Vision Res. Vol. 24, No. 10, pp. 1133–1140, 1984
Printed in Great Britain. All rights reserved

0042-6989/84 $3.00 + 0.00

VERTICAL IMAGE REGISTRATION IN STEREOPSIS

K. R. K. Nielsen and T. Poggio

Massachusetts Institute of Technology, Department of Psychology and Artificial Intelligence Laboratory,
Cambridge, MA 02139, U.S.A.

(*Received* 7 *November* 1983; *in revised form* 14 *April* 1984)

Abstract—Most computational theories of stereopsis require a registration stage prior to stereo matching to reduce the matching to a one-dimensional search. Even after registration, it is critical that the stereo matching process tolerate some degree of residual misalignment. We have studied the tolerance to vertical disparity in situations in which false targets abound and corrective eye movements are eliminated. Our main results are:

(a) vertical disparity of only the central "figure" part of a random dot stereogram can be tolerated up to about 3.5′, and

(b) vertical disparity of the "figure + ground" is tolerated up to about 6.5′ in the presence of monocular cues to vertical disparity.

Our data suggest that this tolerance is attained by two non-motor mechanisms:

(1) the spatial average performed by the receptive fields that filter the two images prior to stereo matching, and

(2) a non-motor shift mechanism that may be driven at least in part by monocular cues.

Stereopsis Vertical disparity Image registration

INTRODUCTION

Recent attempts at building artificial stereo systems have clarified the computational problems in stereopsis. The extraction of disparity information from two images may be decomposed into two principal steps. The first step is the identification of the same object point in the two images—this is the correspondence problem for stereopsis. Julesz's (1971) experiments with random dot stereograms demonstrated that the human visual system can solve this problem even in the presence of abundant false matches and without monocular cues. The second step is the actual measurement of the disparities. Additional information and computations are of course required to convert the resulting disparity map to absolute depths. It is the first step in the computation of the disparity map—the correspondence problem—which is most difficult and it is primarily differences in approaches to its solution that distinguish the various stereo algorithms that have been proposed (e.g. Marr and Poggio, 1979; Mayhew and Frisby, 1981; Baker and Binford, 1981).

One simplification which is commonly made in the solution of the correspondence problem is based on the so-called epipolar constraint. It reduces the correspondence problem to a one-dimensional search by making use of the fact that possible matches for a point in one image lie along a line in the other image known as an epipolar line. Such one-dimensional matching schemes require, however, that the two images be precisely registered. Registration problems are introduced by simple misalignment and by differences in the imaging properties (optics and sampling grid) of the two sensors. Even when the images are globally registered, an additional complication is introduced by the geometry of stereo viewing. All of the epipolar lines are not horizontal, in general. Vertical disparities are introduced when a "horizontal scan line" search scheme is used, for example, on images obtained with convergent visual axes or when the fixation point is not "straight ahead", because the epipolar lines are not horizontal under these viewing conditions. These have proven to be serious problems for computer implementations of stereo algorithms, many of which use the horizontal scan lines of their input cameras as approximations to the true epipolar lines.

Mayhew and Longuet-Higgins (1982) have recently suggested that the vertical disparities resulting from the geometry of stereo viewing, rather than having to be "tolerated", may be used to derive information about the distance and direction of the fixation point in order to calculate absolute depth. If this idea is correct, the correspondence process should be effectively two dimensional. It should, in addition, precisely measure vertical disparities of less than 9′ for most situations and less than 4′ for usual stereo viewing conditions. As we will see, our results, although better interpreted in terms of a 1-D search process, do not rule out Mayhew and Longuet–Higgins' proposal.

There have been a number of other psychophysical studies of vertical disparities reported in the literature. Some of these have been concerned with measuring the vertical extent of Panum's area, that is, the range of vertical disparities that can be fused. Schor and Tyler (1981) measured the horizontal and vertical extents of Panum's area using stimuli consisting of two spatio-temporally modulated sinusoidal lines. Both the horizontal and vertical disparity ranges for fusion vary inversely with spatial frequency. The horizontal range increases at low temporal frequencies while the vertical range is largely independent of temporal frequency. The vertical extent decreases from about 8′ at 0.125 c/deg to about 1.5′ at 2 c/deg. Duwaer (1982) used an afterimage method to measure the non-motor component of the vertical fusion range and found a value of 8–15′. The targets he used consisted of a 2.5° square with or without 50 horizontal lines defining a surround 57° in diameter. Fender and Julesz (1967) used stabilized image conditions to measure horizontal and vertical fusion ranges without eye movements. For line stimuli, they found that fusion occurred when the targets were brought within 9–14′ vertical disparity of each other. For random dot stereograms, they found fusion occurred within 1–9′ vertical disparity.

