A Model-Driven Stereo Correspondence Algorithm Using Dynamic Programming

Y. Shao J. E. W. Mayhew
AIVRU, The Univeristy of Sheffield, Sheffield S10 2TP, UK
email: yuan,mayhew@aivru.shef.ac.uk

Abstract

We describe an edge-based stereo correspondence algorithm in the model-driven vision system. A constraint derived from the 3D model is used to prune false alarms and speed up the matching process. This constraint is based on computional considerations and experimental and psychological observations concerning vertical disparities. An edge constraint is also presented. A numbering scheme is used to facilitate the implementation of the correspondence algorithm using dynamic programming. Experiments on natural images show that the correspondence of an edge can usually be achieved in a few seconds. The computed disparities are further incorporated into object model to refine the estimates of object pose.

Keywords: Correspondence, Constraints, Cost function, Model-driven vision, Dynamic programming

1 Introduction

The work reported here is a part of an ongoing research project "model-driven stereo vision under variable camera geometry", as shown in Figure 1. Three major stages can be identified: *stage 1* uses simple grey-level image processing and 2D templates of 3D object models embedded in a Bayesian statistical reasoning architecture to provide an object localisation system. The outputs of this stage are estimates of the object position and pose; in *stage 2*, a stereo matching algorithm computes the disparities of object templates. It uses the priming preliminary estimates of object pose from the previous stage to facilitate matching. The outputs of this stage are disparities of object template. Details of this stage will be described in this paper; *stage 3* uses an object motion model, disparity information from stage 2, and a smoothing filter to track and foveate the object.

The object used in this research is a Toyota component. Three focus features, corresponding to "bosses" on the part are chosen. Each focus feature consists of two nearly concentric circles and their centre. We use 2D template to represent the projection of the 3D focus features onto image planes. Each template thus consists of two elliptical edge segments and a circular feature, i.e. a blob.

Mapping two 2D images into 3D disparity space is an ill-posed problem. So constraints, derived from *a priori* assumptions about the scene structure and the

