
BINOCULAR CORRESPONDENCE IN STEREOSCOPIC VISION

A. J. PARKER, J. M. HARRIS, B. G. CUMMING and J. H. SUMNALL
Oxford

For binocular stereoscopic vision to be possible, the visual nervous system needs to perform two tasks with the information available from the left and right eyes. First, features visible from the left eye must be paired up with the corresponding feature as seen from the right eye's vantage point. Second, the geometric information from the matched binocular features must be transformed into an estimate of binocular disparity. In this sense, the paper by Barlow, Blakemore and Pettigrew¹ represents the first complete proposal for the neural mechanisms underlying binocular correspondence and the perception of stereoscopic depth. They proposed that the feature selectivity of visually-responsive neurons should have a central role in sorting out which visual features in the left eye should match with those in the right eye and that the same group of neurons should provide signals from which binocular disparity could be extracted to signal depth. In short, for the cat's visual system, the neurons in cortical area V1 were proposed to be involved in both binocular matching and the recovery of stereoscopic depth.

Neurons in V1 have a number of obvious monocular feature selectivities that can be exploited for binocular matching, namely the local orientation, spatial frequency and spatial phase of regions of the image's luminance (black/white) contrast. The behavioural significance of these features is also confirmed by numerous psychophysical studies of the stereoscopic capabilities of human vision. Other feature selectivities in V1, such as colour, may have a role but this is not discussed in detail here. It is also

In this context the term 'binocular correspondence' refers to the process by which features present on one retina are matched to similar features on the retina of the other eye in order to deliver a perception of stereoscopic depth. An initial condition for this process is that the ocular axes are aligned (i.e. not strabismic). Thus, bringing the ocular axes into correspondence can be thought of as the first essential stage in the whole process of binocular correspondence.

Correspondence to: A. J. Parker, University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, UK.

important ultimately to consider how information from stereo disparities is integrated with information from other depth cues. This question is also set aside for the moment.

These monocular feature selectivities permit a number of different ways in which the signals from both eyes can be combined to form a binocularly-sensitive cortical cell, which is a candidate for acting as a detector of stereo disparity. The manner in which neurons combine monocular signals has been explored experimentally in an extensive series of papers by Freeman, Ohzawa and others.^{2,3} This work provides considerable insight into the way in which isolated features (presented simultaneously to left and right eyes) will interact within the receptive fields of disparity-sensitive neurons within the primary visual cortex of cats. It should be noted that this type of experiment has essentially explored how neurons behave when presented with features that are not in anatomical correspondence on the two retinæ. Part of the discrepancy in the present literature between psychophysical work in humans and neurophysiological work in the visual cortex of anaesthetised animals is that there are other aspects of binocular combination that need to be addressed.

Some of these issues have been exposed by studies with random-dot stereograms^{4,5} and by computational studies of stereo vision.^{6,7} The problem posed by the random stereogram is an extreme version of the problem faced by a computer vision system that is attempting to recover stereoscopic depth from a pair of images. In these cases, there is an ambiguity problem. If a reliable monocular feature can be extracted from one eye's image, there will be a multiplicity of features in the other eye's image that are candidates for matching to the chosen feature. This problem cannot be solved by techniques that are restricted to the local analysis of a single feature. Indeed, from the beginning Julesz⁵ emphasised that the solution of the random-dot stereogram by human

vision represented in some sense the need for 'global' processing.

The main approach in addressing this problem has been to investigate whether a simple rule applied consistently within the spatial neighbourhood of a monocular feature might lead to the system being able to fix upon the correct global solution without the creation of too many local mismatches. An effective type of rule is one that places a limit on the spatial gradient of disparity within a particular neighbourhood.⁷ It has been established mathematically that when a three-dimensional scene of solid objects is inspected there is an upper limit on the steepness of the disparity gradient that is present in binocular images of the scene.⁸ Accordingly, this geometric fact can be exploited in computational models of stereopsis to provide a criterion for the elimination of false matches. A limit based on the spatial gradient of disparity means inevitably that, within the context of these models, the relative disparity between nearby features in the visual image is of more importance for binocular correspondence than whether or not those features fall on anatomically corresponding (or non-corresponding) points.

