

Computing convergence angle from random dot stereograms.

K.Prazdny

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

ABSTRACT.

Camera calibration is difficult in random dot stereograms lacking monocularly distinguishable features. An algorithm for computing the convergence angle from purely visual information without a prior solution to the correspondence problem was developed and tested. It is an experimental validation of some ideas first outlined in a previous report.

1. Introduction.

Marr and Poggio's (1979) theory of binocular stereopsis is one of the most visible examples of the success of the so-called computational approach to (visual) perception (Marr, 1982). It is also one of the most widely known modern theories of stereoscopic vision. Its cornerstone is a mathematical analysis of distribution of projected features (oriented edge segments) along a line. This results in a (rather simple, but effective) feature matching schema which makes use of another geometrical convenience: the epipolar constraint. The epipolar constraint explicitly identifies conjugated lines in the two images: features lying on such a line in one image must have their matching counterparts on the appropriate conjugated line in the other image (e.g. Duda and Hart, 1973). The theory (like all projection field theories) reduces the problem of binocular vision to solving the correspondence problem. In this paper it will be taken as an example of an edge-based, epipolar-line oriented approach to stereoscopic matching.

The criticism of the theory (e.g. Julesz and Schumer, 1981; Baker and Blicher, 1982; Mayhew and Frisby, 1981) concentrated mainly on some aspects of the epipolar matching strategy or on the role of eye movements. One of the major but rarely discussed problems with any epipolar-based matching scheme is the requirement that the orientation of the epipolar lines on the image surface be known with high accuracy before matching (i.e. the images must be rectified). While this requirement does not pose a problem when the convergence angle of the two eyes is small (the epipolars are nearly horizontal, in this case), the problem becomes urgent when viewing nearby objects, and must definitely be taken seriously when viewing objects in the near space, e.g. reading a book, or trying to get a thread through the eye of a needle. As the fixation point comes nearer to the observer, the orientation as well as the position of the epipolar lines on the image surface change rapidly (Prazdny, 1982). A common approach (at least in computer vision or automatic photo interpretation) to solve this problem of camera calibration relies on uniqueness of certain monocular features. The correspondence problem is usually solvable for a small set of such monocularly unique and distinguishable image elements which in turn makes camera calibration (and thus image rectification) possible (e.g. Moravec, 1980; Gennery, 1979; Longuet-Higgins, 1981).

Random dot stereograms (RDS) do not, in general, contain such unique features so that it is not possible to perform camera calibration using the above mentioned methods. A related issue

is that the horizontal disparities which are commonly thought to be the carriers of stereoscopic depth perception (e.g. Kauffman, 1974) are by themselves uninterpretable without information about eye position. In the case of RDS considered here, the disparities cannot even be measured without some prior camera calibration. The absence of monocularly distinguishable features, and the associated difficulty of solving the correspondence problem for a subset of features independently of the rest of the image (in RDS, it is not, in general, possible to solve the correspondence problem for a subset of image points independently of the rest of the image) requires that camera calibration be performed without explicitly solving the correspondence problem first.

This paper is concerned with this issue of how to obtain the convergence angle of the two cameras from purely visual information without an explicit solution to the correspondence problem. We report on an implementation of some ideas first outlined in Prazdny (1982) and show that a slightly more restricted camera model (only one camera is free to move, the other is held fixed) simplifies the computation of the convergence angle, and that, in general, the problem reduces to determining the intersection of sets of values computed within a few (typically 5) image neighborhoods.

2. The algorithm.

Computation of the convergence angle is possible because of the existence of a strong global constraint: the convergence angle has to be the same independently of which image points are used to compute it (Prazdny, 1982). The stereo camera model used in the program is illustrated in Figure 1. We assume that the interocular distance (d), and the radius of rotation (c) are known (from an initial calibration which has to be performed only once in the lifetime of the stereo set-up). Furthermore, it is assumed that the focal distances of the two cameras are fixed and the same (fixed focus lenses). Within this geometry, the knowledge of a correct correspondence for one image point enables one to compute the convergence angle α . To see this, choose a visual direction corresponding to an image point (p_l) of the fixed camera (left camera in our case), and hypothesize a matching point (p_r) on the right image plane. This point corresponds to a visual direction which has to "intersect" (in some e.g. least square sense) with the left visual direction. The only way to achieve this is by rotating the right image plane. The rotation of the visual direction corresponding to p_r generates a ruled surface. The problem of computing the angle α reduces thus to the problem of computing the intersection of the ruled surface (hyperboloid of revolution) parameterized by the angle α and some length parameter (p) with the left visual direction (see Figure 2 for a diagram and associated mathematics). Given a left image point p_l , for each hypothesized corresponding point p_r , we obtain thus a value of α . While it is possible to obtain a closed form solution to the above equations (the surface is a quadric), we opted for a numerical approximation because of the possibility of ill-conditioned equations (Faux and Pratt, 1979). The intersection of a line with a ruled surface