BMVC 1996 doi:10.5244/C.10.50

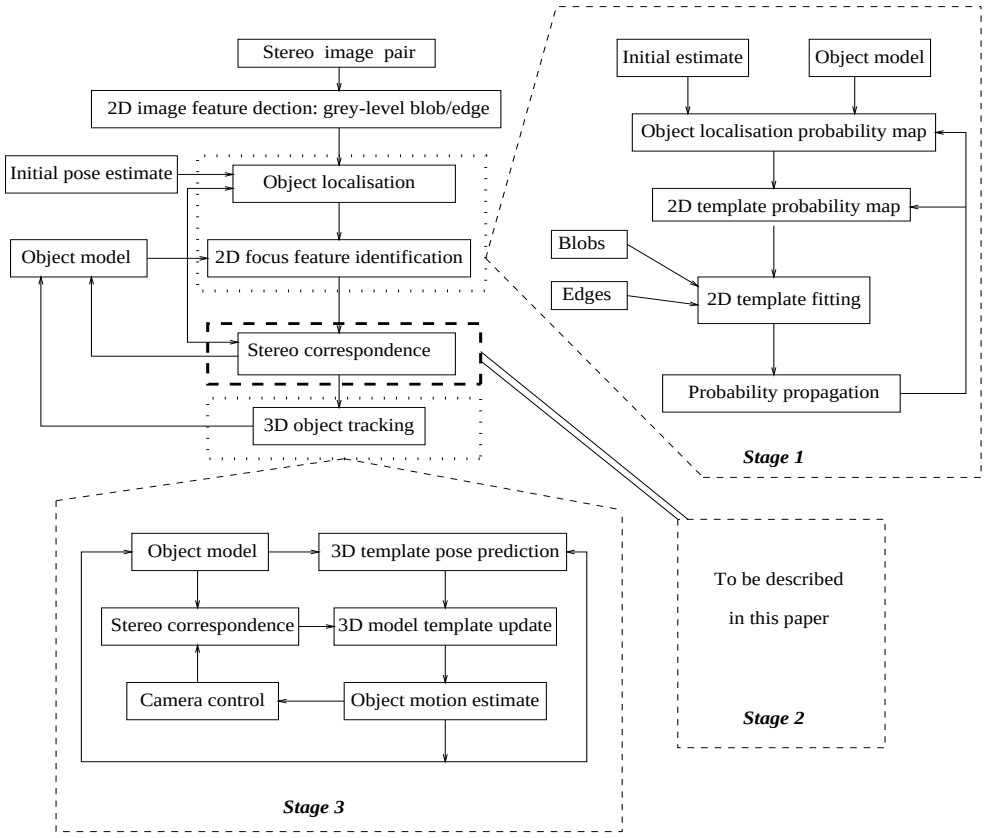


Figure 1: Model-driven stereo vision under variable camera geometry

camera geometry, are used to make the problem solvable. Widely used constraints are *uniqueness* [11], *continuity* [12][9], *smoothness* [1][5] and *the disparity gradient* [13]. To deal with the salient features in a scene, i.e. the discontinuities in depth at boundaries of objects, the discontinuities in surface orientation, and steeply sloping surfaces, Belhumeur [2] suggested a stereo algorithm should internally maintain a detailed map of scene geometry. In his algorithm [3] depth and occluding countours are explicitly represented. Geiger [8] dealt with occlusion by imposing an occlusion constraint, which suspends the smoothness constraint at the boundraies of objects explicitly. Despite the progress made on generic solutions to correspondence process, we suggest that including further constraints derived from the application domain could be used to improve the performance of the matching algorithm. In this paper we use prior knowledge about the scene or objects to prune false alarms and quicken matching. In our model-driven vision [16], the prior knowledge about the scene is encoded in a (weak) model.

Computationally, the solution of the correspondence problem minimises a cost function incorporating constraints. Geiger [7] used dynamic programming (DP) to

address this problem. His algorithm based on DP was basically a 1D process, either along epipolar line or along a given edge. In this paper we propose a numbering scheme to serve representation of the correspondence search space. This scheme allows an intrinsically 1D algorithm (DP) to work in a 2D problem space, i.e. both along epipolar line and along edge features.

The stereo correspondence algorithm takes an object model, (weakly) calibrated camera geometry, focused 2D template and edges as input. After minimizing a cost function under constraints using dynamic programming, it outputs the disparities of an edge template.

2 Notation

Let's denote $e_i^l(e_i^r)$ to edges in left (right) images. And let $p_{mn}^l(p_{mn}^r)$ be points lying on the edge $e_m^l(e_m^r)$. Then the question becomes, for points p_{ij}^l, for all j, on any edge e_i^l, to compute their correspondences with primitives in the right image.

We adopt in this paper matching process M, as used in [8], and extend it into 3 dimension numbering scheme. M is defined as

$$M(j,m,n) = \begin{cases} 1, & \text{if } p_{mn}^r \text{ matches } p_{ij}^l; \\ 0, & \text{else.} \end{cases}$$

Now we define an ordering function X as $X(p_{ij}^l) = j$, and $X(p_{mn}^r) = n$. The physical meaning of X is clear. It tells the order of points p in the edge e. 3D matching space $\mho$ is also defined so that with 3 coordinates representing $X(p_{ij}^l)$, m, and $X(p_{mn}^r)$ respective, illustrated by Figure 2. The correspondence processes $M(j,m,n), \forall j$ become state variables in matching space. The solutions of correspondence problem are then represented as paths through matching space. The question is to search for a constrained path which minimizes the cost.

Since the correspondence algorithm is a minimization process, the key issue is the design of the cost function to be minimized and to achieve this minimization efficiently.

3 Disparity Analysis and Experiment

The binocular viewing geometry can be represented by Figure 3. The image coordinates (x,y) of point $P = (X,Y,Z)^T$ is given by

$$\begin{cases} x = f a_x \frac{X}{Z} \\ y = f a_y \frac{Y}{Z} \end{cases},$$

where (a_x, a_y) are aspect ratios.

Now we move the cyclopean eye by rotation $R = (\omega_x, \omega_y, \omega_z)^T$ and then translation $T = (t_x, t_y, t_z)^T$. Under rigid motion, we have

$$\dot{P} = T - R \times P.$$

By differentiation and substitution of above equations, we acquire horizontal disparity h and vertical one v as

$$\left\{\begin{array}{rcl} h & = & x_r - x_l \approx \dot{x} = fa_x\frac{\dot{X}}{Z} - fa_x\frac{X}{Z}\frac{\dot{X}}{Z} \\ & = & fa_x t_x\lambda - fa_x\omega_y + \frac{a_x}{a_y}\omega_z y - t_z\lambda x + \frac{\omega_x}{fa_y}xy - \frac{\omega_y}{fa_x}x^2 \\ v & = & y_r - y_l \approx \dot{y} = fa_y\frac{\dot{Y}}{Z} - fa_y\frac{Y}{Z}\frac{\dot{Z}}{Z} \\ & = & fa_y t_y\lambda + fa_y\omega_x - t_z\lambda y - \omega_z\frac{a_y}{a_x}x + \frac{\omega_y}{fa_x}xy + \frac{\omega_x}{fa_y}y^2 \end{array}\right. ,$$

where $\lambda = 1/Z$.