In this paper, we study specifically the tolerance to vertical disparity in conditions in which potential false matches abound and corrective eye movements are eliminated.

METHODS

Stimuli

The first experiment was designed to explore the effect of vertical misalignment on the ability of our subjects to fuse random dot stereograms for a range of horizontal disparities spanning Panum's area. The stereograms were 54′ square with 50% density black and white 54″ square elements. Horizontal disparities were applied to the central 25% area ("figure") in the form of a square or diamond. Disparities were in multiples of the dot size and evenly divided between the two halves of the stereogram so that no monocular cues to depth were introduced. Stereograms were generated with 0, 1.8, 5.4, and 10.8′ crossed and uncrossed horizontal disparities and 0, 1.8, 5.4 and 10.8′ vertical disparities. A second set of trials used similar stereograms, but with 0, 3.6 and 7.2′ vertical disparities. Two stereograms were generated for each vertical disparity—one with the left component of vertical disparity upward and the right component downward and one with the right component upward and the left downward. These were presented in random sequence so that subjects could not predict the direction of vertical disparity, thereby precluding compensatory eye movements during the brief stimulus presentation. Note that this procedure for generating and displaying the stereograms also ensured that there were no non-disparity cues (such as the location of an unfused patch) to the "sign" of the horizontal disparity (crossed or uncrossed). This is not the case when (as has sometimes been done) the disparity is introduced by shifting the figure in only one half of the stereogram. Then in the absence of successful stereopsis, the location of an unfused patch could serve highly-trained observers as a cue to the sign of disparity. We emphasize that such cues were not available in our stereograms. Finally, our subjects reported seeing depth in the flashed stereograms.

In order to obtain some indication of the factors that might influence tolerance to vertical misregistration, we used two vertical disparity conditions. In one set of stereograms, only the figure was given vertical disparity ("figure only" condition). In a second set, the figure and ground were vertically misaligned ("figure + ground" condition). There are a number of differences between these two classes of stimuli. In the "figure only" condition, there are no monocular cues to vertical disparity, only 25% of the area of the stereogram has information about the amount of vertical misalignment and no global registration process (such as eye movements) can bring the entire stereogram into registration. In contrast, the "figure + ground" condition (a) does have monocular cues to vertical disparity (the vertical misalignment of the top and bottom edges of the two halves of the stereogram), (b) 100% of the area of the stereogram gives information about the amount of vertical misalignment (possibly important for an area-based registration process) and (c) global registration of the entire stereogram (by vertical eye movements, for example) is possible.

The second experiment was designed to give some indication of which of the differences between the stimuli in the "figure only" and "figure + ground" conditions were responsible for the observed difference in performance. We removed the monocular cue to vertical misalignment of the figure + ground by adding a border of random dots at the top and bottom of the stereograms. The stereograms were now 54′ × 65′. The effect of this on the "figure only" condition was simply to slightly extend the background which was always at 0′ vertical and horizontal disparity. We therefore expected no difference between the results from experiments one and two for this condition. For the "figure + ground" condition, however, we expected a decrease in performance if monocular cues were important since they had been eliminated in this experiment. The two other differences between the two vertical misalignment conditions noted above were largely unchanged. The 0′ horizontal and vertical disparity borders were only 17% of the area of the stereogram, so a global registration mechanism would be expected to favor alignment of the rest of the stereogram. Note, however, that the addition of the border to the "figure + ground" condition makes it similar to the "figure only" condition in that global registration of the entire stereogram is no longer possible. Similarly, we expected little or no difference between the first experiment and this one if tolerance were due to a purely area-based mechanism.

Subjects

Three subjects were run, two with normal acuity and one myope who wore correcting spectacles. They were highly practised at the task. Two of them were naive about the expected experimental outcome and the third subject was one of the authors (K.N.).