are two distinct points, in general. The degenerate case of non-intersection is detectable by the high value of the sum of squares function returned by the algorithm.

The actual minimization of the sum of squares was done using the variable metric descent to a minimum as suggested by Nash (1979, pp.159-160). This method was found to be very fast, reliable, and easy to code.

The algorithm proceeds as follows:

- [1]. Chose (at random) a (non-meridian) point on the left image (p_l) and establish some fixed neighbourhood N around p_l .
- [2]. Choose a point in the right image (p_r) lying within N and compute the angle α as outlined above.
- [3]. Repeat [2] for all right image points p_r within N . Save the results (i.e. the angles α_i) in an array A_l .
- [4]. Chose another (non-meridian) left image point. Perform steps [1]-[3] and save the results in an array A_j .
- [5]. Compute the set intersection of A_l with A_j and store the results in an array I .
- [6]. Copy I into A_l and repeat steps [4]-[5] until the intersection set I contains only one element.

The reason for not considering meridian points in step [1] is that such points do not have vertical disparities (which carry information about α , see Mayhew (1982) for a discussion of the importance of vertical disparities). The neighbourhood radius in [1] is determined by the maximum allowable horizontal disparity. In step [3], each computed value is actually represented by an uncertainty interval (a fixed proportion of the computed value α) and the intersection operation is performed on these intervals because it is unlikely that the obtained values are ever noise free.

Figure 3 shows a (synthetic, wide-angle) stereo image (of a transparent ellipsoid) compatible with the model. The image plane was quantized into a 1000x1000 array. This quantization and numerical errors (double precision arithmetic was not used) are the principal sources of noise. The algorithm has been tested on a series of such synthetic images with good results. The values of α found were within about ± 2 degrees of the actual value. The algorithm only rarely needed to consider more than about 5 image neighbourhoods (this was the case only when some points in either image did not have matching points in the other image).

3. Discussion and Conclusion.

In this paper we investigated the feasibility of computing the convergence angle for a simple stereo camera model and random dot stereograms without solving the correspondence problem (in fact, in order to be able to solve the correspondence problem). The extension to a less restricted case in which both cameras are free to move is, conceptually at least, straightforward. The computations will become more involved, however, because 2 points (and the corresponding matches) are needed to obtain the convergence angle. One must fear that the combinatorial explosion associated with solutions requiring 4 or more points (with associated matches) will make them computationally prohibitive. The set intersection method used by the algorithm outlined here is much faster and less complicated than the histogram updating suggested in (Prazdny, 1982). This is because it requires only a handful of image neighbourhoods (histogram updating requires a large set of values α_i for a reliable identification of the peak).

As mentioned in the antecedent paper (Prazdny, 1982), it is not clear how to extend a method based on the existence of discrete image locations to the general case of smooth (real world) images. This "pointilistic" way of treating images is probably the most serious drawback of the current method.

REFERENCES.