Back to the two view eyes, we can have $(t_x, t_y, t_z) = (-I\cos\gamma, 0, -I\sin\gamma)$ and $\omega_y = t_x/d$. So

$$\left\{\begin{array}{rcl} h & = & -Ifa_x\cos\gamma\lambda + I\sin\gamma x\lambda + g_h() \\ v & = & I\sin\gamma y\lambda + Ay^2 + Bxy + Cx + Dy + E \end{array}\right. ,$$

where $g_h()$ and cofficiencies $A - E$ are dependent only on camera geometry and image coordinates.

Since the deviation from symmetric vergence $I\sin\gamma$ is generally small. The vertical disparity thus depends very weakly on the $\lambda(x, y)$, i.e. the depth structure of the scene. In other words, we can assume that the vertical disparity v encodes only the camera geometry. Thus, we acquire

$$v \approx \hat{v} = Ay^2 + Bxy + Cx + Dy + E.$$

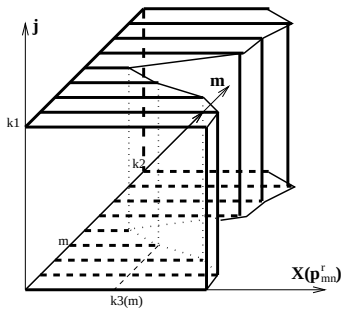


Figure 2: 3D match space for feature-based correspondence, within which search occurs. Axis **j** is the primitives in left image, axis **m** is edges in right image, and $X(p^r_{mn})$ are primitives on those edges, represented by the thick dashed lines

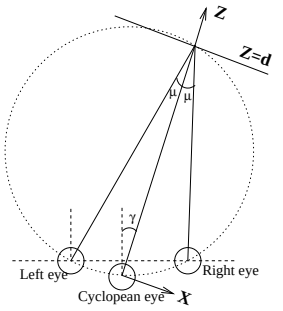


Figure 3: Cyclopean camera geometry

4 Model Prediction

In model driven vision [16], the system possesses knowledge about a scene or objects. This knowledge can then be used to predict correspondences of an edge primitive which is a component of the model templates [16].

Since vertical disparity is very weakly dependent on depth structure of scene around fixation area [6], we use the approximate expression of vertical disparity $\hat{v}$ to serve as a predictor. The cofficiencies $A - E$ are estimated directly from model disparities $v(x, y)$ by linear least squares.

The model constraint can then be defined as

$$| v - \hat{v} | < \zeta$$

where ζ are the deviation limit derived from the vertical disparity approximation cofficiency estimate variances and the camera calibration.

5 Constraints and Cost Function

The cost function is defined by constraints imposed on the correspondence process, which reflect assumptions underlying the matching process. In our stereo correponding algorithm, following constraints are used.

Uniqueness With uniqueness constraint, we assume there is at most one match to any primitive in the left imge. Uniqueness prohibits multiple matches from occuring. The uniqueness constraint can be stated as $\sum_m \sum_n M(j, m, n) = 0, 1, \quad \forall j$.

Monotonicity The monotonicity can be expressed as that the function $X(p^r_{mn})$ is *monotonic* with respect to $X(p^l_{ij})$ for all such j that p^l_{ij} matches p^r_{mn}. It has been pointed out [8] that monotonicity constraint can be violated, such as in the case of the double-nail illusion [10]. Actually this violation may also occur due to noise in images or the lack of robustness of edge operator. The occurances in this case are always at the end points of an edge. Therefore, the monotonicity constraint is suspended for those points whose correspondences are end points of an edge.

Disparity Gradient In the PMF stereo [14], disparity gradient has been exploited to give neibourhood support to a match.