Procedure

The random dot stereograms were generated on a LISP machine and displayed on a high-resolution, noninterlaced video monitor with P4 phosphor with a refresh rate of 60 Hz. Subjects sat in a darkened room and viewed the stereograms from 3 m through two tubes which defined 1.5° separate fields of view for each eye with approximately parallel visual axes. The simplicity of this setup and the ease with which fusion could be obtained using it outweighed the disadvantage of the slight accommodation/convergence mismatch which results from viewing a fixation point at 3 m with parallel visual axes. Each eye should converge by about 0.6-deg to avoid the mismatch. A uniform grey background with mean luminance 8.5 cd/m^2 (generated by turning on every other pixel) was on at all times except during the brief presentation of a stereogram. This served three purposes: (1) to prevent eye movements occurring after presentation of a stereogram from aiding registration by erasing any phosphor persistence, (2) to maintain a state of moderate light adaptation and (3) to prevent masking artifacts by avoiding large changes in the total light flux when the stereograms were flashed. The fixation target was a 10′ × 10′ cross centered in a 54′ × 65′ rectangle. It was turned off during the presentation of the stereograms and was turned on again after the subject had responded. The order of presentation of the stereograms was randomized and the rate was controlled by the subject. Stereograms were flashed for 117 msec to preclude voluntary vergence eye movements. The display flash duration was calibrated using a phototransistor and an oscilloscope to measure the time course. After each stereogram was flashed, the subject had to make two two-alternative forced choices, each of which was indicated to the computer by pushing one of two buttons. The subject first indicated whether the figure was in front of or behind the fixation plane and then indicated whether it was a square or a diamond. The first 16 trials were practice, during which incorrect responses were signalled by a tone. Thereafter, there was no feedback. Each session consisted of the presentation of 256 stereograms and lasted about 30 min. Each subject had six sessions in the first experiment, and four in the second.

The data were analyzed by computing the percent of trials on which the subjects correctly discriminated the sign of depth, the percent of trials on which they correctly discriminated the forms and the percent of trials on which they correctly discriminated both depth and form. These percentages were calculated as functions of horizontal disparity for each of the vertical disparity conditions. Note that 50% correct responses represents chance performance on this two-alternative forced-choice task. Results were pooled for the two "signs" of vertical disparity since the only purpose of these conditions was to allow randomization of the direction of vertical misalignment, thereby preventing registration eye movements (performance was very similar for the two conditions). Tests for the significance of differences between conditions were conducted using the one-tailed t-test for sample means.

Computer simulations

These results were compared with the performance of Grimson's (1981) implementation of the Marr–Poggio stereo algorithm. The binary arrays representing the stereograms were filtered with a Gaussian operator with $\sigma = 24''$ and then sampled at 30″ intervals to model optical blur and sampling by foveal cones. They were then filtered with 3 different operators corresponding to Laplacian of Gaussian ($\nabla^2 G$) channels with central excitatory regions with widths $w = 3.6'$, $7.2'$ and $14.4'$ (see Wilson and Bergen, 1979). The stereo algorithm was run on each channel independently with "eye position" fixed at zero disparity to simulate a flashed presentation. For comparison with the psychophysical results, the disparity assignments made by the program were grouped into three pools: near, fixation plane and far.

RESULTS

The results from the three subjects were very similar, so only averaged results will be presented (with one exception noted below). Under our experimental conditions performance on the form discrimination task ranged from about 40–70% correct with a very weak dependence on the experimental parameters. The relatively poor performance on this task probably reflects the small stereogram size as well as the greater difficulty of form discrimination compared with depth perception in briefly flashed random dot stereograms (see Harwerth and Rawlings, 1977). The performance of one subject on the form discrimination task was not improved by simplifying the experiment to a single forced-choice form discrimination task (i.e. by eliminating the depth discrimination task). This suggests that the poor form discrimination was not due to this judgment always following the depth judgment. Note that while the subjects were not able to discriminate the forms, they did see an amorphous figure on most trials. The results for both depth and form were very similar to the results for depth alone, but were shifted to a lower per cent correct level because of the poor form discrimination. We therefore only present results for the depth discrimination task.

Figure 1 shows the means for the collected data. Consider first the results of experiment one [Fig. 1(a) and (b)]. Note that between 1.8 and 5.4′ horizontal disparity, there is little dependence of performance on horizontal disparity for most of the vertical disparity conditions. Performance decreases at 10.8′ horizontal

