THE ROLE OF EYE POSITION INFORMATION IN ALGORITHMS FOR STEREOSCOPIC MATCHING

K. Prazdny

Laboratory for Artificial Intelligence Research
Fairchild Instrument and Camera Corporation
4001 Miranda Avenue, MS 30-888
Palo Alto, California 94304

ABSTRACT.

The viewing system parameters probably have to be explicitly taken into consideration in formulating computational solutions to the stereoscopic matching problem in binocular vision. An algorithm for computing the directions of gaze from purely visual information without a prior solution to the correspondence problem is outlined. It would seem that the availability of such (or similar) algorithm is a precondition for any stereo matching process using the epipolar constraint.

1. Introduction.

The structure of a 3D scene can be uniquely determined from two projections taken at two spatially separated locations. This fact and the underlying geometry is the basis of binocular stereopsis. Conceptually, the problem has been divided into two sub-problems:

[1] one has to solve the correspondence problem which, when succesfully accomplished, enables one to

[2] compute the relative orientation of the two view directions (the orientations of the two principal rays, that is), and the structure of the viewed scene (depth relationships between various locations in the scene).

In this paper it is argued that an algorithm based on this particular serial decomposition of the problem is flawed: eye orientation parameters probably have to be computed before an attempt to solve the correspondence problem. The origin of the present analysis is the belief that the major goal of binocular stereopsis is to obtain absolute depth information (i.e. distances scaled by the interocular separation) in a near space (i.e. at distance up to about 1-2 metres) for the purposes of fine hand-eye coordination. This specification of the goals of binocular stereopsis, while certainly not novel leads, however, to reformulation of the problem. This is due to the fact that absolute depth cannot be computed without accurate information about viewing system's parameters. Because the eye position information from extraretinal sources is assumed either to be unavailable to the visual system, or to contain errors too large to be usefull (the experimental evidence seems to show that only very rough information about eye position and motion is available [see e.g. Bridgeman, 1980]), one should look for a way to compute the directions of gaze from purely visual information.

The problem of computing the "camera parameters" is conceptually not very difficult and has been investigated in the past (e.g. Panton, 1977; Gennery, 1979; Longuet-Higgins, 1981). The problem with this classical approach is the requirement that the correspondence problem be somehow solved for some minimal subset of image elements. The eye orientation parameters are then computable with a precision that depends mostly only on the accuracy with which the position of the image "points" can be specified.

2. The epipolar constraint.

A large amount of literature on stereo vision has been devoted to the problem of establishing a "point-to-point" correspondence between the left and right images to extract the disparity information (e.g. Dev, 1975; Julesz, 1963, 1971; Nelson, 1975; Marr and Poggio, 1976, 1979; Mayhew and Frisby, 1981). The usual approach is exemplified by Marr's statement that "because the two eyes lie horizontally, we need consider only all the possible matches along horizontal lines; therefore, we can reduce the problem to the simple one-dimensional case..." (Marr, 1982, p.116). This statement is unfortunately true only in a special case when the directions of gaze are (nearly) parallel, i.e. when the eyes fixate a distant object. In general, the epipolar lines refered to above are neither horizontal nor parallel (see Figure 1). The epipolars refer, in this context, to a set of pairs of straight lines such that all points on a straight line in one image have matching points along a straight line in the other image. A planar projection surface is assumed here. In the case of a spherical retina, a straight line would correspond to a great circle. The idea of an epipolar line pair is a powerful computational constraint which also seems to have a strong physiological support. For example, in a recent paper, Kroll and van de Grind (1980, p.667) succesfully argue that the interactions between binocular neurons are limited to neurons with the same elevation specificity. While elevation here means presumably a horizontal scanline another, more plausible, interpretation is possible: elevation could be interpreted to mean the appropriate epipolar line. The orientation of epipolars in the image plane is a function of the orientation of the directions of gaze (see section 4 for detail).