For two matches $M(j_0, m_0, n_0) = 1$ and $M(j_1, m_1, n_1) = 1$, their disparity gradient δd is defined as $\delta d = \frac{2\|(p^r_{m_0 n_0} - p^l_{i_1 j_1}) - (p^r_{m_0 n_0} - p^l_{i_0 j_0})\|}{\|(p^r_{m_1 n_1} + p^l_{i_1 j_1}) - (p^r_{m_0 n_0} + p_{i_0 j_0})\|}$. Then the disparity gradient can be represented as $\delta d_{jj+1} \leq \delta d_{max}$, where δd_{max} is disparity gradient limit, usually 1.1 or 1.2 [13]. As mentioned in [13], imposing a limit on disparity gradient implies a limit on the reorientation of edge segments in order that they be allowed to be matched.

Epipolar Constraint Ideally the correspondence of a primitive should lie in the epipolar line. In practice, due to errors in estimate of camera geometry and errors of imaging, deviation from the epipolar line should be allowed, though at a penalty, provided this deviation is within a limit.

Edge Constraint In an edge-based correspondence problem, it is safe to assume that neighbouring points of a single edge in one image would more likely

match points of another single edge in another image. But in practice applying an edge filter on the projection of a 3D edge may result in several segments. Consequently 2 neighbouring points of one edge, with one matching to either end point of an edge, may not match to two points of one edge.

So we can define the edge constraint as that *correspondences of neighbouring primitives lie on some edge except for those primitives whose correspondences are end points of a edge.*

The cost function can thus be defined as

$$C = C_1 + C_2 + C_3 + C_4 + C_5.$$

The first item C_1 is based on the correlation at intensity level, i.e.,

$$C_1 = \sum_j \sum_m \sum_n M(j,m,n) Corr(j,m,n),$$

where $Corr$ defines the correlation process. The second item C_2 is a penalty for unmatched primitives. Denote a_u to penalty for one unmatched primitive, then

$$C_2 = a_u \sum_j (1 - \sum_m \sum_n M(j,m,n)).$$

The third term is a penalty for deviation from the epipolar constraint.

$$C_3 = \sum_j \sum_m \sum_n M(j,m,n) f_d(j,m,n),$$

where $f_d(j,m,n)$ defines the epipolar deviation penalty. The forth term is an effective piecewise smooth function [8].

$$C_4 = \mu \sum_j \sqrt{\mid D_{j+1} - D_j \mid},$$

where D_j is the disparity with respect to primitive p^l_{ij}. It implies correspondences of neibouring primitives are usually neibouring and also discourage staircase-like disparity corresponding to interlaced matched and occluded primitives. Let $D(j,m,n)$ be the disparity between the points p^r_{mn} and p^l_{ij} and rewrite C_4, we have

$$C_4 = \mu \sum_j \sqrt{\mid \sum_m \sum_n M(j+1,m,n) D(j+1,m,n) - \sum_m \sum_n M(j,m,n) D(j,m,n)}.$$

The fifth cost item C_5 is from an edge constraint which states correspondences of neibouring primitives usually lie within some edge. So we define

$$C_5 = \lambda \sum_j u(\sum_m \sum_n (M(j+1,m,n) - M(j,m,n)) m),$$

where function u is defined as $u(x) = \begin{cases} 0, & \text{if } x = 0, \\ 1, & \text{else.} \end{cases}$ It should be noticed that both the cost components C_4 and C_5 make contribution to a solution of smooth surface. But C_4 is defined on pixel coordinates, while C_5 is defined on the numbering scheme, i.e. the matching space.

6 Dynamic Programming

To minimize cost function C with respect to M is a problem of $\sum_j \sum_m \sum_n$ variables. We here use dynamic programming to transform it into $\sum_j \sum_m \sum_n$ subproblems each of which consists of one variable. Since computations increase exponentially with the number of variable, but only linearly with the number of subproblems, the advantage of using dynamic programming is tremendous.

Searching using dynamic programming is illustrated by Figure 4. Suppose the current searching point is (j, m, n), then the required subproblems, without any constraint, should be all possible points lies on the plane $Z = j-1$. While the edge constraint reduces them to the shaded line, i.e. on the same edge, the epipolar constraint and the model constraint cut the line with two curves. Finally, applying the monotonicity constraint defines the short piece of thick line.