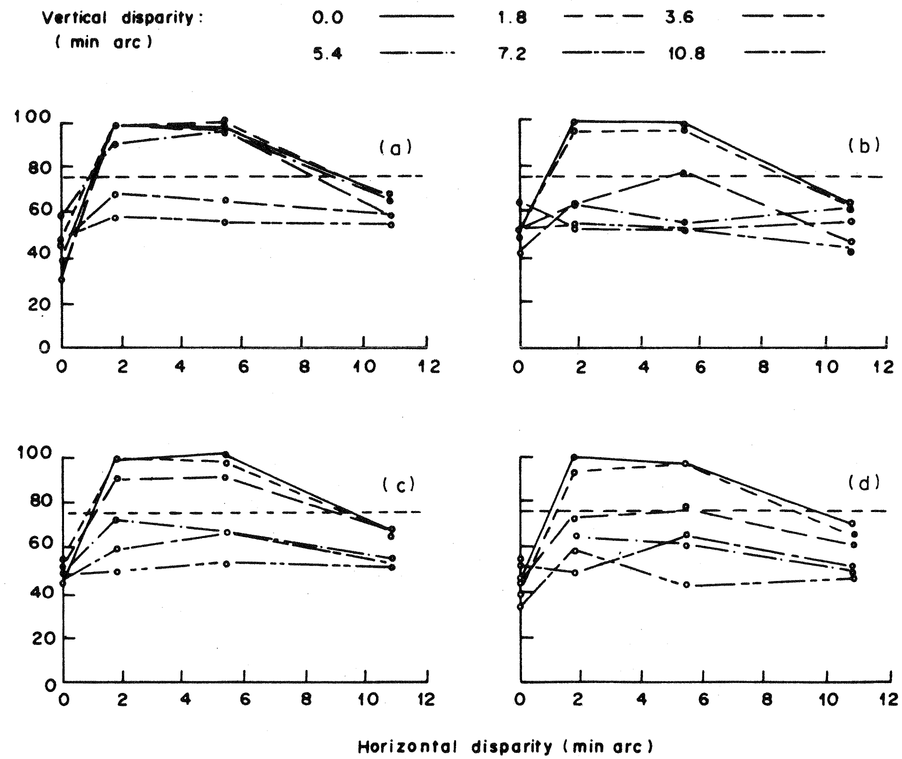

Fig. 1. Mean experimental results for the three subjects. (a) and (b) present data from the first experiment. In (a) vertical disparity was given to the figure and ground. In (b) vertical disparity was applied only to the figure. (c) and (d) show the corresponding results for the second experiment where monocular cues to vertical misalignment in the "figure + ground" condition were removed. The six curves on each graph correspond to the six vertical disparity conditions as indicated by the symbol key. For each vertical disparity, there are a total of 48 trials at 0′ and 96 trials at 1.8, 5.4 and 10.8′ horizontal disparity. The standard deviations range from 0 to 15%. The 75% threshold is indicated by a dashed line.

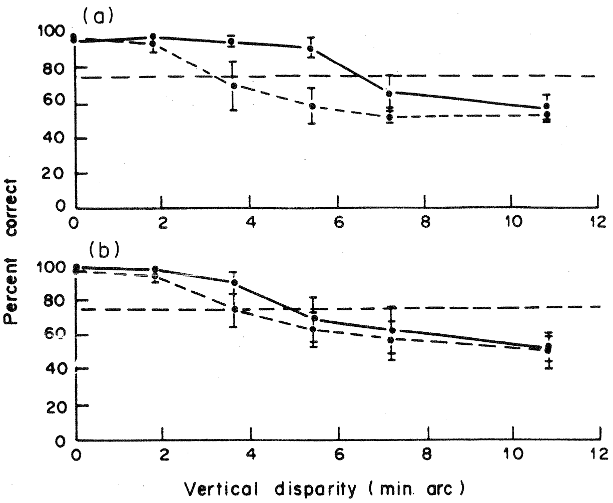

Fig. 2. Data from Fig. 1 replotted as function of vertical disparity. Data points are means and standard deviations of the collected results from the three subjects for 1.8 and 5.4′ horizontal disparity for each vertical disparity. (a) Shows the results from experiment one while (b) corresponds to experiment two where monocular cues were eliminated. The solid curves are for the condition where the figure and ground are vertically misaligned. The dashed curves are for the condition where only the figure is vertically displaced. The 75% threshold is indicated by a dashed line.

disparity, indicating that this value exceeds the fusion range, in agreement with published values for the horizontal extent of Panum's area. The 0 and 1.8′ vertical disparity conditions give very similar results for both the misalignment of the figure and ground [Fig. 1(a)] and the misalignment of figure only [Fig. 1(b)] cases, with nearly perfect performance. Note also that the 75% threshold is less than 1.8′ horizontal disparity, in contrast to the more than 2′ reported by Harwerth and Rawlings (1977). Data from additional experiments (not shown here) indicate that the threshold is less than 1′. There are two new features in the curves for the 3.6′ vertical disparity condition. First, when the figure and ground are misaligned there is no significant decrement in performance while misalignment of only the figure by the same amount results in a large drop in performance to below the 75% correct response threshold. Second, for the case of vertical disparity of the figure only there is a hint of a dependency of per cent correct responses on horizontal disparity so that performance improves as horizontal disparity is increased from 1.8 to 5.4′. Performance is only slightly better than chance when the vertical disparity of the figure alone is 5.4′ or more. When the vertical disparity is given to the figure and ground, then significant impairment only occurs for values greater than 7.2′ with performance falling to chance at 10.8′.

