 1 how the performance deteriorates with increasing dot-density. Such a deterioration, with increasing dot-density, has also been described in human vision (see Akerstrom & Todd, 1988).

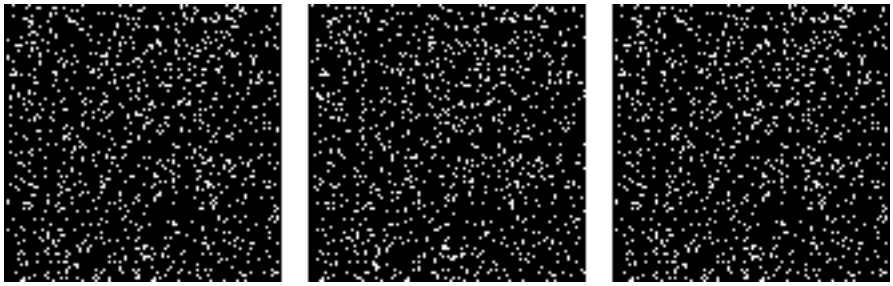

Figure 14: RDS containing two transparent planes, separated by a 4 pixel disparity. Dot density 10%.

6.5 Example 5: RDS - Gaussian "needle" with one transparent plane

In example 5 (figure 15) a transparent layer have been added to the needle in example 3. The needle constitutes an opaque background, while the transparent layer cuts through the peak half way up.

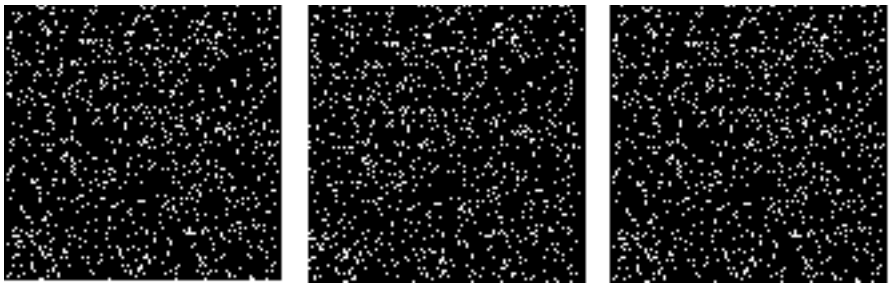

Figure 15: RDS of an opaque background surface that has a peak in the middle, which sticks through an additional flat transparent surface. Dot-density 10%.

6.6 Example 6: Random-Disparity Random-Dot Stereogram (RDRDS)

In the random-dot stereogram, figure 16, the disparity between each pair of corresponding dots was randomly set, and lies in the interval from -3 to 3 pixels around zero disparity. This type of stereogram is clearly a challenge to models that rely on assumptions about surface smoothness, since there is no correlation whatsoever between the disparity values of any two neighbouring dots. Surprisingly, this type of stimuli seem quite easily fused as long as the dot-density is relatively low, and the disparity interval is not too large (according to a highly informal study where two members at our department participated as subjects).

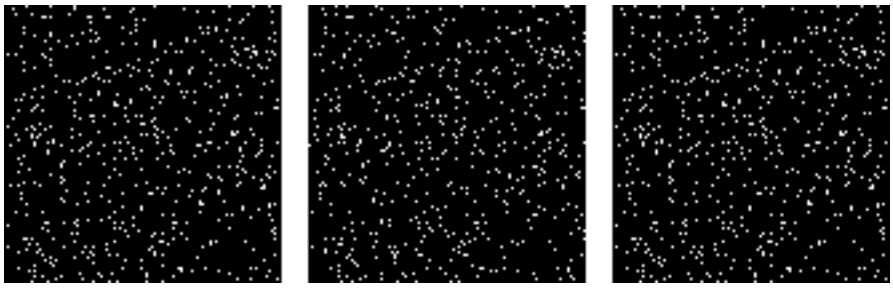

Figure 16: Random-dot stereogram with random disparities between corresponding dot-pairs. Dot-density 5%.