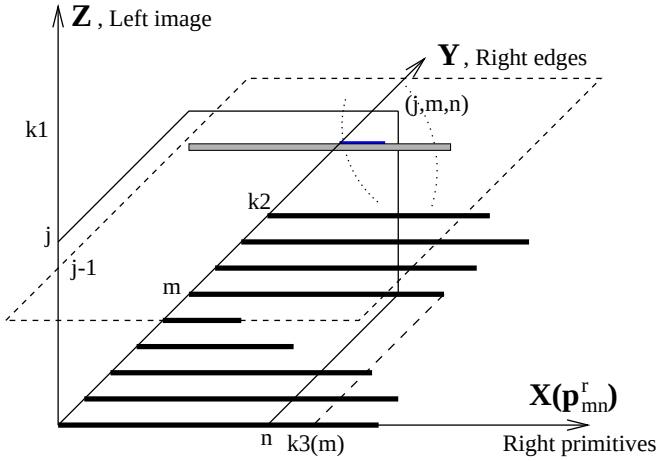


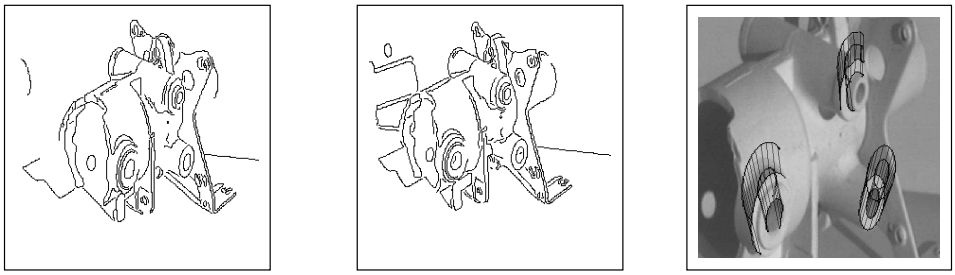
Figure 4: Correspondence searching using dynamic programming

When the edge constraint cannot hold, as in the cases of end primitives, the required subproblems are given by epipolar constraint, preferably beginning primitives of edges where the epipolar constraint hold. This is also used to provide starting searching points at $Z = k_1$, where no edge constraint is available.

7 Experiment

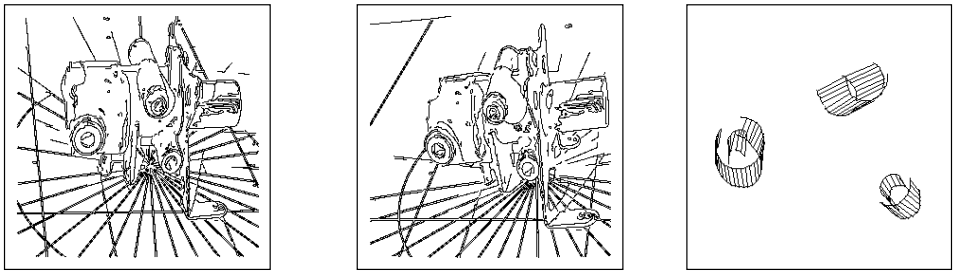
The algorithm was implemented using the TINA vision system [15]. The experiments on an industrial object was shown in Figure 5. Figure 5(c) has been zoomed to give a more intelligible illustration. The correspondence results are satisfying and are achieved in a few seconds. Figure 6 shows an experiment on the same object but at another pose and on a more cluttered background.

We use the computed disparities to recover the scene structure using the calibrated camera geometry and prior model information. The accuracy of this recovered pose estimate can be used as a measure of the performance of the matching



(a) Edges in the left image (b) Edges in the right image (c)correspondence

Figure 5: Matching results for edges fitting with 2D templates of the object focus features. Thin lines in (c) gives disparity for every fifth points along the curve



(a) Edges in the left image (b) Edges in the right image (c)correspondence

Figure 6: Matching results for edges fitting with 2D templates of the object focus features with the object at another pose. Thin lines in (c) gives disparity for every fifth points along the curve

algorithm.

The recovered three templates of the object are shown in Figure 7 and Figure 8, corresponding to Figure 5 and Figure 6 respectively.

8 Summary

An edge-based correspondence algorithm is described for use in a model-driven stereo vision context. The algorithm make an extensive use of prior knowledge about the object. A model constraint, which sets bounds of vertical disparities, is developed to prune false alarm and speed up the search. Dynamic programming is used to search the correspondence efficiently.

