Prediction of Stereo Disparity using Optical Flow

P.A.Beardsley, J.M.Brady & D.W.Murray

Department of Engineering Science
University of Oxford
Oxford OX1 3PJ

This paper describes a scheme in which optical flow information is used to guide stereo correspondence matching. The epipolar constraint, which is used in static stereo to confine correspondence search to a line, is replaced by a mechanism which confines search to an area around a point.

The experimental setup consists of a pair of cameras mounted on a mobile robot vehicle. A sequence of stereo image pairs is taken during vehicle motion, and the images are processed by a corner detector. One of the first steps in extracting useful information from the corners is to solve the correspondence problem. For a stereo sequence, this requires both the temporal matching of corners along the sequence, and also the stereo matching of corners between left and right images. The results of the temporal matching (which is easier than the stereo matching) are used to guide the stereo matching. The cameras are assumed to lie in a plane perpendicular to the direction of straight-forward motion, but no other assumptions are made about their orientation, and no knowledge of their extrinsic parameters need be provided as input to the system. The system is currently limited to straight translational motion in a static scene.

One of the fundamental tasks in an autonomous vehicle navigation system is to create a model of the scene around the vehicle. This model, possibly in combination with a predefined map of the environment, provides the basis for obstacle avoidance and route planning. As in most areas of AI, the choice of representation for the model is a key issue [1]. In vehicle navigation, a 3D representation of a scene is often used i.e. points in the scene are represented by 3D position (see, for instance, [2, 3, 4]). In contrast to this approach, we intend to investigate the possibility of carrying out navigation without **explicit** 3D information, using quantities which are more directly related to image plane measurements yet which still encode the information required for the navigation task.

This idea is undeveloped at the moment. However, it has provided the motivation for the work described here. The purpose of the work is to carry out temporal and stereo correspondence matching in an integrated way. It is certainly true that this integration can be carried out effectively in 3D-based systems [5], but the method in this paper performs the integration without reference to explicit 3D information. The overall approach was suggested by [6], and the main features of the integration are listed below. The current system is limited to **straight translational motion** in a **static scene**.

- Temporal correspondence matching is used to support the stereo correspondence matching. This approach was adopted for two reasons. Firstly, the the combination of rate of image capture and typical vehicle speed is such that corners are only displaced by small amounts in consecutive images in a sequence so only a small search area is required for temporal matching - in contrast, the search area during stereo matching is along an epipolar line so matching ambiguities are more likely to arise. Secondly, the corner detector used [7] is sensitive to change in viewpoint. This affects the 'corner strength' attribute which is generated to characterise a corner, and hence affects the matching processes. Temporal matching is less affected than stereo matching because the change in viewpoint between two consecutive frames of a sequence is much smaller than the change in viewpoint between the left and right cameras.

- Two attributes are used for corner comparison during stereo matching - corner strength and a motion-based attribute.

- Stereo matches found in one stereo pair are 'cascaded' forwards to initiate matching in the next stereo pair in the sequence.

- Finally, the main subject of the paper is a method by which the optical flow at a corner is used to predict its stereo disparity. Given this prediction, the search area for stereo matching is confined around a point instead of along an epipolar line.

Section 1 describes system initialisation, section 2 the basic concepts underlying the main processing, section 3 the main processing itself, and section 4 the results.

1. SYSTEM INITIALISATION

This section describes the initialisation carried out before the main processing can begin. The purpose of

BMVC 1990 doi:10.5244/C.4.46

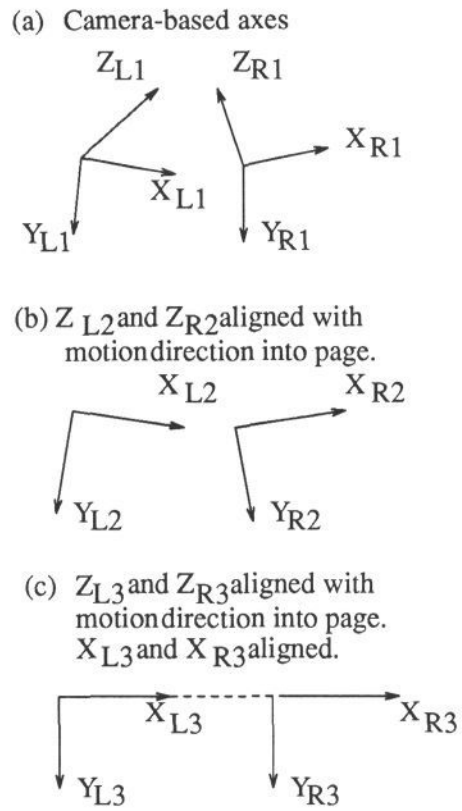

Figure 1: *Co-ordinate frames.*

the initialisation is to determine a rotation matrix which will transform image plane co-ordinates into a new co-ordinate frame suitable for the main processing. The cameras' intrinsic parameters (focal length, point of intersection of the optical axis with the image plane, scaling factors) are assumed to be known before initialisation begins. They were found using [8]. No knowledge of the translation and rotation between the two cameras need be provided as input to the system.