The second experiment was designed to give some indication of the mechanism underlying the difference between the two misalignment conditions. This experiment removed the monocular cue to vertical misalignment in the "figure + ground" condition by adding borders of random dots at the top and bottom of the stereograms to fill in blank space left by the vertical misalignment of the figure and ground. Figure 1(c) and (d) show the results. There are no significant differences between the first experiment and this experiment for the vertical disparity to "figure only" condition [Fig. 1(b) vs 1(d)]. For the vertical disparity to the "figure + ground" condition, the only significant difference (t-test, $P < 0.05$) is for 5.4′ vertical disparity [Fig. 1(a) vs 1(c)]. Performance is at about the 90% level when there are monocular cues to vertical misalignment but falls to about the 65% level when the monocular cues are removed. One of the subjects did not show the effect, responding at about 85% correct depth discrimination for 5.4′ vertical disparity in both experiments.

In order to highlight the effect of vertical disparity, the above results have been replotted in Fig. 2 with vertical disparity on the abscissa. Since the data have only a weak dependence on horizontal disparity between 1.8 and 5.4′, each data point represents the mean of the collected results for the three subjects for 1.8 and 5.4′ horizontal disparity with its standard deviation indicated by the bars. The results from the first experiment are shown in Fig. 2(a). When only the figure has vertical disparity, performance falls to 75% correct at about 3.5′ vertical disparity, while 6–7′ vertical misalignment is tolerated at the 75% level when the figure and ground are shifted. Figure 2(b) shows the results for the second experiment. There is no significant difference for the vertical disparity to "figure only" condition, but the maximum vertical disparity tolerated in the "figure + ground" condition is now less than 5.4′.

These results were compared with the performance of Grimson's (1981) computer implementation of the Marr–Poggio stereo algorithm for four vertical disparity conditions with 3.6′ uncrossed horizontal disparity. The disparity assignments made by the algorithm for the four vertical disparities are shown in Fig. 3. For ease of comparison with the two-alternative, forced-choice experimental data, the results are presented in three groupings of disparity assignments: fixation plane ±36″ ("FIX"), more than 36″ crossed disparity ("NEAR"), and more than 36″ uncrossed disparity ("FAR"). The channels are identified by the widths of the central excitatory regions (w's) of the $\nabla^2 G$ operators with which the stereograms were filtered. The values 14.4, 7.2 and 3.6′ were chosen as representative of a reasonable range for human foveal vision (the size of the channels is expected to increase with eccentricity, Wilson and Bergen, 1979; see also Marr and Poggio, 1979).

Figure 3(a)–(d) show the results for a square-shaped figure with 3.6′ uncrossed horizontal disparity and 0, 1.8, 3.6 and 5.4′ vertical disparity, respectively. The fixation plane is well-defined as can be seen in the columns labelled "FIX". The central "hole" in the fixation plane assignments of the two smaller channels ($w = 3.6'$ and $w = 7.2'$) also show that the stereogram has a central square-shaped figure not in the fixation plane. In the absence of vertical disparity [Fig. 3(a)] the disparity of the figure is clearly assigned to the "FAR" pool by all three channels. The algorithm makes a few false assignments to the "NEAR" pool around the discontinuity between the figure and ground. When the figure has 1.8′ vertical disparity [Fig. 3(b)] the large channel ($w = 14.4'$) correctly assigns the figure to the "FAR" pool. The two smaller channels are consistent with this assignment, although they probably would not be sufficient on their own for a correct response to be made. A correct forced choice may be made in the 3.6′ vertical disparity case by comparing the number of assignments to the "near" and "far" pools in the larger channels. It is unlikely that the algorithm can yield correct responses for any larger vertical disparities. [A more recent version of the matching algorithm (Grimson, 1984) that exploits figural continuity is not expected to perform significantly better since the sensitivity to vertical disparity mainly depends on the properties of the filters].

DISCUSSION

If the vertical disparities that could be tolerated by the matching process—in the presence of abundant