There are a number of demonstrations of sensitivity to relative disparity that are available from human psychophysical experiments. These will be reviewed briefly. A striking example of sensitivity to relative disparity is demonstrated by an experiment of Westheimer.^{9,10} In this study, subjects were asked to fixate a binocular target and judge whether a second, briefly-flashed stereoscopic stimulus was presented in front of or behind the depth of the fixation target. Two conditions were explored. In the first, the binocular fixation target was continuously visible and human stereo thresholds were of the order of 10 seconds of arc (10"). In the second condition, the binocular fixation target was extinguished just before the second stereoscopic stimulus was flashed on. Thresholds for stereo depth judgements were then 6–10 times poorer, even though the whole experiment was arranged with brief stimulus presentations so that the deviation of the eyes from the fixation point would be the same, regardless of whether or not the fixation point was continuously visible. In terms of retinal stimulation, the absolute disparity (i.e. the anatomical correspondence) of the stereoscopic stimulus was notionally the same in the two conditions. However, removal of the opportunity for an estimate of relative disparity increased the thresholds considerably.

There are other demonstrations of the importance of relative disparity. Erkelens and Collewijn¹¹ examined sensitivity to motion-in-depth under conditions in which the whole visual field is moved in opposite horizontal directions in the left and right eyes. Although such motion could induce convergent

and divergent eye movements, no sensation of motion-in-depth was induced. By contrast, a strong sense of motion-in-depth was induced when limited regions of the visual field were moved in the left and right eyes in such a way as to simulate motion-in-depth. In another kind of study, Mitchison and Westheimer¹² examined the sensitivity of humans to stereoscopic depth in configurations where a collection of points formed a plane in depth. Sensitivity for deviations of a single point out of coplanarity is in general higher than sensitivity for changes in the tilt or slant of the plane as a whole. A further example of the possible role of relative disparity is the suggestion (made repeatedly from psychophysical studies^{13–15} and occasionally from neurophysiological results^{16,17}) that the visual system might be specifically sensitive to local differences in orientation and spatial frequency between the two eyes, a sensitivity which would give a direct way of estimating stereoscopic tilt and slant in three dimensions.

In summary, there are a number of results in the psychophysical and computational literature that point to the importance of interactions between the disparities of spatially neighbouring features. This theme will be developed through a description of recent work that has measured the statistical efficiency of stereoscopic vision in human observers. It will be argued that these studies can provide a model-free estimate of the quality of binocular correspondence in more complex stereoscopic configurations.

HOW GOOD IS STEREO VISION WITH RANDOM-DOT FIGURES?

When an observer obtains a stereoscopic percept from a random-dot stereo figure, just what is going on internally within the visual system of the observer? This is obviously the central question in understanding binocular vision. But, before this question is approached, it would be as well to determine what level of performance is being achieved when stereo information is abstracted from the stereogram.

One highly useful measure of performance can be achieved by considering the susceptibility of stereo depth judgements to the addition of noise in the disparity domain. This measure turns out to be fundamental in the sense that if known levels of variability are added to any parameter it is possible to calculate an absolute, statistical limit on the ability of any system (man-made, biological or purely theoretical) to estimate that parameter. There is a long tradition of exploiting this approach in the study of human spatial vision, notably the studies by Andrews¹⁸ and Barlow.¹⁹

Consider a stereo task in which the observer is presented with a random-dot stereogram that depicts

a vertical step-edge in depth.²⁰ The observer has a clear binocular fixation mark before the trial to assist convergence and indicate where the step-edge will be located. The observer must judge on each trial whether the edge steps from right to left or the converse. If no noise has been added to the disparities of the dots forming the stereogram, all the dots lie in depth exactly on the planes of the step-edge and human thresholds for indicating the polarity of the step-edge are close to those for conventional stereo acuity. If noise is added individually to each dot, in the form of an additional disparity selected at random from a Gaussian distribution, then performance declines as the noise increases.

