MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 558 July 1980

SOME COMMENTS ON A RECENT THEORY OF STEREOPSIS

David C. Marr and Tomaso Poggio

Abstract. A number of developments have taken place since the formulation of Marr and Poggio's theory of human stereo vision. In particular, these concern the shape of the underlying receptive fields, the control of eye movements and the role of neuronal pools in the so-called pulling effect. These and other connected matters are briefly discussed.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Project Agency of the Department of Defense under Office of Naval Research contract N00014-75-C-0643 and in part by National Science Foundation Grant MCS77-07569.

1. Some Comments on a Recent Theory of Stereopsis

We recently proposed an algorithm for the matching process in human stereopsis (Marr and Poggio, 1977, 1979). A number of points have developed since the theory was formulated in 1977 and we feel that they are interesting enough to deserve being discussed explicitly. In this paper we raise these points and discuss the present stance of the theory with respect to psychophysical, physiological and computational data.

2. The Shape of the Underlying Receptive Fields

According to the theory each image is filtered by oriented second derivative (bar) masks of four different sizes and the zero-crossings are matched between the two filtered images. Under rather weak assumptions, it can be shown that the oriented masks can be replaced by circularly symmetric masks (Marr and Hildreth, 1979) and we feel that the physiological implementation is based on this scheme. Specifically we believe (Marr and Hildreth, 1979; Marr and Ullman, 1979) that the LGN center-surround cells are the filters, whereas cortical simple cells are oriented zero-crossing detectors and should not be thought of as oriented bar mask filters. Such a scheme has been implemented on the MIT Artificial Intelligence Laboratory computer (Grimson and Marr, 1979) and it leads to an efficient stereo matching program. Mayhew and Frisby (1978) have independently found psychophysical evidence that the underlying filters in stereopsis are not oriented.

The statistical analysis that we gave for the interval distribution of zero-crossings of oriented filters is not valid for center-surround receptive fields. This is important, because the results of this analysis enable one to make quantitative predictions about the extent of Panum's fusional area under various conditions.

The reason why one needs to modify the analysis is that the Fourier transforms of the oriented and non-oriented filters are different. That of the oriented filter consists roughly of two infinite vertical stripes, symmetrically placed either side of the w_y axis. The cross-section looks like two camel humps (see Marr and Poggio, 1979, Table 1). The Fourier transform of the center-surround filter, on the other hand, looks like a ring around the origin in the Fourier plane (w_x, w_y). In order to computer the interval distribution between zero-crossings, measured along horizontal scan lines in the image, one has to use the one-dimensional Fourier spectrum obtained by projecting the filter's two-dimensional transform onto the w_x -axis. The projections are clearly different for the two filters, and hence the results of the statistical analysis differ.

The results of changing to center-surround filters are as follows. If no restriction

is placed on the relative orientations of the zero-crossings that are matched between the two images, the results are numerically about the same as for oriented filters expressed in the usual units of $w_{1\text{-}D}$. This is the width of the central part of the receptive field as projected onto a line, and it is what Wilson measured in his experiments.

If one places restrictions on the relative orientations of the matched zero-crossings, the distance between compatible zero-crossings increases dramatically. For example, if one requires that their orientations must be within 30° before they can match, the probability of finding two candidate zero-crossings within $\pm w_{1\text{-}D}$ decreases from about 50% for the two previous cases to about 10% (Grimson, in preparation). Furthermore, Eric Grimson has found, in statistical measurements made on the results of running his program on 50% random dot stereograms, that the empirical findings match closely the theoretical predictions.

3. The Pools are only for Pulling

The theory is not definite on the neural implementation of the algorithm or the neural representation of disparity. Disparity could be represented either by many neurons each sharply tuned to a particular disparity or by two or three pools of more broadly ($\sim w$) tuned neurons. What the theory does require is the existence of three pools of the second kind for the purpose of disambiguation. According to the theory, in up to 50% of the cases possible matches could be ambiguous, and this ambiguity is resolved by "pulling," that is by consulting the sign (convergent, divergent, near zero) of neighboring successful matches.

4. Some Consequences of Pulling

The operation of pulling amounts to a kind of local averaging and, therefore, tends to reduce the rate at which the system can follow spatial changes in disparity (corrugations). Tyler (1978) for example, showed that the bandwidth of stereopsis is limited to 3-4 cycles/degree. This may help to determine the neighborhood size over which pulling takes place.

If there are cells that implement the pulling process they might for example have the following characteristics:

1) They may be sensitive to the disparity of a thin bar, but rather insensitive to its position in the receptive field.