6.7 Example 7: RDS - Weinshall stimuli

Figure 17 shows an example of the basic stimuli used in Weinshall's studies (1991, 1993). What is particularly interesting with this stimuli is that it shows that the human visual system treats the same stimuli different when it is seen in isolation from when it is seen *en masse*. Recalling that in the (single) case of the double-nail illusion, subjects saw only the two ordered matches. In the "Weinshall-stimuli", on the other hand, where the same stimuli has been copied multiple times into the same stereogram, subjects saw up to four planes (corresponding to the disparities of both the ordered and the ghost matches). A number of variations of this stimuli were tested on the model (see Table 2), where the dot-density was the same as in the Weinshall study (9% x 2 = 18% total), and the inter-dot separation (G_L and G_R) was varied. Each column/example in table 2 shows the averaged result of three different stereograms with the same values of G_L and G_R. Interestingly, but unanticipated, the model produced results that had the same basic characteristics as that of human vision in this respect. When the inter-dot separation in one image was zero, but greater than zero in the other image (i.e. multiple copies of Panum's limiting case), the great majority of matches was concentrated at zero disparity, and the disparity corresponding with the non-zero inter-dot separation. The remaining matches were fairly evenly distributed over all other disparities. When the inter-dot separation was the same (G_L=G_R) in the two image halves, the output was concentrated to one single plane at zero disparity. When G_L and G_R was different from each other (and both greater than zero), again the majority of activity/matches was in the two planes corresponding to the ordered matches, but there was also some activity at several other disparities with clear peaks at the two planes corresponding to the two ghost matches.

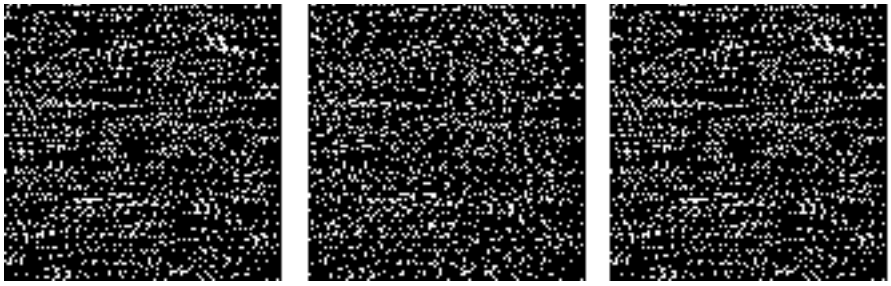

Figure 17: Example of the stimuli used by Weinshall (1991, 1993), with GL=4, and GR=8. Dot-density 18%.

In Weinshall's study (1991) subjects saw between two and four planes with stimuli like this. Seeing to the absolute number of matches only (in the model output), there were relatively few matches in the two planes corresponding to the two ghost matches, and even fewer than in some of the other layers which should not produce a perception of a plane. However, first of all, it only takes a dot-density as low as 0.001 (Weinshall, 1991) to produce a sensation of a plane in these stimuli, which in the examples above corresponds to about 10 dots. So clearly, the number of dots in the ghost planes are above that value in both of the examples, in table 2, of this type (G_L=4, G_R=7 and G_L=6, G_R=3). Further, it is not clear whether it is the absolute number of matches, or some relative measure that creates the sensation of a plane in human perception. In the example below with G_L=4 and G_R=7, there are for example only 16 matches in the "ghost" layer (at disparity 7), but more than that in several of the other layers (4, 2, 1, -1) which should not produce the perception of a plane. On the other hand, of all the planes that contain more than 10 matches in them, only in the layers -4, 0, 3 and 7 does the number of matches reach local peak values. Whether or not this may explain the psychophysical observations for this type of stimuli remains to be seen.

In the final two examples (last two columns in table 2), all accidental matches were removed from the stereograms. In Weinshall's study (1991), accidental matches were avoided by spacing the dots so that the disparities of possible accidental matches were larger in absolute values than the possible disparities for each corresponding (left-right) dot-pair. With the same procedure for removing accidental matches the model output was concentrated exclusively to the two planes corresponding to the ordered matches of each dot-pair, which is in accordance with the results obtained by Weinshall. However, not surprisingly - considering that the ordering principle is the key constraint in the model - the model produced the same output as long as the separation between any two (following) dot-pairs was at least one pixel larger that the value of *max*(G_R,G_L). This spacing

preserves the relative ordering of the dots, but it does not guarantee that the possible disparities (in absolute values) between the dots within each (left-right) dot-pair is smaller than that between possible accidental matches.