In automatic photo interpretation or machine vision, the problems are usually approached without explicit concerns about biological constraints. The stereoscopic matching is conceptualized as a genuine two-dimensional problem (see e.g. Barnard and Thompson, 1980). In many cases, the camera orientaions are available a priori. If not, a common approach is to apply an "interest operator" which extracts a set of (monocularly) distinguishable (ususally in terms of grey level distribution) features in each image. The correspondence problem may be solvable for this set which in turn makes the "camera calibration" (computing the viewing system's parameters) possible. Unfortunately, this approach breaks down in situations where it is not possible to solve the correspondence problem for a subset of image "tokens" based on the uniqueness of certain features in the monocular input. An example are random dot stereograms. The ability of the (human) visual system to establish correspondence in this case implies the existence of a mechanism based on different considerations.

To conclude, although it is well known that horizontal disparities are by themselves uninterpretable, what is less commonly realized is that they cannot be even measured unless some information about eye position is explicitly used to guide the stereo-matching process. Previous conceptualizations of the task have (usually) seen the computational process as decomposeable into two serial modules: first extract the disparities, and then interpret them. This cannot work because in order to solve the correspondence problem utilizing the epipolar constraint one needs accurate information about the direction of gaze. This information cannot be obtained from extraretinal sources and has to be computed from visual information. But in order to do this one is supposed to have solved the correspondence problem for some minimal subset of image feature points. This is the famous "chicken and egg" problem in another disguise. It is interesting to note that the problems of this class arise frequently in the investigation of visual perception. For example, the computation of "shape-from-shading" and obtaining the direction of illumination seem to be similarly interrelated.

3. Possible solutions.

Three solutions to the problem identified above present themselves immediately.

[1] One can eliminate the epipolar constraint completely and conceptualize the stereoscopic matching in a way similar to motion computation in monocular motion parallax situation, i.e. as a genuine two-dimensional matching problem (see e.g. Barnard and Thompson, 1980).

[2] One may try to develop a scheme in which it is possible to make the stereoscopic matching and the computation of directions of gaze run concurrently as an intrinsically cooperative process. One way to do this may be to make the matching process a function of the eye orientation.

[3] The third possibility is to compute the eye orientation angles prior, and independently of, the subsequent matching process which may then use the epipolar constraint.

In section 5 it will be shown that it is, theoretically at least, possible that an algorithm of this latter class, directed by the global requirement that the directions of the gaze of the two eyes must be the same which ever points are used to compute them (this is a very strong global constraint holding over the whole image) could be used to accomplish this. The computational complexity of the proposed algorithm can be enormous if the minimal subset of feature point pairs (potential matches) needed to compute candidate directions of gaze is large. In the next section it will be shown that when some simple assumptions are made, this subset can be as small as 2 or 3 feature points. It will be seen that under these circumstances, the whole process can usefully be conceptualized as maintaining a 1-dimensional histogram and locating its highest peak.

4. Computing the directions of gaze.

Refer to Figure 2. It is assumed that the eye moves in such a fashion that the results of its motions can be described by a rotation utilizing only two of the possible three degrees of freedom (Donder's law). In our case, the description consists of rotations about an axis perpendicular to a common plane which itself is restricted to rotations about the interocular axis (vector $\vec{T}$ in Figure 2).

An image point P_i and the interocular distance $O_l\ O_r$ define a plane which cuts each image along an epipolar line. In the case of a planar projection surface, these two lines are straight lines. All epipolars in an image meet at one common point. The locus of zero "vertical" disparities coincides (in our case) with the image x-axis. Refering to Figure 2 it is easy to see that

$\vec{T}=(\sin\beta,0,\cos\beta)$
$\vec{z}_r = (-\sin\alpha,0,\cos\alpha)$
$\vec{z}_l = (0,0,1)$
$\vec{x}_r = (\cos\alpha,0,\sin\alpha)$
$\vec{x}_l = (1,0,0)$
$\vec{y}_r = \vec{y}_l = (0,1,0)$

(the focal distances are assumed to be 1). After some algebra it is seen that the x-coordinates of the epipolar intersections are specified by $(\tan\beta,0)$ and $(\tan(\alpha+\beta),0)$ in the left and right image coordinate frames respectively (the y-coordinates are zero).