- Baker H. Blicher P. 1982
Review of Grimson's "From Images to Surfaces"
Sigart Newsletter #82 (October), 12-14
- Duda R.O. Hart P.E. 1973
Pattern Classification and Scene Analysis
New York: Wiley
- Faux I.D. Pratt M.J. 1979
Computational geometry for design and manufacture
Chichester: Ellis Horwood Ltd.
- Gennery D.B. 1979
Stereo Camera Calibration
Proc. DARPA I.U. Workshop 101-107
- Grimson W.E.L. 1981
From Images to Surfaces
Cambridge: MIT Press
- Julesz B. 1971
Foundations of Cyclopean Perception
Chicago: The University of Chicago Press
- Julesz B. Schumer R.A. 1981
Early Visual Perception
Am. Rev. Psychol. 32, 575-627
- Kauffman L. 1974
Sight and Mind
New York: Oxford University Press
- Longuet-Higgins H.C. 1981
A computer algorithm for reconstructing a scene from two projections
Nature 293, 133-135
- Marr D.
Vision
San Francisco: Freeman
- Marr D. Poggio T. 1979
A computational theory of human stereo vision
Proc. R. Soc. London b-204, 301-328
- Mayhew J.E.W. 1982
The interpretation of stereo-disparity information
(submitted to "Perception")
- Mayhew J.E.W. Frisby J.P. 1981
Psychophysical and computational studies towards a theory of human stereopsis
Artificial Intelligence 17, 349-385
- Moravec H.P. 1980
Obstacle avoidance and navigation in the real world by a seeing robot rover
PhD. Thesis, Stanford University Rept. AIM-304a
- Nash J.C. 1979
Compact numerical methods for computers
New York: Wiley
- Prazdny K. 1982
The role of eye position information in algorithms for stereoscopic depth
Proceedings AAAI, 1-4

Figure 1.

Geometry of the stereo camera model used in the paper. Left image plane is fixed, and right camera is allowed to rotate (on the plane with normal $\vec{z}_L = \vec{z}_R$) about a pivot point PP_R (not necessarily) distinct from the centre of projection. FP is the fixation point determined as the intersection of the two principal rays ($\vec{y}_L, \vec{y}_R$). O_L and O_R are the centres of the left and the right coordinate systems, respectively. The mutual orientation of the left and right image planes is determined by the angle α .

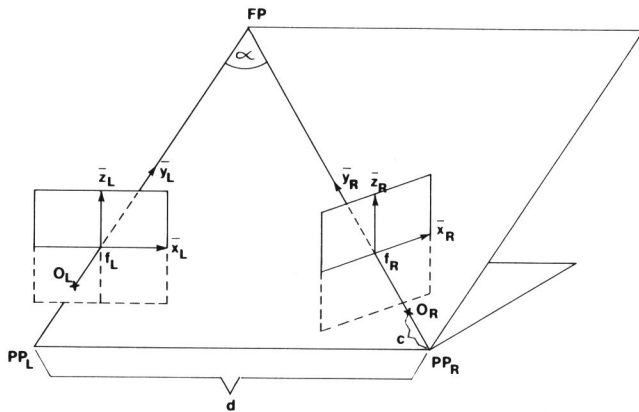


Figure 2.

Consider a right image point (x_R, z_R) with 3D coordinates $\vec{R} = (x_R, f_R, z_R)$. As the (right) camera rotates about the pivot point PP_R , the line determined by $\vec{r}$ going through O_R generates a (ruled) surface. The equation of this surface is:

$$c\vec{y}_R(\alpha) + p(x_R\vec{z}_R(\alpha) + f_R\vec{y}_R(\alpha) + z_R\vec{z}) + d\vec{x}_L - c\vec{y}_L = \mathbf{RS}(\alpha, p)$$

where $\vec{z} = \vec{z}_L = \vec{z}_R$ is the "vertical" (unit) vector common to both coordinate systems. The intersection of this surface $\mathbf{RS}(\alpha, p)$ with a line $\mathbf{L}(\lambda)$ through O_L (determined by the visual direction corresponding to a left image point) can be computed by minimizing the sum of square function

$$|\mathbf{RS}(\alpha, p) - \mathbf{L}(\lambda)|^2.$$

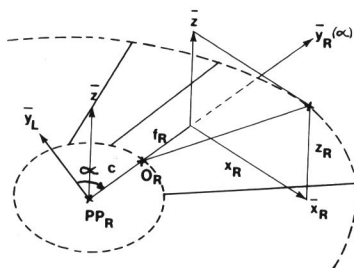


Figure 3.

A (synthetic) wide-angle stereo image pair of a transparent ellipsoid generated using the model in Figure 1 ($d=210\text{mm}$, $c=70\text{mm}$, $f_l=f_r=60\text{mm}$, $\alpha=20$ degrees). The distance of the centre of the ellipsoid is 450mm measured from O_L . Each image contains approximately 400 points.