	G_L=0, G_R=3 (**0,3**) It: 105	G_L=2, G_R=2 (-2 **0**,2) It: 93	G_L=4, G_R=7 (**-4,0,3,7**) It: 135	G_L=6, G_R=3 (**-6,-3,0,3**) It: 139	G_L=6, G_R=3 (-6,**-3,0**,3) No accid. It: 63	G_L=2, G_R=4 (-2,**0,2**,4) No accid. It: 67
Disparity (pixels)	**Number of active matches**					
12	17	0	2	0	0	0
11	18	0	1	1	0	0
10	25	0	2	1	0	0
9	25	0	2	3	0	0
8	20	0	2	1	0	0
7	25	0	**16**	1	0	0
6	28	0	9	1	0	0
5	32	0	10	3	0	0
4	31	0	28	3	0	0
3	**468**	0	**593**	**29**	0	0
2	52	0	107	15	0	**729**
1	48	0	107	18	0	0
0	**512**	**1567**	**625**	**617**	**694**	**747**
-1	32	0	32	108	0	0
-2	36	0	11	113	0	0
-3	33	0	10	**564**	**647**	0
-4	32	0	**26**	16	0	0
-5	25	0	2	15	0	0
-6	23	0	1	**31**	0	0
-7	18	0	2	5	0	0
-8	21	0	1	3	0	0
-9	19	0	1	3	0	0
-10	12	0	0	4	0	0
-11	14	0	1	2	0	0
-12	10	0	1	1	0	0

Table 2: Model output with different values for the left (G_L), and right (G_R), image inter-dot separation, in stereograms of the same type as used by Weinshall (1991, 1993). The number of active matches, and the number of iterations (It.) required, shown in each column are the average results for three different stereograms (with the same values of G_L and G_R in each trial). The numbers within parenthesis (second top row) mark at what disparities matches are possible, if only corresponding (left-right) dot-pairs are considered; i.e. all possible matches between the dots in a left image dot-pair, and the dots in the corresponding right image dot-pair (the copy). The numbers in bold correspond to disparities where Weinshall's subjects reported of seeing depth planes. No accid. stands for no accidental matches.

7 Discussion

7.1 Relation to other models

The one major feature of the proposed model that separates it from, possibly all, previous models of stereopsis (and other algorithms devoted to the correspondence problem) is its more relaxed version, or interpretation, of the uniqueness constraint. Because multiple matches are not penalised when a one-to-one correspondence does not hold between the left and right image halves, the model naturally explains why humans perceive double (e.g. Panum's limiting case) or multiple (e.g. "Weinshall-stimuli") instances of the same stimuli in the examples reviewed above. This feature of the model and the fact that no direct assumption is made about surface smoothness (only indirectly via the ordering constraint) also means that the model handles transparency fairly well (possibly in line with human performance), as long as the relative ordering of the stimuli is not broken up too much.

Due to the basic difference in interpretation, and implementation, of the uniqueness constraint compared to the traditional interpretation (Marr & Poggio, 1976), which has been the prevailing one, it is somewhat difficult to make a direct comparison to any previously described model. Apart from the uniqueness constraint, two key features of the model is that it 1) relies on the ordering constraint, and 2) is cooperative.

The ordering constraint has been explicitly used in previous models by for example Baker and Binford (1981) and Ohta and Kanade (1985). Although these models differ in detail, both models find a solution to the correspondence problem by matching the edges (Baker & Binford, 1981) or the intervals between edges (Ohta & Kanade, 1985), in the input stereogram, according to the ordering principle. And moreover, they also share the central idea of utilising edges that run vertically across each image half to guide this process. Neither of these models however address the phenomena of multiple matching, since they were not designed to explain human perception, but designed on rather strict computational grounds. Moreover, Ohta and Kanade use interpolation to determine the disparity at positions where there are no edges available. Consequently, it is unlikely that their model can handle stereograms that contain transparency. It is unclear how the "Baker -Binford" model would handle transparency.