Within this geometry, the eye orientation parameters (the angles α and β) are computable from the correspondences defined at three (the fovea and two arbitrarily choosen) image points, and solving a single fourth-degree equation. To see this, consider a (left) image point p'_i, with a corresponding point in the right image p''_i. The transformation from the right to the left image coordinate system is define simply as

$$P_i'\vec{P}_i'=(P_i''\vec{P}_i''+\mathbf{T})\mathbf{A} \qquad (1)$$

This says that the vector $\mathbf{P}_i'=\mathrm{P}_i'\vec{\mathrm{P}}_i'$ (P_i' is the magnitude and $\vec{P}_i'$ is the unit direction vector) defined in the left coordinate system, is abtained from its counterpart, defined in the right coordinate system, by translating and rotating it respectively. $\mathbf{T}=\mathrm{I}\vec{T}$ where I is the interocular separation. The rotation $\mathbf{A}$ is determined by the matrix

$$\mathbf{A}=\begin{bmatrix}\cos(\alpha) & 0 & \sin(\alpha)\\ 0 & 1 & 0\\ -\sin(\alpha) & 0 & \cos(\alpha)\end{bmatrix}$$

Equation (1) can also be written as

$$P_i'\vec{P}_i'\mathbf{A}^{-1}=(\mathrm{P}_i''\vec{\mathrm{P}}_i''+\vec{\mathrm{T}}) \qquad (2a)$$

Multiplying equation (2a) by the unit vector $\vec{P}_i''$ (vector product is denoted by a "x"), and scalar multiplying (denoted by a ".") the result by $\vec{T}$ produces

$$P_i'(\vec{P}_i'\mathbf{A}^{-1}\ \mathrm{x}\ \vec{\mathrm{P}}_i'')\ .\ \vec{\mathrm{T}}=0 \qquad (2b)$$

This means that $(\vec{P}_i'\mathbf{A}^{-1}\ \mathrm{x}\ \vec{\mathrm{P}}_i'')$ is perpendicular to $\vec{T}$. Consequently, the vector product $\mathbf{n}_{ij}=(\vec{\mathrm{P}}_i'\mathbf{A}^{-1}\ \mathrm{x}\ \vec{\mathrm{P}}_i'')\ \mathrm{x}\ (\vec{\mathrm{P}}_j'\mathbf{A}^{-1}\ \mathrm{x}\ \vec{\mathrm{P}}_j'')$ points in the direction $\vec{T}$. The same must hold for any pair of points, i.e. $\mathbf{n}_{ik}=(\vec{\mathrm{P}}_i'\mathbf{A}^{-1}\ \mathrm{x}\ \vec{\mathrm{P}}_i'')\ \mathrm{x}\ (\vec{\mathrm{P}}_k'\mathbf{A}^{-1}\ \mathrm{x}\ \vec{\mathrm{P}}_k'')$. Thus

$$\mathbf{n}_{ij}\ \mathrm{x}\ \mathbf{n}_{ik}=0 \qquad (3)$$

It is advantageous to chose point p_i to be the fixation point, i.e. $\vec{P}_i'=(0,0,1)$ and $\vec{P}_i''=(0,0,1)$. [It is assumed that the principal rays through the two centres of fovea do indeed intersect.] Expanding (3), and after a good deal of algebra, we obtain

$$ax^4+bx^3+cx^2+dx+e=0 \qquad (4)$$

where $x=sin(\alpha)$, and a, b, c, d, and e are some constant expressions. This is because $\vec{P}_i'\mathbf{A}^{-1}=(\sin(\alpha),0,\cos(\alpha))$ so that $\vec{P}_i'\mathbf{A}^{-1}\ \mathrm{x}\ \vec{\mathrm{P}}_i'')=(0,\sin(\alpha),0)$. Equation (4) has a closed form solution and four distinct roots, in general. An important point to note is the fact that equation (3) does not involve radial distances, and that it holds for **all** points on the epipolar line (i.e. if equation (4) applied to points j and k results in a value x, the same value x is obtained using points j and l provided k and l lie on an epipolar line in the right image).

