# An Adjustment-Free Stereo Matching Algorithm *

Ruihua MA Monique THONNAT Marc BERTHOD
INRIA Sophia-Antipolis
B.P.93, 06902 Sophia-Antipolis Cedex FRANCE
email: ruihua@sophia.inria.fr

**Abstract**

Matching is recognized as the most difficult step in stereo vision. Many algorithms have been proposed but none of them is robust enough. The primary reason for this is that one of the most useful cues, namely *shape similarity*, is not appropriately exploited. Making full use of shape similarity for disambiguation requires determining an appropriate criterion and an adequate implementation. We show that a disparity gradient (DG) limit as small as 0.2 can be used to ensure shape similarity and permits the majority of scene surfaces to be correctly handled. This DG limit is a compromise between the quantity of matches and the quality of matching. The implementation of a DG limit is made effective by explicitly taking into account uncertainty in image point positions. A very effcient voting scheme is proposed. Many tests have been performed, mostly on complex real scenes, but none of the parameters has been adjusted. The results are satisfactory both in terms of the quantity of matches and the correctness of these in normal circumstances.

## 1 Motivation

In stereo, the correspondence problem is recognized as the critical and the most difficult step[2, 3]. Much effort has been spent on this problem during the past decade and numerous matching algorithms proposed[3, 6]. These algorithms present a great diversity involving the input data, the computational scheme, and the way that available constraints are exploited. However there is still room for improvement, in particular in *robustness*, a key performance of vision algorithms.

We believe that effective use of available information is crucial for designing a good matching algorithm. In fact, it is not unusual to see that an algorithm fails to match or yields false matches at locations where available information should allow good matching. In particular, *shape similarity* is not fully exploited, whereas both by intuition and from a information point of view shape similarity is one of the most discriminating cue. However, in order to take advantage of shape similarity, one must first of all know *how to express it, how to measure it, and how to take into account noise and geometric distortion in data.* Therefore, in this paper we focus our effort on answering these questions, based on which a very simple contour-based matching scheme is proposed. Instead of simply describing the algorithm and simplistic analysis, as is ususally the case in the literature, we will give full justification on the various choices during the design.

Performance evaluation is a difficult task in vision because of lack of ground truth. We adopt a visual method using matched contours and 3-D views of the reconstructed scenes (if the imaging system is calibrated). In total, more than 50 stereo pairs were processed among which many are complex outdoor scenes. The results are quite good in terms of the percentage of matched points and the

*This work is supported by the Eureka Project Prometheus.

BMVC 1993 doi:10.5244/C.7.61

quality of matches. But the most important point is that no adjustment has been necessary.

## 2 Generalities of 2-D Matching

Matching is a search. It can be carried out using constraints. In 2-D matching, constraints are *local similarity*, *continuity*, *uniqueness*, *ordering*, etc.. For stereo, the epipolar constraint is also generally available.

Local properties, or features, involve typically graylevel values, edge strength, edge orientation, etc. These do not contain sufficient information for matching, because of various noises and geometric distortion. However, local similarity alone can only be used to select candidate matches in order reduce the search space. More global constraints governing the consistency between matches can be applied for disambiguation.

In the above, matching has been cast into a two-step process, in which local and global constraints are applied separately. An alternative to this is to exploit constraints in a combined way, implicitly or explicitly. Typically, local similarity and continuity constraints are combined, as in most correlation methods. Conceptually such methods need segmentation to ensure that the involved primitives in each of the images result from a same 3-D surface. Ignoring this leads without surprise to deficiencies at occluding boundaries, as with correlation methods.

As pointed out by Marr[10], p.115, available matching constraints are, mathematically speaking, only necessary, but not sufficient conditions. In fact, matching is a underconstrained problem, no matter how we combine the constraints[19]. Repetitive patterns represent a pertinent example. In such a context, a compromise between number of matches and the quality of matching is inevitable.