Three other models that share at least some traits with the one proposed here are the ones suggested by: Marr and Poggio (1976), Pollard et al. (1981) and Prazdny (1985). What all of these models share is that they i) are cooperative, ii) are mainly concerned with the correspondence problem as such, and not the nature (or composition) of the matching primitives, and iii) were all designed - at least in part - to explain human stereopsis, or aspects thereof. Each of these will be briefly discussed below.

The model of Marr and Poggio (1976), while not being the first cooperative one, is probably one of the most well known. The basic idea, or assumption, in their model is that surfaces in general are opaque and cohesive, i.e. smoothly changing in depth. This assumption led to the formulation of the uniqueness constraint, and the cohesitivity (or continuity) constraint; which in their implementation, in turn, were translated into rules stating that matches representing different disparities lying along the same lines-of-sight should inhibit each other (uniqueness), while matches that were close to each other (within some radius) and had the same disparity should support each other (cohesitivity). Their implementation could successfully solve certain types of random-dot stereograms (preferably consisting of smooth opaque surfaces), but it had significant shortcomings. First, their particular interpretation, and implementation, of the cohesitivity constraint (i.e. support between neighbouring matches with the same disparity), makes the model biased towards fronto-parallel surfaces. Consequently, it is not well suited for resolving stereograms of, for example, tilted or jagged surfaces. Another aspect of this shortcoming (discussed earlier in the context of example 3; the Gaussian "needle"), which the authors also pointed out, is that "the width of the minimal resolvable area increases with disparity". This simply means that, for example, a small surface patch in front of a background surface, becomes increasingly difficult to resolve (from the background) as the disparity difference between the surfaces increase.

Finally, as already discussed, their strict formulation of the uniqueness constraint does not allow the model to account for, neither, the phenomena of multiple matching in human perception, nor transparency, since the model can not resolve multiple overlapping surfaces or even represent them.

The PMF model proposed by Pollard et al.(1985) is highly similar to the model of Marr and Poggio (1976) in the interpretation of the uniqueness constraint, and in that neighbouring matches (representing similar disparity values) support each other. However, a major difference is that while the support region, in the latter

model, is basically a 2-D disc centred around each match, it is a 3-D space in the former. More specifically, in the PMF model, the width (in disparity) of the support region grows with the distance from a match, and is bounded by the surface were the disparity gradient is equal to 1. The disparity gradient is defined, given two binocularly visible points in space, as the difference in disparity (between the points) divided by their cyclopean separation (Burt & Julesz, 1980). The choice of using a disparity-gradient limit of 1 was made partly due to the computational argument that binocularly corresponding features in most natural scenes seldom exceed a disparity-gradient of 1, and partly due to the argument that the human visual system seems to have such a limit (Pollard et al., 1985). A major advantage, over using a "flat" support region, is that the PMF model is not biased towards fronto-parallel surfaces, but handles e.g. slanted and jagged surfaces (within the disparity limit of 1) well. However, regarding multiple matching and transparency, the PMF algorithm too falls short; since at any given image point, only the strongest match is kept and all others discarded. Thus, multiple matches and transparent surfaces can not be represented simultaneously. It is unclear if, or how, the PMF model could be modified to handle multiple matching and transparency as well, and at the same time keep its disambiguating power since, for example, in the Panum's limiting case the disparity gradient between the two bars is exactly 2, and hence well beyond the imposed limit. See also Weinshall (1993) for why the PMF model fails to account for her results.

Prazdny's (1985) model does handle transparency and can (if modified), to some extent, also account for the results in Weinshall's studies (see Weinshall, 1993). The disambiguating power of Prazdny's model has the same basic motivation as the previously two described models, i.e. that the disparity difference between neighbouring points on a surface, in general, will be small. In his model, matches support each other according to a Gaussian measure that essentially decreases with increasing distance and/or disparity-difference between the matches. The major difference from the previous two models, however, lies in that Prazdny does not assume opaque surfaces, and therefore there is no (explicit) penalty, or inhibition, in his model between matches that represent different disparities. This allows the model to represent, and resolve, stereograms that contain transparent surfaces, since each surface can be said to support itself without any interference from possible others. On the other hand, the model does not allow multiple matches at any single image point. If there are still ambiguous matches left at any image point, after the matches with the strongest support have been selected, these are simply merged into one and the disparity is determined by interpolation. Thus, in its original form, the model can not account for multiple matching.