9 Acknowledgement

Y. SHAO is sponsored by the Sino-British Friendship Scholarship Scheme. Authors feel indebted to all members in AIVRU at the University of Sheffield.

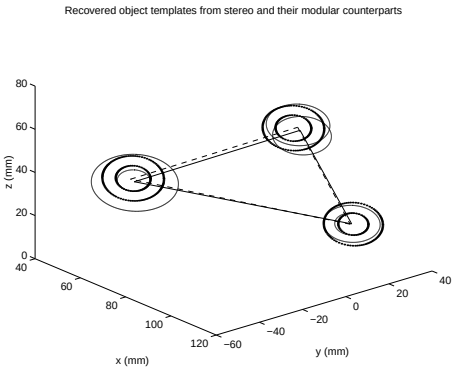


Figure 7: Recovered object templates and their true counterparts. The dashed lines and circles by thickened lines are for the object templates, while the solid line triangle and circles by normal lines are those recovered. the position error is $3.4mm$ and the pose error is $1.1°$

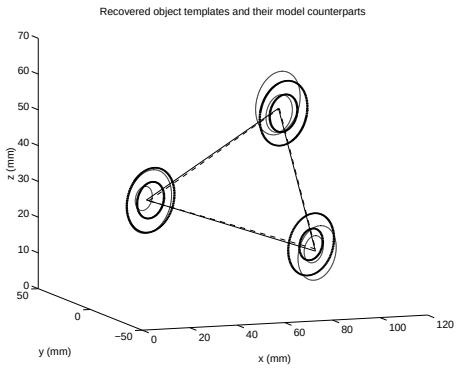


Figure 8: Recovered object templates and their true counterparts with object at another pose. The position error is $1.0mm$ and the pose error is $2.7°$

References

[1] S. T. Barnard, Stochastic stereo matching over scale, *Int'l J Computer Vision*, **3**(1):17-32

[2] P. N. Belhumeur, A binocular stereo algorithm for reconstructing sloping, creased and broken surfaces in the presence of half occlusion, *Proc. 4th Int'l Conf on Computer Vision*, 1993, 431-438

[3] P. N. Belhumeur and D. Mumford, A bayesian teatment of stereo correspondence problem using half-occluded region, *Proc IEEE Conf CVPR*, Urbana, IL, June 1992

[4] J. Canny, A computational approach to edge detection, *IEEE Trans PAMI*, 1988, **8**(6):679-698

[5] B. Cernuschi-Frias, D. B. Cooper, Y. P. Hung and P. N. Belhumeur, Toward a model-based bayesian theory for estimation and recognizing parameterized 3-D objects using two or more images taken from different positions, *IEEE Trans PAMI*, 1989, **11**(10):1028-1052

[6] J. Gårding, J. Porrill, J. P. Frisby, and J. E. W. Mayhew, Uncalibrated relief reconstruction and model alignment from binocular disparity, *Proc. 4th ECCV*, Apr. 1996, Cambridge, England

[7] D. Geiger, A. Gupta, L. A. Costa and J. Vlontzos, Dynamic programming for detecting, tracking, and matching deformable contours, *IEEE Trans PAMI*, **17**(3):294-302

[8] D. Geiger, B. Ladendorf and A. Yuille, Occlusions and binocular stereo, *Int'l J of Computer Vision*, 1995, **14**:211-226

[9] W. E. L. Grimson, Computational experiments with a feature based stereo algorithm, *IEEE Trans PAMI*, 1985, **7**(1):17-34

[10] J. D. Krol and W. A. Van der Grind, The Double-nail illusion: Experiments on binocular vision with nails, needles And pins, Perception, 1982 **11**:615-619

[11] D. Marr and T. Poggio, Cooperative computation of stereo disparity, *Science*, **194**:283-287, Oct. 1976

[12] J. E. W. Mayhew and J. P. Frisby, Psychological and computational studies towards a theory of human stereopsis, *Artificial Intelligence*, 1981, **17**

[13] S. B. Pollard, Identifying correspondences, *PhD Thesis*, University of Sheffield, Nove. 1995

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

[15] J. Porrill, S. B. Pollard, T. P. Pridmore, TINA: The Sheffield AIVRU vision system, *Proc. of 10th Int'l Jiont Conf on Artificial Intelligence*, 1987, **2**:1138-1144

[16] Y. Shao, J. E. W. Mayhew, Object localisation using model-driven vision, *AIVRU Memo-106*, University of Sheffield, March 1996