## 3 Using Shape Similarity for Disambiguation

We cnsider matching points of contours. Every contour conveys shape specifque information. Shape similarity is used here for matching but this does not imply that the shape of each contour should be explicitly characterized. Shape similarity between corresponding contours is reflected in the continuous variation of disparity along the contours and this variation lies within a limit, which is independent of the shape of the contours. On the contrary, two contours do not have these properties if they do not correspond and/or are not similar in shape. In other words, good matches are normally consistent (or support) each other in terms of disparity continuity and variation, whereas bad ones do not, or at least not always so. This fact consistutes the basis of voting scheme for disambiguation.

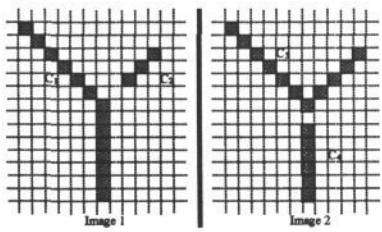


**Figure 1.** Voting is robust with respect to data imperfections.

One of the advantages of this scheme is that it is robust with respect to data imperfections, namely missing corresponding points or different configurations of corresponding edge points at junctions, as illustrated in Fig.1. Unless there exist other contours of similar shape for $\mathcal{C}_1$ in Image 2 (repetitive pattern), the number of votes of a good match of any point on $\mathcal{C}_1$ receive will be always largely superior to that of a bad match, although it may be smaller than in the ideal case.

### 3.1 Determining a shape similarity (continuity) criterion

For a parallel stereo system, if the normal of any 3-D surface is assumed to follow a uniform distribution, an approximate probability density function (PDF) of DG can be derived[1, 17, 13]. The resulting cumulative probability for a given DG

limit $DG_0$ is then:

$$P(DG < DG_0) = \int_0^{DG_0} p_{DG}(DG, \bar{z}) d(DG) = \frac{\bar{z} \cdot DG_0}{\sqrt{1 + \bar{z}^2 \cdot DG_0^2}} \quad (1)$$

where $\bar{z} = z/B$ is what we call *relative depth*, with $B$ denoting the baseline length of stereo system and $z$ the depth. It is drawn in Fig.2. It shows that, for a given DG limit, this probability increases rapidly toward *1.0*. This means the further away the scene surfaces are, the wider the range of surfaces (in terms of slant and tilt) the given DG limit allows us to deal with. For $DG_0 = 0.2$, at $\bar{z} = 5$, more than 70% of surfaces can be dealt with. This percentage is as high as 91% at the same depth given $DG_0 = 0.5$.

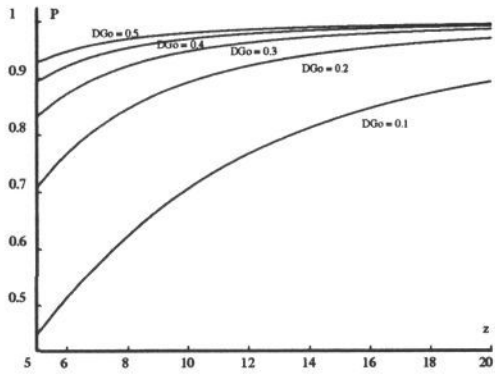


**Figure 2.** The cumulative probability on a DG limit.

At this stage, to choose an appropriate DG limit, a compromise must be made between the quantity of matches and the reliability of matching. If $\bar{z} = 5$ is considered as the minimal depth of these, which is equivalent to $2m$ for $B = 40cm$, and less than 30% of them lie at $\bar{z} \leq 5$, then more than 90% of the total scene points of interest will be correctly dealt with with $DG_0 = 0.2$. This DG limit is appropriate in the sense that it offers a good disambiguation power and at the same time allows to deal with the majority of surfaces in a scene.

### 3.2 Implementation issues

Now we discuss some implementation problems[1], which are not less important than theoretical ones!

#### *3.2.1 Applying DG limit*

Imposing a DG limit $DG_0$ is equivalent to imposing a maximum disparity difference $\Delta disp_0(\Delta dist)$. That is

$$\Delta disp < \Delta disp_0 = DG_0 \cdot \Delta dist. \quad (2)$$

We preferred this form to $DG < DG_0$ for two reasons. First, it is more efficient: a division is replaced by a multiplication; allowable disparity difference can be calculated and stored in a look-up table. Second and more importantly, the noise which is present in an image point position due to the discretization in image formation and dislocation in edge detection can be appropriately coped with by adding a constant term $\Delta d$ to $\Delta disp_0$ (see (2)), which is usually in the order of 1 or 2 pixels.