Also worth mention is that although there are no explicit inhibitory connections in Prazdy's model, there are indeed such at the effective level (or level of implementation); both in the process of choosing the supporting matches (i.e. given any match, i, only the best supporting match, j, contributes to the activity increment of i), and in the final selection of the matches that have the highest activity. From a computational perspective it changes nothing whether you choose to call an operation "selection" or "inhibition" as long as it performs the same function, but the point here is that, from a biological perspective, it does matter how such a selection processes could be implemented in neural circuitry. What perhaps is particularly questionable (from this perspective) is why *only* the best supporting match should contribute to any given match's activity level, and how this could be neurally implemented. Considering any possible match in a dense random-dot stereogram, it is difficult (but of course not impossible) to see how a neural mechanism could be so finely balanced as to, from a set of many possible matches with similar support values, let through only the support from the strongest, and at the same time block the support from all others. However, this critique may not be serious to Prazdy's model, and it certainly does not diminish the elegance of it, but it certainly would be interesting to see if a more biologically oriented implementation could perform equivalently.

From the perspective of human stereopsis, a more general critique can be directed, not only to the three models discussed above, but to all models that rely on surface smoothness for disambiguating power. While it is true that surfaces in general are smoothly changing in depth, and that this clearly can be a powerful constraint; it is not necessarily the case that the human visual system make use of this constraint, at least not at the level where stereopsis is achieved. Surface perception (reconstruction) does not necessarily have to be as tightly coupled, or integrated, with stereopsis as these models suggest, but it is quite possible that these are fairly separate processes. The type of stimuli used in example 6 (figure 16), the random-disparity random-dot stereogram (RDRDS), should be a good probe for testing whether human stereopsis is actually biased towards producing matches that appear as cohesive surfaces, even when no obvious surface cues are available. If we are not so biased then there is no reason to believe that it is an important constraint in human stereopsis.

7.2 Possible neural mechanisms

In the current model and implementation, a number of important simplifications have been made, particularly regarding the preliminary matching stage, which makes it difficult to say to what degree it can be considered a model of human stereopsis. Needless to say, however one chooses to look at it, the model lacks too many of the known features of human stereopsis to be called a complete model of human stereopsis. On the other hand, despite the many simplifications made, the close agreement between the model output and the reviewed psychophysical data does seem to suggest that - while the model as a whole may not map onto human stereopsis - the proposed mechanism (i.e. the dual inhibitory near-far AND-gates) may very well have some correlate in human depth perception.

Exactly what this correlate might consist of, or correspond to in human vision, is of course more difficult to specify. It is however tempting- perhaps dangerously so - to make the association from the Near and Far AND-gates proposed above, to the Near and Far cells described by Poggio and Fisher (1977). Could these two mechanisms have anything in common other than parts of their names?

By measuring the impulse activity of cells in foveal striate (A17) and prestriate (A18) cortex of the Rhesus monkey, Poggio and Fisher (1977) classified neurones, depending on their response to binocular stimuli, into four different categories: Tuned Excitatory, Tuned Inhibitory, Near or Far. The Tuned Excitatory cells were ocularly balanced and gave an excitatory response over a narrow range near the fixation distance. The Tuned Inhibitory cells were ocularly unbalanced and responded to stimuli from the dominant eye over a wide range of disparities except near the fixation distance. The Near cells were also (predominantly) ocularly unbalanced and responded to stimuli, over a broad range, in front of the fixation distance, but not at, or beyond it. Finally, the Far cells mirrored the Near cells, in that they were ocularly unbalanced and responded to stimuli over a broad range, but differed in that they only responded to stimuli beyond the plane of fixation.

The Near and Far cells have two properties that, in combination, makes them seem particularly fit as components in a neural implementation of the proposed Near/Far AND-gate mechanism. First, they only respond to stimuli that lies either in front of (Near cells), or behind (Far cells) the plane of fixation. Second, they respond to stimuli over a broad range of disparities, which may suggests that they receive their response properties by summation of a number of subunits sensitive to different disparities. Given these two properties it is not difficult to imagine how a group of two Near, and two Far, cells could subserve one AND-gate each, as depicted in figure 18. What is missing in the data from Poggio and Fisher's study is any direct evidence of cells with response properties like the proposed AND-gates themselves. However, this is not surprising since they did not use ambiguous stimuli in their study. Using only a single bar, or some similar stimuli, there will never be any ambiguity and hence never any simultaneous activation of either two near, or two far, neurones that could drive the postulated AND-gates. If any direct evidence of these AND-gates is to be found, some ambiguous, preferably repetitive, stimuli will have to be used.