This decline in performance would be expected, whatever system is analysing the stereo information, because as the noise increases the task becomes a harder statistical discrimination. A better measure of human performance is to consider how well humans perform relative to the absolute statistical limits. These absolute limits define an 'ideal observer' and human performance can be assessed as the efficiency of humans relative to the ideal observer. The ideal observer has an efficiency of 100%. Thus, a human efficiency of 5% means that humans notionally require 20 times the number of statistical samples (in this case, dots in the stereogram) to achieve the same level of performance as the ideal observer.

In practice, human efficiency for this task is surprisingly low, often in the range 3–5% and rarely above 20% depending upon the exact stimulus conditions.^{20,21} There are other anomalies of human performance. For example, as the total number of dots in the stereogram is increased, the performance of the ideal observer improves steadily because the number of statistical samples is increasing. The performance of human observers does not improve steadily and soon reaches a more-or-less fixed level, at a value that is consistent with exploiting the information from only a few dots on either side of the step-edge.²⁰

WHAT LIMITS HUMAN STEREO PERFORMANCE?

There are two possible classes of limitation on human performance relative to that of the ideal observer. The first is that humans may suffer an additional source of noise, internal to their own visual system, that weakens their performance. The second possibility is that humans fail to use some of the information in the target that is exploited by the ideal observer. For example, if humans were to analyse a small sub-region of the step-edge target rather than the whole target, their performance would be inevitably worse than the ideal. In fact, our experimental evidence and computational mod-

elling show that the main limitation on human performance is of the second kind, but the limitation is more subtle than simply ignoring some spatial regions of the step-edge target.

Clearly, before a dot in the random-dot pattern can contribute to the estimation of stereo depth, binocular correspondence between the left and right eyes must have been established for that dot. This is not a problem for the ideal observer, because it has been arranged that it should have access to all the binocular information from each dot in the pattern. It follows that a failure of binocular matching by human observers would inevitably contribute to the observed pattern of inefficiency. In practice, our data indicate that humans lose efficiency, not only by a failure to achieve completely accurate binocular matching, but also by a loss of information during the combination of disparity estimates from points that have been matched correctly.^{21,22}

Further specific information on the limitations of human binocular matching can be obtained by analysing what happens as the level of noise in the step-edge pattern is increased. For the ideal observer, although performance will decline as noise increases, this still represents an efficiency of 100% because the task is statistically harder. However, the efficiency, not just the performance, of human observers declines as the level of noise increases. This suggests that as the noise is increased, there is a corresponding fall in the proportion of dots that are successfully matched. Also, as noise levels increase in these random-dot figures, the depicted surface changes from being smooth to jagged. In a jagged surface, there are more binocular pairs of dots that have a steep spatial gradient of disparity between each other.

Hence, this aspect of human stereo psychophysics may reflect the importance of the computational principle of setting a limit on the gradient of disparities during the binocular matching process. The important property of the computational models is that binocular matching of points takes place not only on the basis of having consistent features between the right and left eyes, but also with regard to the final spatial distribution of disparities within the binocular scene. Thus the spatial context in which a candidate pair of points (from left and right eyes) is found affects whether the algorithm will assign the pair as a true binocular match. Only those pairs that are matched correctly would be available for statistical pooling during depth judgements within our noisy stereograms.²²

In summary, humans appear to have two types of limit on their efficiency with these stereoscopic stimuli. The results with random-dot stereograms to which disparity noise has been added indicate that the spatial gradient of disparity is one important

limitation on the efficiency of human observers. The other limitation probably reflects a failure in extracting a single estimate of disparity from a number of disparity samples (i.e. in extracting disparities from correctly matched dots).

FEATURE-BASED SOLUTIONS IN STEREO CORRESPONDENCE

Up to this point, two types of constraint on stereo matching have been considered. One is the feature-based constraint identified from neurophysiological experiments; the other is the geometric constraint on the relationships between spatially-nearby stereo matches. It would be of considerable interest to know whether these two types of constraint may interact within stereo matching. Recently, we have obtained evidence that information from tokens or features involved in binocular matching can be treated separately during the stage at which binocular correspondences are sorted out. Thus feature-based information can assist the assignment of binocular matches, provided that there is a rich range of visual information within the scene.