[1] For details of the implementation, see [9].

### 3.2.2 *Reinforcing the shape similarity constraint*

Most algorithms such as [15][16] which make use of DG limit weight the support (or vote) by the inverse of distance. Their argument was that when $\Delta disp_0(\Delta dist)$ increases, false matches may also get maximal global support by chance. Thus the possibility of choosing false match matches increases, especially when the chosen DG limit is large ($\geq 0.5$). However weighting support has an undesirable effect: the disambiguating power of shape similarity is attenuated.

To avoid this, the solution consists in using a small DG limit as far as this is possible. In our case, as discussed above, a DG limit of 0.2 is used. Further, zero disparity difference is favored over non-zero ones. We simply double the vote (termed as *support* below) of candidates inducing zero disparity difference.

### 3.2.3 *Support collection*

Suppose we have a contour $\mathcal{C}$ composed of $L$ ordered points: $\mathcal{C} = \{p_i, i = 1..L\}$, each point having a list of candidate matches: $\mathcal{M}_i = \{c_{ij}, j = 1..n_i]\}$, where $n_i$ the number of candidates of the $i$th point $p_i$. Thus the total amount of support $c_{ij}$ receives, we call it *score*, is defined as follows:

$$Score\ (c_{ij}) = \sum_{p_k \in \mathcal{N}(p_i)} max\{Support\ (c_{ij}|c_{kl}), c_{kl} \in \mathcal{M}_k\} \tag{3}$$

where $\mathcal{N}(p_i)$ denotes $p_i$'s neighborhood and $Support\ (c_{ij}|c_{kl})$ the support $c_{ij}$ receives from $c_{kl}$ of $p_k$.

Eq. (3) is similar to PMF[14]. Two radical differences are to be noted. The first one is already discussed, namely, support is not weighted here. The second is that local similarity measures (edge norm and orientation), which have been used to select candidate matches, do not take part in the score. In fact, because of the presence of noise, we should no more discriminate selected candidates.

A large neighborhood $\mathcal{N}$ allows a good use of shape similarity and it should be as large as possible. However this also increases the computational cost considerably. Indeed, scoring is the more expensive part. To compensate for this, a sampling technique may be considered[2].

### 3.2.4 *Other problems*

We use a *coarse-to-fine* control structure. Each image of the original stereo pair is consolidated to form a pyramid of images representing different resolution levels. Feature extraction is carried out independently for each level. Disparity values computed at coarser levels are used as a prediction for subsequent levels.

As in [14, 7], matches are validated symmetrically. As would be expected, this strategy of validation allows to greatly reduce the number of false matches. Already matched points are no more considered as candidate of any unmatched point, in order to impose the uniqueness constraint.

We note that *coarse-to-fine* control structure and simultaneous match selection are both mainly intended for a more reliable matching.

Interpolation is performed along a contour to fill unmatched gaps. Before this, we check to see whether at some locations disparity continuity is satisfied on only one side, to make sure that a contour does not belong to more than one surface. If this is the case, the contour is split at such locations.

To conclude this section, we point out that the same constraint (*e.g.*, DG limit) can be implemented quite differently in different algorithms. However, it does not

[2] Used in the latter version of the implementation.

make sense to tune parameters derived in the continuous domain for ideal cases without being able to cope with noise properly. This largely explains why different algorithms using similar constraints perform quite differently.

# 4 Experiments

The algorithm has been tested on numerous stereo pairs, including indoor, outdoor and synthetic scenes. The three scenes we show here are chosen for their complexity. Depth variations in the scenes are important, and this permits to see whether occluding boundaries can be well dealt with.