5. An algorithm.

To compute the eye orientation angles (and thus the slope of the epipolar lines on the image surface) one may use the enormous amount of redundancy afforded by the global constraint that the directions of gaze have to be the same whatever set of points is used to compute them. The following simple algorithm accomplishes this.

[1] For each pair of points p'_i, p'_j in the left image, consider all possible matches (within some neighbourhood) in the right image, and for each such possibility, compute the value $sin(\alpha)$ as outlined above [equations (3) and (4)].

[2] Make a histogram of those values of $sin(\alpha)$ lying within the allowable range (e.g. the principal directions cannot point behind the head, nor even too much sideways). If a

particular value of $sin(\alpha)$ occurs n-times, the histogram value will be n. In practice, because of errors of various nature, one would update the frequencies within some neighbourhood of a particular value of $sin(\alpha)$ (according to some, perhaps gaussian, weighting function).

[3] Chose the value under the highest peak of the histogram as the true value of $sin(\alpha)$, compute the angle β (from 2b), and from this the orientation of the epipolar lines.

It is clear that because of the more or less random nature of the competing "incorrect" matches (i.e. matches not lying on the correct epipolar lines), the histogram will indeed peak above the correct value of $sin(\alpha)$.

6. Discussion.

The method, in its present form, has several shortcomings. One is its "pointilistic" way of treating the images. The assumption that an image can usefully be described by point descriptors is valid in the case of a random dot stereogram but may not be very useful in the general case.

The second problem is exemplified by the disparity field in Figure 1. While the disparities associated with elements of the frontal surface are within the fusable range specified by the Panum's area, the disparities associated with the background surface are clearly too large (this is why the vergence eye movements are essential to fusion). Consequently, if the image as a whole writes into the same histogram, and there are regions with disparities larger than the testing neighbourhood used in step [1], one may encouter problems. One solution may be to keep local histograms and test for agreement. This in turn suggests an intriguing possibility of segmenting the image on the basis of binocular combination before any depth perception. [This problem may turn out to be a non-problem in view of some evidence suggesting that binocular interactions can occur between points separated by as much as 7 degrees of visual arc (see Marr, 1982, p.154).]

The proposed algorithm is local and parallel in nature: the computation of α, and the histogram updating can easily be done at each image locus at the same time and independently of computations at other retinal loci. It should also be observed that nothing has been said about the classic ambiguity problem; it is orthogonal, in a sense, to the question of the orientation of the epipolar lines on the projection surface. False matches along an epipolar cannot be resolved without employing some other considerations (see e.g. Marr and Poggio, 1979). Once this has been done, however, the computation of absolute depth is, conceptually at least, trivial.

7. Conclusion.

The discussion in the beginning of this paper is related to the concept of the "plasticity of retinal correspondence" (Nelson, 1977): two loci on the two retinae do not signal a fixed disparity, the disparity signalled is a function of eye position (Richards, 1971). This modulation cannot, however, be a simple function of eye position obtained from extra-retinal sources: we argued that a precise information not obtainable from these sources is essential. Such conclusion is a direct consequence of formulating the goal of binocular stereopsis as obtaining an **absolute** depth information about the near space and the available evidence that the epipolar constraint is used in stereoscopic matching.

A simple algorithm for computing the eye orientation parameters was outlined to suggest a way in which a visual system could obtain accurate information about the orientation of the epipolar lines on the projection surface from purely visual sources before it attempts to solve the correspondence problem. The availability of such an algorithm is a precondition for any stereo matching algorithm using the epipolar constraint.

Acknowledgement.

I wish to thank Prof. John Mayhew and Dr. J.M.Tenenbaum for their comments and discussion.