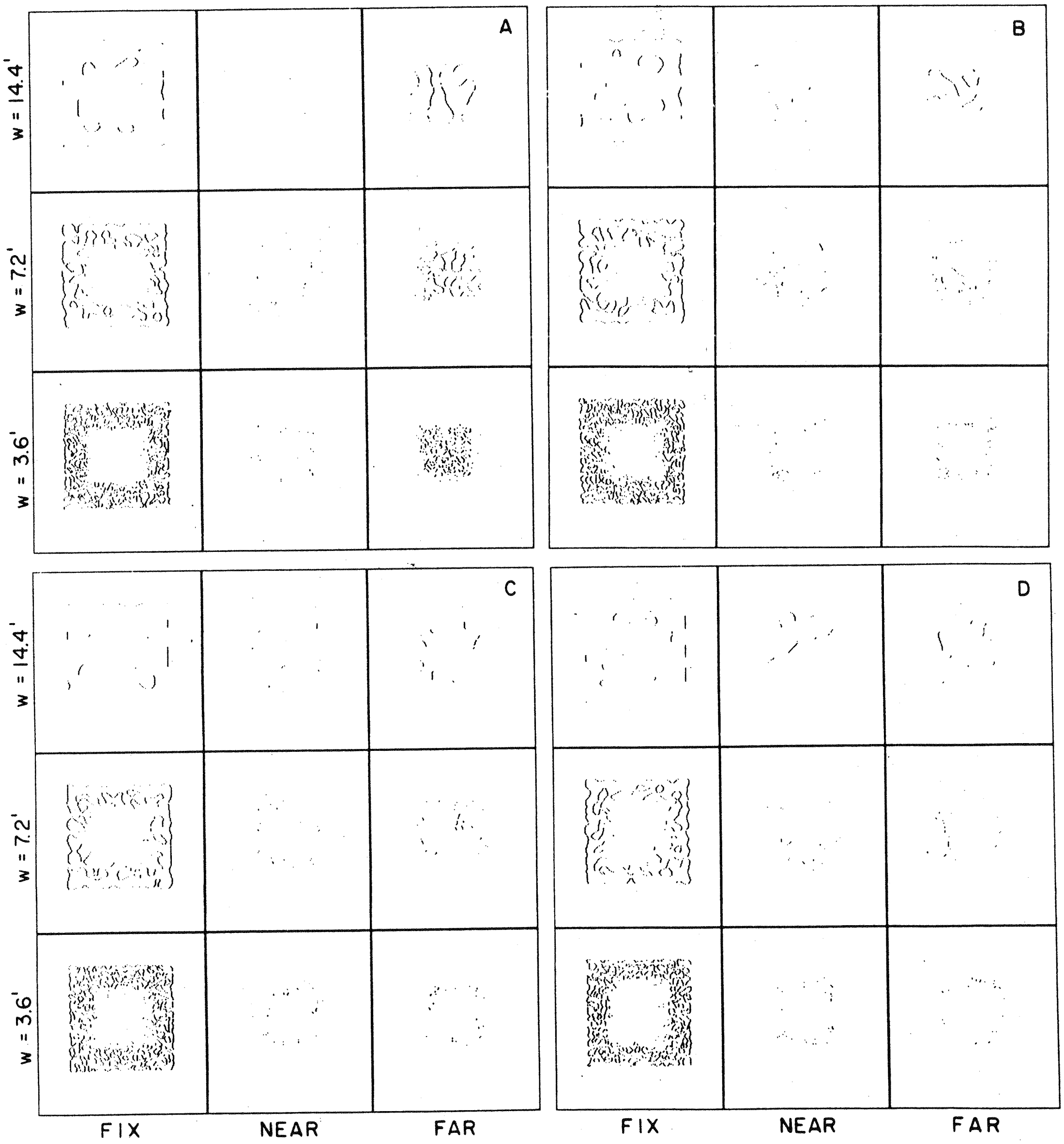

Fig. 3. Disparity pool assignments from Grimson's computer implementation of the Marr–Poggio stereo algorithm. The three rows correspond to the three operator sizes used, $w = 14.4'$, $w = 7.2'$ and $w = 3.6'$. The three columns show assignments made to approximately zero disparity (FIX), convergent disparity (NEAR) and divergent disparity (FAR). The horizontal disparity in all the stereograms is 3.6′ divergent. (a)–(d) show results for a square with vertical disparities 0, 1.8, 3.6 and 5.4′, respectively.

false matches—were large, one would be forced to conclude that human stereo matching were based on a 2-D search and significantly more complex than most existing theories assume. How large then are the vertical disparities that can be tolerated without eye movements? The answer given by our experiments is clear: under our experimental conditions, the maximum vertical disparity that allows a correct depth judgement (front vs behind) is small (3–7′). Human stereopsis seems therefore to use the epipolar constraint and to restrict matching to a roughly 1-D search.

The obvious next question is, how is this small tolerance to vertical disparity achieved? A simple possibility is the following. Recent theories of stereo matching assume that the image is first filtered with receptive fields (with center-surround organization) of several sizes (Marr and Poggio, 1979; Mayhew and Frisby, 1981) in order to obtain suitable primitive features to be used in the matching stage. These schemes have a small, intrinsic tolerance to vertical disparity because of the spatial averaging, or blurring, performed by the filtering stages. A comparison with a specific stereo matching scheme suggests that the results from our experiments with vertical misalignment of the "figure only" can be accounted for in this way. This conclusion is likely to hold true for several other similar stereo algorithms, since it follows more from the properties of the filters than from the matching process itself. We have in fact run an implementation of another matching scheme with the same filtering stages but which makes only near, far or approximately zero disparity measurements (Nishihara, 1983). The results were quite similar. Notice that this tolerance mechanism does not require monocular cues and is not restricted to global shifts of one image relative to the other. This proposal leads to the prediction that images that stimulate only channels with small w's will yield smaller values for the tolerance to vertical misalignment. Conversely, significant responses in larger channels would mediate a larger tolerance to vertical disparity.