Evaluation of stereo algorithms is always a tough task due to lack of well-defined procedures, criteria and database with ground truth. So we give here both numerical results and images for qualitative analysis. Different views of the results of 3-D reconstruction are also provided to allow a visual evaluation. Analysis, qualitative and/or quantitative, is made as well, whenever possible.

## 4.1 Input data and Parameters

The input consists essentially of contours made up of connected edges. Edges are detected by the edge detector described in [5]. Edge linking is performed by using the algorithm presented in [8], and the contours serve as support for global consistency checking. Each edge point is associated with its magnitude and orientation.

The parameters involved in the algorithm are: 1. Local similarity criteria (i.e., norm and orientation of each edge point); 2. Neighborhood size for scoring; 3. Radius of search region for estimated disparity; 4. Disparity gradient limit; 5. Disparity range; 6. Number of iterations for the scoring and match validation.

During the whole test, *none* of the parameters has been changed from scene to scene, except the disparity range. In fact, the parameters scarcely affect on the matching results, provided that those lie in a reasonable interval (for details, see [9]). The failures reported result from the characteristics of the images (*e.g.*, repetitive patterns) rather than from an inappropriate choice of some parameters.

## 4.2 Results

**Synthetic scene** This scene contains a cylinder, a cone, an ashtray, a torus, and a calibration grid. The surfaces of the first three objects are textured. From Fig.3, we can see that almost all edge points visible to both views are matched. The 3-D visualization[3] shows that no false match is found. Zigzags in the reconstructed scene are due to the discrete effect in edge position. Notice that textures (including squares on the grid in a large sense) have been correctly matched. This is because the directions of repetition do not coincide with that of sloping epipolar lines, and this does not mean that our algorithm is able to deal with repeating patterns (see Section 5).

**Rocks and grid** This scene contains rocks and a calibration grid on the ground in front of a wall. Due to the small depth, the perspective distortion is important, especially for the rocks. This explains why only 55% (=9372/17047) of edge points find their potential matches (see Table 1). Still, edges resulting from the main structures of the scene are well detected on both images and matched.

Nonetheless, the matching suffers from the repeating squares on the grid. Only a portion of edge points of these are matched and this owing to the coarse-to-fine control structure. Part of the contours on the top of the grid are not well matched, due to two factors: (1) they lie on epipolar lines; (2) the calibration and thus the computation of epipolar lines are not precise. This is also why zigzags in the reconstructed scene are more noticeable than those in the synthetic scene.

[3] Reconstructed points of a same chain are linked. The same is true for all 3-D visualizations.