The required transformation has two stages - see figure 1. In (a), the left and right camera co-ordinate frames have arbitrary relative orientation. In (b), the Z-axis of each camera is aligned with the direction of motion. In (c), the X-axis of each camera is aligned to pass through the optical centre of the other camera. Conditions (b) and (c) together require that the optical centres of the two cameras lie in a plane which is perpendicular to the direction of motion.

The transformation from $X_1Y_1Z_1$ to $X_2Y_2Z_2$ is carried out independently for each camera in the following way. The focus of expansion is found for straight-forward motion of the vehicle. An iterative method is used that alternately updates estimates of point-depths and of the position of the FOE in order to converge to a least squares solution [9]. At this stage, two significant vectors are available - one along the optical axis (the Z_1-axis) determined from the intrinsic parameters, and one in the direction of motion (the Z_2-axis) determined from the FOE. The rotation between the two vectors can be regarded as the combination of two rotations - one around the X_1-axis and one around the Y_1-axis. A rotation matrix is computed for each of these rotations, and the two rotation matrices are then combined into a single matrix which effects the transformation $X_1Y_1Z_1$ to $X_2Y_2Z_2$.

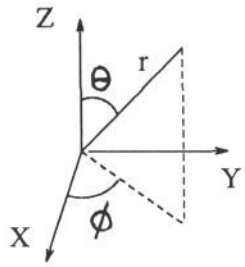

Figure 2: *Spherical co-ordinates.*

Two degrees of rotational freedom between the cameras are removed by the transformation above, but there is still a degree of freedom around the Z_2 axis. The transformation $X_2Y_2Z_2$ to $X_3Y_3Z_3$ to remove this degree of freedom can be determined if stereo matches are known. However, this transformation proved unnecessary. In our system, the cameras are verged and pointed slightly downwards, but they are mounted at the same vertical height on the vehicle and there is little rotation (in terms of the 'uprightness' of the cameras) around the axis along the forward direction of vehicle motion. Thus, $X_2Y_2Z_2$ and $X_3Y_3Z_3$ are nearly parallel in our system, and experimental results suggest that they are sufficiently close for this transformation to be ignored. However, more work will be carried out in this area in the future.

Having determined the rotation matrix, it is available for use during the main processing to transform image plane co-ordinates into the $X_3Y_3Z_3$ co-ordinate frame. Corners are held in spherical co-ordinates in the $X_3Y_3Z_3$ frame - see figure 2. No physically meaningful value can be assigned to the r component, and only the (θ, ϕ) components are used. All references to corner positions in the remainder of the paper should be assumed to refer to (θ, ϕ) co-ordinates in the $X_3Y_3Z_3$ frame.

2. BASIC CONCEPTS

The first part of this section describes the stereo camera geometry, and the remaining parts give the derivation of various useful quantities. The actual use of these quantities is discussed later, in section 3.

2.1. The Stereo Camera Geometry

Figure 3 shows a point P which has spherical co-ordinates (θ_L, ϕ_L) with respect to the left camera, and (θ_R, ϕ_R) with respect to the right camera. P is at distance Z along the Z-axis, and the figure shows the XY plane in which P lies. B is the length of the baseline between the

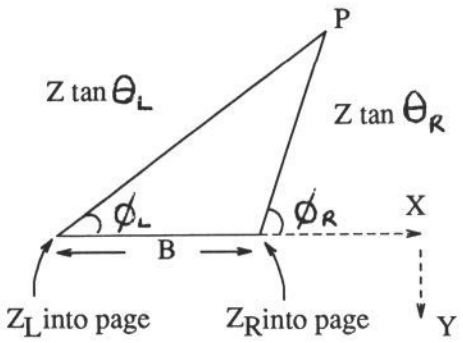

Figure 3: *Stereo camera geometry.*

two cameras.

By the assumptions made in section 1 when setting up the co-ordinate frame, the displacement of P along the Z-axis is the same for both cameras. Also note that the perpendicular distance from the left Z-axis to P is $Z\tan\theta_L$ (refer to figure 2), and a similar expression applies for the right Z-axis.