The additional tolerance found for the "figure + ground" case cannot be explained in the same way. The difference is small—an additional 3′ depending on the criterion used—but significant. It is also functionally important since it brings the overall tolerance to about 7′—which is about what is required to compensate for the vertical fixation variability of the vergence system (which is around 7–9′, Fender and Julesz, 1967; see also Motter and Poggio, 1983). The only obvious difference between experiments one and two is the presence of monocular cues in the first case and their absence in the second. A simple hypothesis is that there is a mechanism compensating for global shifts of the image, possibly driven by monocular cues. The second experiment confirms that monocular cues have a significant role in the additional tolerance to vertical disparity [Fig. 2(a) vs 2(b)]. One of the three subjects, however, does not show the effect, indicating that there may be individual variability in the strategies used to compensate for binocular distortions.

The importance of monocular cues was also demonstrated by replacing the random dot stereograms with just the outlines of the figure and ground. One of the authors (K.N.) viewed these stereograms in both flashed and prolonged presentations. The observations are fully consistent with our interpretation. Vertical disparities of the "figure only" of up to 5.4′ were easily fused with no residual diplopia. The outline of the figure and of the ground could not be simultaneously fused, however, when the vertical misalignment of the figure was increased to 10.8′. The 7.2′ vertical disparity case was intermediate with brief periods of complete fusion alternating with small amounts of diplopia of the figure.

The properties of the correction mechanism would be very similar to eye movement properties and one wonders whether eye movements may indeed play a role. We believe that this is unlikely because of the short exposure time we used, since the reported latency of vergence movements of this small size is significantly slower (Schor and Ciuffreda, 1983). But this possibility cannot be completely excluded, since we did not record eye movements. The discovery that corrective vergence movements were possible within a time interval of less than 120 msec would be surprising. If we assume that this is not the case, we are left with the intriguing idea that a shift mechanism, independent of eye movements (possibly cortical), may underlie the residual 3′ tolerance to vertical disparity. Notice that 3′ disparity corresponds to about 6 foveal cones, a small but nonnegligible disparity. In the light of increasing evidence for a "focus of attention" in monocular visual information processing capable of moving across the visual field (Posner, 1980) a "shift" mechanism for vertical disparity may not be too far-fetched. Notice that, despite the term, no real shift of one image in the cortex or elsewhere needs to take place.

Fender and Julesz (1967) proposed a cortical-registration process to account for the large amounts of hysteresis they found for stereopsis under stabilized image conditions. This process was proposed to operate after matches had been made to maintain correspondence while the images were moved on the retinae. The registration mechanism we are suggesting operates before all matches have been assigned and uses monocular cues to correct for local image misalignments before a one-dimensional search establishes the remaining matches.

Our experiments do not provide any clue to what drives eye movements in normal conditions or to which vertical disparities and vertical disparity gradients can be eliminated by eye movements. The simplest hypothesis consistent with available data is that eye movements guided by monocular cues correct for vertical disparity. An observation made during our experiments also supports this point of view. In the case of vertical misalignment by more than 3.6′ of only the figure, where there are no monocular cues, prolonged viewing time does not significantly improve depth discrimination. The stereograms remain very difficult to fuse. However, in the case of vertical misalignment of the figure and ground, where there are monocular cues, fusion occurs effortlessly and under prolonged viewing conditions the subject is not aware of the vertical disparity. In the third condition, where the figure and ground are vertically misaligned but the monocular cues are eliminated by adding a random dot border, prolonged viewing gives rise to alternating percepts. The dominant percept is rivalrous, but occasionally the figure and ground are fused and can be seen clearly. These moments of fusion probably occur when the shifts in eye position which accompany binocular fixation align the stereograms by chance.