2) The effect of introducing a second bar within the receptive field should depend on whether its disparity has the same sign (+, −, 0) as that of the first bar.

3) The disparities involved here are small, on the order of the optimal width of the bar. These cells are not to be identified with the "near" and "far" neurons described physiologically by Poggio and Fisher and other authors (Poggio and Fisher, 1977; von der Heydt et al., 1978) and which we discuss in the next section.

5. Control of Vergence Movements and Rough Sense of Depth

Our theory concentrated on stereo fusion and we neglected the fact that vergence movements may be controlled not only by the matching of zero-crossings, but also by more general and weaker estimates of disparity. Thus, our prediction 13, that for disparities outside the range of the largest channel, the control of additional vergence movements should exhibit the behavior of a random search, is not a necessary consequence of the theory.

There are various simple ways of estimating whether the visible surface lies convergent or divergent to the current fixation. These methods, however, would deliver only information about the sign of the disparity and not about the shape of the surface. One of these ways is simply to compare the number of possible convergent and divergent matchings. Such methods can be quantified. For example, using standard signal detection criteria, if the number of disparity detectors falls off inversely with disparity, then the criterion for detecting the existence of a convergent (divergent) surface at disparity d varies with the square root of the area of that surface. Such a relation has been suggested by B. Julesz and P. Burt, working with dynamic random-dot stereograms (personal communication). The important characteristics of such estimators are:

a) They can provide a rough sense of depth and hence drive eye movements.

b) They can provide no evidence about the precise shape of the region.

c) No depth discrimination between two convergent, or two divergent planes will be possible.

These points are essentially contained in our prediction 7 (p. 322, Marr and Poggio, 1979).

The neural substrate of such operations may perhaps be identified tentatively with the "near" and "far" neurons described by several physiologists. These have typical disparity ranges of a degree or more (in the macaque) which is much larger than the few minutes required for the pools we discussed earlier (see the diplopic detectors of Figure 7 and text, Marr and Poggio, 1979).

6. The Trigonometry of Stereopsis

The recovery of depth and surface orientation from disparity and its rate of change across the image is a matter of simple trigonometry. The formulae raise some points of interest, however, and so we give them here.

Figure 1(a) shows a top view and Figure 1(b) a side view of the geometry of the situation. It is straightforward to check that the rate of change of disparity ϕ measured vertically across the image is

$$\partial\phi/\partial\psi_v = -A\,l/[(B+l)^2+A^2]\ \cot\theta$$

where θ is the vertical component of surface orientation.

The horizontal situation is given by

$$\partial\phi/\partial\psi_h = [A^2+B(B+l)-A\,l\cot\theta]/[A^2+(B+l)^2]$$

Notice that this expression reduces to 1 when the surface coincides with the line of sight from the right eye.

The interesting feature of these formulae is that the right hand side essentially has a factor of $1/l$, A and B being small. In other words, perceived surface orientation for a fixed rate of change of disparity across the image should depend upon the current estimate of distance from the viewer to the fixation point. This observation is reflected in the perceptual fact that as one increases the viewing distance to a random-dot stereogram, the perceived surface orientation steepens.

7. Remarks about the Domain of the Matching Process

According to our theory the items that are matched between the two images are zero-crossings and terminations obtained from Wilson's four channels. This theory has

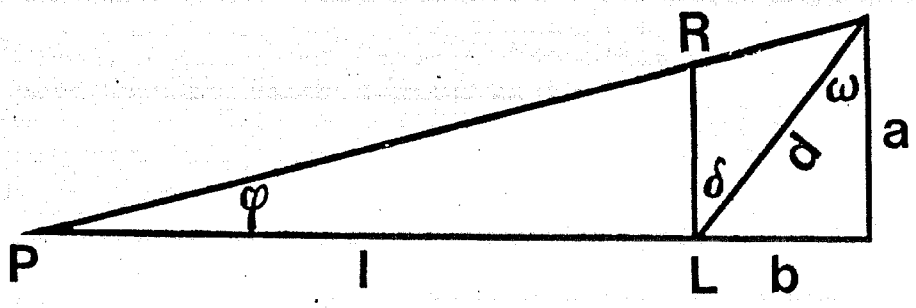
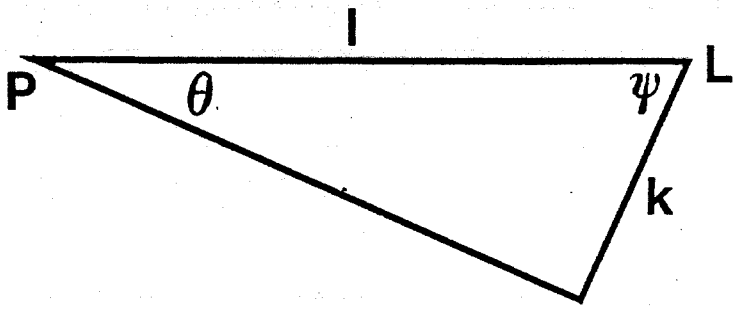