The logic of the experimental investigations is as follows. The measure of efficiency introduced earlier can be interpreted as though the human observer uses only a percentage of dots in the stereogram as a basis for his or her judgement. Let us generalise the notion of the random-dot stereogram into a random-element stereogram, where the elements are tokens or micro-patterns that have attributes such as contrast, orientation, direction of motion, colour or other features. Suppose that we measure human efficiency for random-dot stereograms with a certain number of elements (all of the same class) and a particular level of added disparity noise. This is the baseline level of performance. If we now present stereograms that have the same distribution of disparities and spatial arrangement of elements but have more than one kind of element or token, then any change in efficiency must be due to the fact that the binocular matching stage treats the disparity information arising from different tokens as independent. This must reflect a property of the human visual system, since the same distribution of disparities is delivered to the ideal observer introduced above, regardless of whichever token signals a particular disparity.

This creates an opportunity to test whether different tokens are treated independently by the human stereoscopic system. The word 'independent' is being used here in the sense of 'statistically independent', since the measure is statistical efficiency. An increase in the efficiency of human stereoscopic vision when the same information about disparity is distributed over two tokens rather than one implies that each token is being processed

to some degree independently by the human binocular system. An exact doubling of efficiency would be consistent with complete independence of processing of the two tokens. The functional interpretation of statistical independence is not only that the binocular system refuses to match inappropriate tokens to one another, but also that geometric constraints, such as the disparity gradient limit, are effective only within a single class of token and not between classes of tokens.

Experimentally, we have investigated three possible manipulations of token-type in recent experiments and have obtained one positive and two negative results. In the positive case, dots of opposite sign of contrast appear to be treated independently: human efficiency doubles in stereopatterns where half the dots are bright and half the dots are dark relative to the background.²³ By comparison, in one of the negative cases, manipulating the level of contrast for dots of the same polarity of contrast (i.e. all dark or all light, only variations in degree of darkness or lightness) does not lead to improvements in efficiency.²³ Finally, in the other negative case, manipulating direction of motion (half the dots moving leftwards and half rightwards) does not lead to improvements in efficiency.²⁴ Needless to say, the experimental search for other possibilities continues.

There are two general points about the relation of feature-based correspondence to geometric-based correspondence that can be made. First, with an exact (but rather naively formulated) link between neurophysiology and psychophysics, one might expect that each type of physiological feature-detector that has been found should correspond to a statistically independent mechanism that can be identified psychophysically. This is unlikely to be true because the physiological feature-detectors may well share sources of noise that make their responses statistically dependent. However, it is a principle that should guide further study of this type. Second, in natural scenes there is a rich variety of features that are spatially distributed over the scene. Since complete statistical independence for the processing of features leads to a doubling of efficiency, this points to a possible explanation of the low efficiencies found in random-dot stereograms. If the human stereoscopic system could extract five or six different types of feature independently (i.e. in parallel) from natural images, then the overall efficiency of human binocular vision under natural circumstances would be much higher, approaching the region of 50% found for some other pattern vision tasks.

CONCLUSIONS

The early neurophysiological studies of binocular neurons indicated the importance of feature extrac-

tion for establishing stereoscopic correspondence. Later psychophysical and computational studies have pointed to the importance of the geometric distribution of stereo disparities within the scene as an important constraint on binocular matching. The importance of geometric constraints has had relatively little influence on neurophysiological studies of stereoscopic vision, in which even recent work²⁵ has studied disparity almost exclusively in terms of anatomical correspondence. The correspondence problem in binocular vision is especially difficult in the case of a random-dot figure when there are many similar features that need to be considered before the correct solution can be obtained, but it is easier if the features are different so that they can be labelled and matched separately. Theoretically, there is a continuum of possibilities for binocular stereopsis ranging from the complete avoidance of the correspondence problem by the use of highly elaborate features (e.g. at the level of full object recognition) to complete elimination of the need for any feature extraction and the application of purely geometrical constraints. Human vision does not occupy either extreme of this continuum.