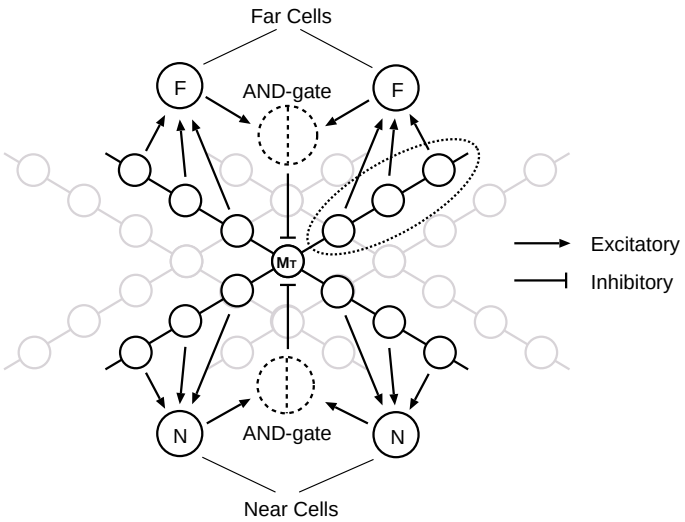

Figure 18: Schematic view of how two Near, and two Far cells, could be arranged to subserve inhibitory cells with possible AND-gate properties.

Finally, regarding the fact that the great majority of the Near and Far cells were ocularly unbalanced; it seems that such a distinguishing feature should reflect some major functionality of the cells. Unfortunately, the study of Poggio and Fisher did not reveal sufficient details about the nature of this property to allow for any (here) meaningful speculation of how it could fit with the proposed model, but it is nevertheless worth mentioning that several different reconcilable interpretations are possible.

References

Marr, D & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science, 194*, 283-287

Mayhew, J.E.W. & Frisby, J. P. (1981). Psychophysical and computational studies towards a theory of human stereopsis. *Artificial Intelligence, 17*, 349-387

Baker, H.H. & Binford, T.O. (1981). Depth from edge and intensity based stereo. *In Proceedings of the 7th IJCAI* (Los Altos, CA: William Kaufmann), 631-636

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

Pollard, S. B., Mayhew, J. E. W. & Frisby, J. P. (1985). PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception, 14*, 449-470

Krol, J. D. & van de Grind, W. (1980). The double-nail illusion: experiments on binocular vision with nails, needles, and pins. *Perception, 9*, 651-669

Weinshall, D. (1991). Seeing “ghost” planes in stereo vision. *Vision Research, 31*, 1731-1748

Weinshall, D. (1993). The computation of multiple matching doubly ambiguous stereograms with transparent planes. *Spatial Vision, 7*, 183-198

Marr, D. & Poggio, T. (1979). A computational theory of human stereo vision. *Proc. R. Soc. Lond. B. 204*, 301-328

Marr, D. (1982). *Vision*. New York: W. H. Freeman and Company.

Nakayama, K. & Shimojo, S. (1990). Da Vinci stereopsis: Depth and subjective occluding contours from unpaired image points. *Vision Research, 30*, 1811-1825

MaKee, S. P., Bravo, M. J., Smallman, H. S. & Legge, G. E. (1995). The “uniqueness constraint” and binocular masking. *Perception, 24*, 49-65

Akerstrom, R. & Todd, J. T. (1988). The perception of stereoscopic transparency. *Perception & Psychophysics, 44*, 421-432

Ohta, Y. & Kanade, T. (1985). Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-7*, 139-154

Prazdny, K. (1985). Detection of binocular disparities. *Biological Cybernetics, 52*, 93-99

Poggio, G. F. & Fisher, B. (1977). Binocular interaction and depth sensitivity in Striate and Prestriate Cortex of behaving Rhesus monkey. *Journal of Neurophysiology, 40*, 1392-1405