Several points should be kept in mind in interpreting our data and our conclusions. First, our experiments are strictly foveal since the overall area of the stereograms is slightly over 1° square. It is likely that larger stereograms may yield larger tolerances. We think, however, that the differences will not be dramatic and may be fully explained by the effect of eccentricity (cone spacing doubles at 4°) and by the improved signal to noise ratio in detecting the relative number of *near* vs *far* matches over a larger area. Second, it may be argued that our forced-choice test results in an over-estimate of the effective tolerance to vertical disparity. This is because vertical misalignments of the visual axes of the two eyes at the onset of the stimulus are expected to increase the range of vertical disparities that yield more than 50% correct responses beyond the actual fusional range. In the worst case, a subject may have a vertical fixation disparity which allows him to fuse half of the vertical disparity cases (say right up and left down) and, hence, perform at the 100% correct level. He would then perform at the 50% level by chance on the other half of the vertical disparity cases (right down, left up in this hypothetical example). This would result in an overall performance of 75% correct even though he would only have to be able to fuse half of the vertical disparity range nominally being tested. Our estimate of vertical disparity tolerance, however, refers to a criterion of at least 75% correct responses. Furthermore, examination of the data from our subjects did not reveal an asymmetry in performance with respect to the "sign" of the vertical disparity.

Third, form is barely recognized at all in our conditions, even without vertical disparity, possibly an indication that full stereo matching is not achieved. There may be several reasons for this, for example, the size of the stereogram may be too small and the exposure time too short. Additional experiments should clarify whether a process specialized for the detection of form needs a longer time than the 117 msec available in our experiments between onset of the stimulus and onset of the "masking" pattern. In any case, it is worthwhile to stress that our experimental conditions uncover only one aspect of stereopsis, responsible for relative depth discrimination.

Finally, additional experiments are necessary to characterize the possible role of vertical disparity. The vertical misalignments used in our experiments are rare under natural viewing conditions, except when produced by vertical fixation disparity. Therefore, it would be important to repeat our experiments at different fixation distances with the kind of vertical disparity and deformations that are associated with perspective projection. It may well be that the system easily corrects for them from extraocular information about fixation distance. Alternativeiy, the visual system may recover depth directly without the need of extraocular information, as proposed by Mayhew and Longuet-Higgins. In this case, vertical disparity should be not only tolerated but also precisely measured.

Acknowledgements—We thank E. Grimson and K. Nishihara for valuable discussions and for making available their computer implementations of stereo matching algorithms. E. Hildreth also made helpful comments on a draft of this paper. We are grateful to the subjects for their participation. We thank two anonymous reviewers for their careful reading and criticism of a version of this paper. K.R.K.N. was supported by NEI post-doctoral fellowship 5F32EY05584. Additional support came from a grant from the Whitaker Foundation to T.P.

REFERENCES

Baker H. H. and Binford T. O. (1981) Depth from edge- and intensity-based stereo. *Proc. 7th Int. Joint Conf. on A.I.*, Vancouver, BC, pp. 631–636.

Duwaer A. L. (1982) Nonmotor component of fusional response to vertical disparity: a second look using an afterimage method. *J. opt. Soc. Am.* **72,** 871–877.

Fender D. and Julesz B. (1967) Extension of Panum's fusional area in binocularly stabilized vision. *J. opt. Soc. Am.* **57,** 819–830.

Grimson W. E. L. (1981) A computer implementation of a theory of human stereo vision. *Phil. Trans. R. Soc. Lond. B* **292,** 217–253.

Grimson W. E. L. (1984) Computational experiments toward automatic stereo vision. M.I.T. A.I. Memo.

Harwerth R. S. and Rawlings S. C. (1977) Viewing time and stereoscopic threshold with random dot stereograms. *Am. J. Optom. Physiol. Opt.* **54,** 452–457.

Julesz B. (1971) *Foundations of Cyclopean perception.* The Univ. of Chicago Press, Chicago, IL.

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

Mayhew J. E. W. and Frisby J. P. (1981) Psychophysical and computational studies toward a theory of human stereopsis. *Artif. Intell.* **17,** 349–385.

Mayhew J. E. W. and Longuet-Higgins H. C. (1982) A computational model of binocular depth perception. *Nature* **297,** 376–378.

Motter B. C. and Poggio G. F. (1983) Binocular fixation in the Rhesus monkey: spatial and temporal characteristics. *Expl Brain Res.* In press.

Nishihara H. K. (1983) PRISM: A practical realtime imaging stereo matcher. *SPIE: Intelligent Robots: Third Int. Conf. on Robot Vision and Sensory Control*, Cambridge, MA.

Posner M. I. (1980) Orienting of attention. *Q. J. exp. Psychol.* **32,** 3–25.

Schor C. M. and Ciuffreda K. J. (1983) *Vergence Eye Movements: Basic and Clinical Aspects*, pp. 323–326. Butterworths, London.

Schor C. M. and Tyler C. W. (1981) Spatio-temporal properties of Panum's fusional area. *Vision Res.* **21,** 683–692.

Wilson H. R. and Bergen J. R. (1979) A four mechanism model for spatial vision. *Vision Res.* **19,** 19–32.