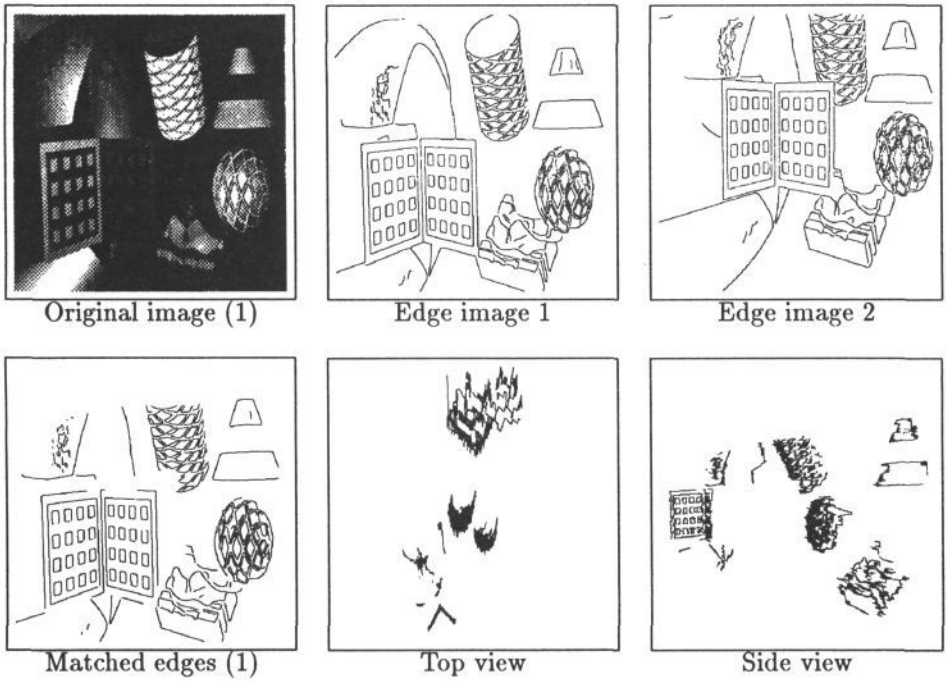


**Figure 3.** The synthetic scene.

**Road** This is a complex road scene composed of moving objects (cars, van, cyclist), road markings, trees, etc.. Edges are well matched. The line indicated by an arrow in Fig.5(f) results from the false edge points on the right side of the images.

Some numerical results are displayed in Table 1. CPU time is measured on a Sun SPARCworkstation IPC.

**Table 1.** Numerical results.

| Scene | Disp range | Nb of points | | Nb of matches | | | | | CPU time (s) |
|---|---|---|---|---|---|---|---|---|---|
| | | T | C | V | N | P | I | F | |
| Synthetic | [0, 70] | 14374 | 12074 | 9990 | 15 | 649 | 1548 | 12143 | 94.3 |
| Rocks and grid | [50, 97] | 17047 | 9372 | 7451 | 68 | 475 | 2834 | 10598 | 115.0 |
| Road | [-40, 10] | 18265 | 12278 | 10510 | 44 | 541 | 2304 | 13204 | 97.9 |

T: total number of edge points in the left image (or View 1; the same for the rest);
C: number of points having at least one candidate match;
V: number of matches obtained by the match validation procedure;
N: number of noisy matches among V (suppressed);
P: number of matches picked up from candidate lists of unmatched and satisfying the disparity continuity constraint;
I: number of matches obtained by disparity interpolation;
F: final total number of matches.

From Table 1 we see that over 80% of points having candidate matches are validated within 3 iterations, with about 80% in the first iteration. Among final


**Figure 4.** Rocks and grid.

matches, up to 25% are obtained by interpolation. Chain splitting guarantees the safety of this operation.

## 5 Evaluation of general performance

### 5.1 Robustness

Robustness has always been our primary concern during the development of this algorithm. Except for the problems discussed in the next section, the robustness of the algorithm is attested by the following points:

1. The parameters in the algorithm are practically insensitive to different types of scenes.
2. The use of DG limit allows us to cope with stereo images which have undergone relatively significant perspective distortions.
3. By relying more on the global consistency constraint and only imposing a rough similarity criterion, occluding contours can be correctly matched.
4. The algorithm is not restrictive with respect to the shape of contours. In fact, a group of points lying on a same chain, as long as they come up to form candidate matches with their counterparts in the other image, will be maximally scored, regardless how the counterparts are grouped, i.e., the shape of contours these counterparts form. This can also be seen in Eq. (3).
5. Except for special cases the algorithm is unable to cope with (Section 5.3), the majority of matchable points are in fact matched.
6. The error percentage is quite small, owing to the use of the small DG limit which is made possible by the constant term taking into account the dislocation of edges in discrete images.