REFERENCES

Barnard S.T. Thompson W.B. 1980
Disparity analysis of images
IEEE Trans.Pat.An.Mach.Intell. 4, 333-340

Bridgeman B. 1980
Direct perception and a call for primary perception
The Brain and Behavioral Sciences, 3, 382-383

Dev P. 1975
Perception of depth surfaces in random-dot stereograms
Int.J.Man Mach.Studies 7, 511-528

Gennery D.B. 1979
Object detection and measurement using stereo vision
Proceedings DARPA Image Understanding Workshop 101-107
Science Applications Inc.
Arlington, Virginia

Julesz B. 1963
Towards the automation of binocular depth perception
Proceedings IFIPS Congr., Munich 1962
ed.C.M.Popplewel
Amsterdam: North Holland

Julesz B. 1971
Foundations of Cyclopean Perception
Chicago: The University of Chicago Press

Kroll J.D. Grind W.A. van de 1980
The double-nail illusion
Perception 9, 651-669

Longuet-Higgins H.C. 1981
A computer algorithm for reconstructing a scene
from two projections
Nature 293, 133-135

Marr D. 1982
Vision
San Francsico: Freeman

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

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

Mayhew J.E.W. Frisby J.P. 1981
Psychophysical and computational studies towards
a theory of human stereopsis
Artificial Intelligence 17, 349-385

Nelson J.I. 1977
The plasticity of correspondence
J.Theor.Biol. 66, 203-266

Panton D.J. 1977
A flexible approach to digital stereo mapping
Photogrammetric Engineering
and Remote Sensing 44, 1499-1512

Richards W. 1971
Anomalous stereoscopic depth perception
J.Opt.Soc.Am. 61, 410-414

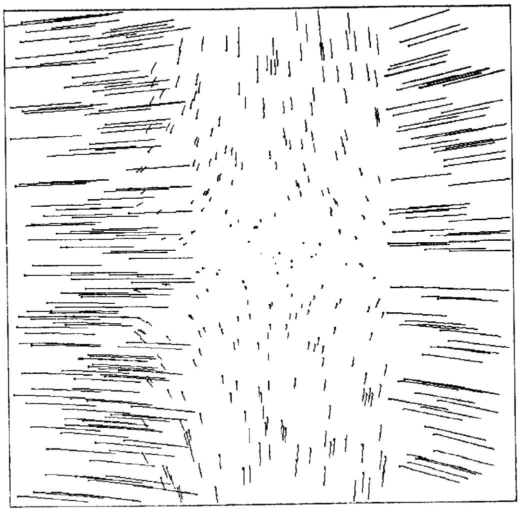

Figure 1.

A disparity map produced by viewing two slanted planes in depth. The observer fixates the front surface located 6 interocular distances away (approximately the reading distance). The distance of the background surface is 8 interocular separations. The field of view is 20 degrees of visual arc. The image is 1000x1000 pixels. Vertical disparity ranges from about -50 to +50 pixels.

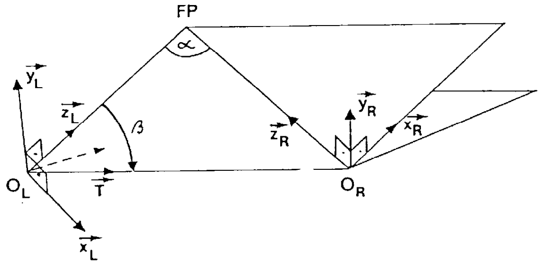

Figure 2.

Binocular geometry used in section 4. The two coordinate frames centered at O_l and O_r (the nodal points of the left and right eye, respectively) are fixed with respect to the retina and are constrained to move such that $\vec{y}_l = \vec{y}_r$, and $\vec{x}_r$, $\vec{z}_r$, $\vec{x}_l$, $\vec{z}_l$ lie in the same plane (i.e. obeying the Donder's law). $\vec{z}_l$ and $\vec{z}_r$ are the principal rays (fixation directions). FP is the fixation point. The (mutual) orientation of the two eyes (image planes here) is determined by the angles α and β.