Figure 1. (a) A topview of the geometry of the two eyes looking at a point P a distance l from the left eye. The line of sight is not necessarily perpendicular to the line joining the two eyes L and R, and the difference is described by the angle ω as illustrated. The true inter-ocular distance for this line of sight is δ_t, and the effective inter-ocular distance for this line of sight is $\delta_t \cos \omega$. The angle between the lines of sight from the two eyes is ψ, and it is the differences in the values of ψ for different points P that are normally called disparities. The lengths $A = \delta_t \cos \omega$, and $B = \delta_t \sin \omega$, are useful geometrical quantities.



(b) A side view of the same situation. The point P is shown lying on a plane that slopes vertically, and its slope at P is described by the angle θ. Only the left eye L is shown in this diagram, and again the distance l refers to the distance from the left eye. In order to recover surface orientation, one needs to recover the angle θ.

now been implemented and actually works on natural images and on random-dot stereograms (Grimson and Marr, 1979). It is, however, not necessary that matching take place at this early stage. It could be that one first finds edges and groups and then matches these, which would amount to matching two primal sketches (Marr, 1976) rather than two sets of zero-crossing segments.

An argument against this is one of precision. When zero-crossings are first extracted their localization may be extremely precise and, hence, measurements of stereo disparity can easily be made very accurately. The longer one waits and the more complicated are the grouping operations that one carries out before matching, the more difficult it would be to maintain precision. While this is not an important consideration for rough estimates of disparity, it is probably crucial for the fusion process itself. We therefore feel that although the matching of higher order primitives is very likely to be used for guiding eye movements and obtaining rough sensations of depth, it would be surprising if the fusional process itself involved primitives substantially more complex than zero-crossings and terminations.

REFERENCES

Grimson, E. and Marr, D. (1979) "A Computer Implementation of a Theory of Human Stereo Vision," *Proceedings of ARPA Image Understanding Workshop,* S.R.I., pp. 41-45.

von der Heydt, R., Adorjani, C., Hanny, P. and Baumgartner, G. (1978) "Disparity Sensitivity and Receptive Field Incongruity of Units in the Cat Striate Cortex," *Exp. Brain Res.*, Vol. 31, pp. 523-545.

Kidd, A. L., Frisby, J. P., and Mayhew, J. E. W. (1979) "Texture Contours Can Facilitate Stereopsis by Initiating Vergence Eye Movements," *Nature,* Vol. 280, pp. 829-832.

Marr, D. (1976) "Early Processing of Visual Information," *Phil. Trans. R. Soc. Lond.,* Vol. 275, pp. 483-524.

Marr, D and Hildreth, E. (1979) "Theory of Edge Detection," *M.I.T. A.I. Memo No. 518.*

Marr, D. and Poggio, T. (1977) "Theory of Human Stereo Vision," *M.I.T. A.I. Memo No. 451.*

Marr, D. and Poggio, T. (1979) "A Computational Theory of Human Stereo Vision," *Proc. R. Soc. Lond.*, Vol. 204, pp. 301-328.

Marr, D. and Ullman, S. (1979) "Directional Selectivity and its use in Early Visual Processing," *M.I.T. A.I. Memo No. 524.*

Mayhew, J. E. W. and Frisby, J. P. (1978) "Stereopsis Masking in Humans is not Orientationally Tuned," *Perception,* Vol. -7, pp. 431-436.

Peichl, L. and Wassle, H. (1979) "Size, Scatter and Coverage of Ganglion Cell Receptive Field Centres in the Cat Retina," *J. Physiol.*, Vol. 291, pp. 117-141.

Poggio, G. F. and Fisher, B. (1977) "Binocular Interaction and Depth Sensitivity of Striate and Prestriate Cortical Neurons of the Behaving Rhesus Monkey," *J. Neurophysiol.*, Vol. 40, pp. 1392-1405.

Tyler, C. W. (1977) "Spatial Limitations of Human Stereoscopic Vision," *Three Dimensional Imaging, SPIE 120.*