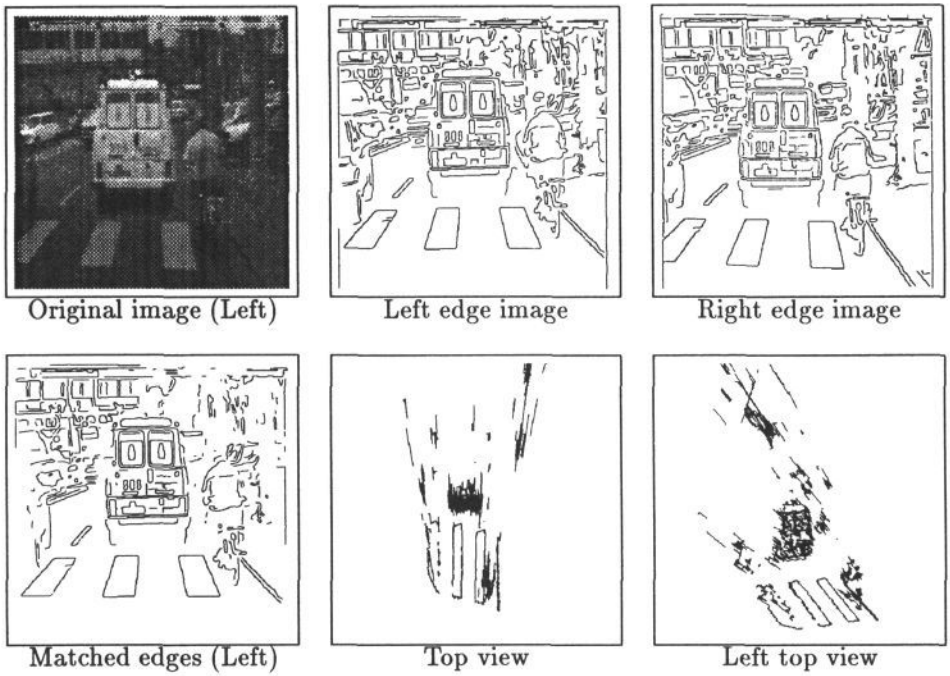


**Figure 5.** The road scene.

### 5.2 CPU time

The CPU time is divided into two parts: extraction of primitives and matching. Matching time is predominantly the extraction of primitives, despite the additional time needed to obtain edge images at coarser resolutions.

During matching, score computation is the most time-consuming. We observe that among validated matches, over 70% result from the first iteration. So, in principle, there is no problem if the number of iteration is set to a minimal number, say, 2. The rest of the matching can be completed by the following steps – selecting matches among candidates and interpolation.

As discussed in Section 3.2.3, an alternative to further reduce scoring time is to use some kind of sampling technique. Moreover, score computation is parallelizable. So to further speed up the matching process, special hardware and/or parallel architectures can be used.

We indicate finally that no special effort has been done to optimize the source code.

### 5.3 Failures and ill-fitted cases

Failures are found principally with repeating patterns. Here we consider a contour segment parallel to the epipolar lines as a special case of repeating patterns, with the edge point being the smallest element.

As for problems in the presence of significant perspective distortion, we think that the problem lies rather in what to match than in how to match. In fact, in such cases, corresponding features are not available because of the dissimilarity of image patches. However, it is evident that in most application contexts, a stereo system using this type of method should be arranged in such a way that perspective distortion, if present, is reasonably small.

Another failure case concerns scenes some portions of which do not contain

sufficient salient features or markings (including apparent occluding contours), as in the rocks scene.

In summary, our algorithm fails in two typical cases: repeating patterns, or lack of salient features. The first problem is inherent to binocular stereo and is still open; the second is a proper problem of the algorithm (*cf.* area-based methods work well on textured regions, which do not necessarily have salient features).

## 6 Conclusion and Future Work

In this paper, we have made a thorough analysis of stereo vision, concentrating on the correspondence problem. We pointed out that one of the most useful cue for matching, namely the *shape similarity*, was not efficiently exploited in other existing algorithms. To make full use of shape similarity, a limit in shape variation must be determined. Moreover, noise in data and geometric distortion must be appropriately handled.

The limit in shape variation is expressed in disparity gradient limit. It is a compromise between the quantity of matches and the quality of matching. We have shown that a DG limit as small as 0.2 can be used, which allows the majority of scene surfaces to be correctly coped with. From an implementation point of view, such a small DG limit is made possible, to some extent, by appropriately expressing the uncertainty in an image point position in the allowable disparity difference.

A simple voting scheme has been proposed. In order to take advantage of the ability of shape similarity for disambiguation, in addition to the use of a small DG limit, support from neighboring points for a candidate match is not weighted An advantageous property of the voting scheme is that it is not restrictive with respect to input data. It is not affected by different configuration of contour points, which frequently encountered in practical cases.

The algorithm is computationally efficient. Of more interest is that its most time consuming part is parallelizable.