This work has been supported by the Wellcome Trust, MRC, EPSRC (SERC) and the Oxford McDonnell-Pew Centre for Cognitive Neuroscience.

Key words: Stereopsis, Binocular vision, Statistical efficiency.

REFERENCES

1. Barlow HB, Blakemore C, Pettigrew JD. The neural mechanism of binocular depth discrimination. *J Physiol (Lond)* 1967;193:327–42.
2. Freeman RD, Ohzawa I. On the neurophysiological organization of binocular vision. *Vision Res* 1990;30:1661–76.
3. Ohzawa I, DeAngelis G, Freeman RD. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science* 1990;249:1037–41.
4. Julesz B. Foundations of cyclopean perception. Chicago, Ill.: University of Chicago Press, 1971.
5. Julesz B. Binocular perception without familiarity cues. *Science* 1964;145:356–62.
6. Marr D, Poggio T. Co-operative computation of stereo disparity. *Science* 1976;194:283–7.
7. Pollard SB, Mayhew JEW, Frisby JP. PMF: a stereo correspondence algorithm using a disparity gradient limit. *Perception* 1985;14:449–70.
8. Trivedi HP, Lloyd SA. The role of disparity gradient in stereo vision. *Perception* 1985;14:685–90.
9. Westheimer G. The Ferrier Lecture. Seeing depth with two eyes: stereopsis. *Proc R Soc Lond B* 1994;257:205–14.
10. Westheimer G. Cooperative neural processes involved in stereoscopic acuity. *Exp Brain Res* 1979;36:585–97.
11. Erkelens CJ, Collewijn H. Motion perception during dichoptic viewing of moving random-dot stereograms. *Vision Res* 1985;25:583–8.
12. Mitchison G, Westheimer G. The perception of depth in simple figures. *Vision Res* 1985;24:1063–73.
13. Blakemore C. A new kind of stereoscopic vision. *Vision Res* 1970;10:1181–99.
14. Koenderink JJ, van Doorn AJ. Geometry of binocular vision and a model for stereopsis. *Biol Cybern* 1976;21:29–35.
15. Rogers BJ, Cagenello R. Disparity curvature and the perception of three-dimensional surfaces. *Nature* 1989;339:135–7.
16. Blakemore C, Fiorentini A, Maffei L. A second neural mechanism of binocular depth discrimination. *J Physiol (Lond)* 1972;226:725–49.
17. Nelson JI, Kato H, Bishop PO. Discrimination of orientation and position disparity by binocularly activated neurons in cat striate cortex. *J Neurophysiol* 1976;40:260–83.
18. Andrews DP. Perception of contour orientation in the central fovea. II. Spatial integration. *Vision Res* 1967;7:999–1013.
19. Barlow HB. The efficiency of detecting changes of density in random dot patterns. *Vision Res* 1978;18:637–50.
20. Harris JM, Parker AJ. Efficiency of stereopsis in random dot stereograms. *J Opt Soc Am* 1992;9:14–24.
21. Harris JM, Parker AJ. Objective evaluation of human and computational stereoscopic systems. *Vision Res* 1994;34:2773–85.
22. Harris JM, Parker AJ. Constraints on human stereo dot matching. *Vision Res* 1994;34:2761–72.
23. Harris JM, Parker AJ. Identification of independent mechanism for bright and dark information in binocular stereopsis. *Nature* 1995;374:808–11.
24. Sumnall JH, Cumming BG, Parker AJ. The contribution of motion information to stereo matching: a statistical efficiency approach. *Invest Ophthalmol Vis Sci (Suppl)* 1995;36:S667.
25. Poggio G. Cortical neural mechanisms of stereopsis studied with dynamic random-dot stereograms. *Cold Spring Harbor Symp Quant Biol* 1990;55:749–58.