The algorithm has been fully tested on a great variety of scenes, in particular complex outdoor ones. Its robustness is evidenced by that fact that *none* of the parameters has been adjusted during those tests, and the results are satisfactory both in terms of the quantity of matches and the correctness of these in normal circumstances.

A possible limitation of this algorithm is that it yields disparity information which might be too sparse to some applications. However, due to the high quality of matches, we think it is particularly promising to combine this algorithm with other schemes such as area correlation based ones, which suffer from occluding contours.

The same voting scheme is also applicable to the temporal matching problem. The only difference between the two problems is that the epipolar constraint is absent in the latter. It suffices then to impose the DG limit in both vertical and horizontal directions.

## Acknowledgment

The synthetic scene is due to Zhengyou Zhang and the original 3-D visualization program due to Luc Robert, Robotvis Project, INRIA Sophia-Antipolis.

## References

[1] R.D. Arnold and T.O. Binford. Geometric constraints in stereo vision. In *Proceedings of SPIE Image Processing for Missile Guidance*, volume 238, pages 281–292, San Diego, 1980.

[2] D.H. Ballard and C.M. Brown. *Computer Vision.* Prentice-Hall, Englewood Cliffs, NJ, 1982.

[3] S.T. Barnard and M.A. Fischler. Computational stereo. *ACM Computing Surveys*, 14(4):553–572, December 1982.

[4] P. Burt and B. Julesz. A disparity gradient limit for binucular fusion. *Science*, 208:615–617, 1980.

[5] R. Deriche. Using Canny's Criteria to Derive a Recursively Implemented Optimal Edge Detector. *International Journal of Computer Vision*, 1(2):167–187, 1987.

[6] U.R. Dhond and J.K. Aggarwal. Structure from stereo – a review. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1489–1510, November 1989.

[7] P. Fua. A parallel stereo algorithm that produces dnse depth maps and preserves image features. *Machine Vision and Applications*, 1991. Accepted for publication, also available as INRIA research report 1369.

[8] G. Giraudon. An Efficient Edge Following Algorithm. In *Proceedings of 5th Scandinavian Conference on Image Analysis*, pages 547–554, Stockholm, 1987.

[9] R. Ma and M. Thonnat. A robust and efficient contour-based stereo matching algorithm. Research Report 1860, INRIA, 1993.

[10] D. Marr. *Vision.* W.H. Freeman, New York, 1982.

[11] D. Marr and T. Poggio. A theory of human stereo vision. Technical Report AI Memo No. 451, MIT, November 1977.

[12] J.E.W. Mayhew and J.P. Frisby. Psychophysical and computational studies toward a theory of human stereopsis. *Artificial Intelligence*, 17:349–385, 1981.

[13] S.B. Pollard, J.E.W. Mayhew, and J.P. Frisby. Disparity gradient, lipschitz continuity, and computing binocular correspondance. Technical Report AIVRU010, AI Vision Research Unit, University of Sheffield, 1985.

[14] S.B. Pollard, J.E.W. Mayhew, and J.P. Frisby. Disparity gradient, lipschitz continuity, and computing binocular correspondance. In *Proceedings of the 3rd International Symposium of Robotics Research*, pages 19–26, Gouvieux, France, 1985.

[15] S.B. Pollard, J.E.W. Mayhew, and J.P. Frisby. PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14:449–470, 1985.

[16] D. Sherman and S. Peleg. Stereo by incremental matching of contours. *IEEE Transactions on PAMI*, 12(11):1102–1106, 1990.

[17] C.V. Stewart. On the derivation of geometric constraints in stereo. Technical Report 91-26, Department of Computer Science, Rensselaer Polytechnique Institute, Troy, New York, August 1991.

[18] H.P. Trivedi and S. A. Lloyd. The role of disparity gradient in stereo vision. In J.E.W. Mayhew and J.P. Frisby, editors, *3D Model Recognition From Stereoscopic Cues*, chapter 2, pages 47–50. The MIT Press, 1990.

[19] J. Weng. A theory of image matching. In *Proceedings of ICCV'90*, pages 200–209, 1990.