From the geometry of the figure -

$$\tan\theta_L \cos\phi_L - \tan\theta_R \cos\phi_R = \frac{B}{Z} \tag{1}$$

$$\tan\theta_L \sin\phi_L = \tan\theta_R \sin\phi_R \tag{2}$$

2.2. Computation of $\Delta Z/Z$

$\Delta Z/Z$ is a ratio which can be measured at an image corner. ΔZ is the displacement of a physical feature along the Z-axis (i.e. along the direction of motion) between the previous image and the current image. Z is the distance of the physical feature along the Z-axis at the time of the current image. The ΔZ value is common to all corners at any particular time, but the Z value varies from one corner to another.

$\Delta Z/Z$ is derived in the following way. Consider a camera moving towards a point P with the Z-axis aligned with the direction of motion (alignment of the X- and Y- axes is immaterial to this analysis). Equivalently, the camera can be regarded as static, and P in motion parallel to the Z-axis - see figure 4.

At time t,

$$\tan\theta = \frac{\sqrt{X^2+Y^2}}{Z+\Delta Z} \tag{3}$$

At time (t+Δt),

$$\tan\theta' = \frac{\sqrt{X^2+Y^2}}{Z} \tag{4}$$

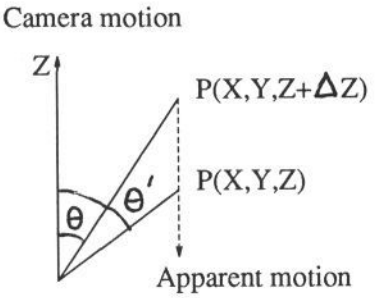

Figure 4: *Computation of $\Delta Z/Z$.*

From equations (3) and (4)

$$\frac{\Delta Z}{Z} = \frac{\tan\theta'}{\tan\theta} - 1 \tag{5}$$

2.3. Computation of $B/\Delta Z$

$B/\Delta Z$ is a ratio which can be measured from a stereo match, provided $\Delta Z/Z$ is known at at least one of the corners in the match. B is the length of the stereo baseline. ΔZ was described in the previous section. B is a constant, and ΔZ is the same for all corners at any particular time. Therefore, at a particular time, $B/\Delta Z$ is invariant with respect to the choice of the stereo match used to generate it.

$B/\Delta Z$ is derived in the following way. Consider a stereo match between corners with co-ordinates (θ_L, ϕ_L) and (θ_R, ϕ_R). If $\Delta Z/Z$ is known at at least one of the corners then, using equation (1)

$$\begin{aligned} \frac{B}{\Delta Z} &= \frac{B}{Z}\cdot\frac{1}{\frac{\Delta Z}{Z}} \\ &= \frac{(\tan\theta_L \cos\phi_L - \tan\theta_R \cos\phi_R)}{\frac{\Delta Z}{Z}} \end{aligned} \tag{6}$$

If there were no sources of error, then the corner in the left image and its match in the right image would both have the same value of $\Delta Z/Z$ (since ΔZ is the same for all corners in both images at a particular time, and a physical feature has the same Z with respect to both cameras in the co-ordinate frame set up in section 1). In a real system, however, the left and right corners of the stereo match have different values of $\Delta Z/Z$ due to error, so some decision is required on what value is inserted into the denominator of equation (6). This is discussed in section 3.1.

2.4. Prediction of Stereo Disparity

Consider a corner (θ_c, ϕ_c) at which $\Delta Z/Z$ is known, but for which no stereo match has been found. If $B/\Delta Z$ is known then, using (1),

$$\frac{B}{\Delta Z}.\frac{\Delta Z}{Z} = \frac{B}{Z} = \tan\theta_c \cos\phi_c - \tan\theta_p \cos\phi_p \quad (7)$$

where (θ_p, ϕ_p) is the position of the predicted match for the corner.

Using (2),

$$\tan\theta_c \sin\phi_c = \tan\theta_p \sin\phi_p \quad (8)$$

Using (7) and (8)

$$\phi_p = \arctan\left(\frac{\tan\theta_c \sin\phi_c}{\tan\theta_c \cos\phi_c - \frac{B}{\Delta Z}.\frac{\Delta Z}{Z}}\right) \quad (9)$$

$$\theta_p = \arctan\left(\frac{\tan\theta_c \sin\phi_c}{\sin\phi_p}\right) \quad (10)$$

or

$$\theta_p = \arctan\left(\frac{\tan\theta_c \cos\phi_c - \frac{B}{\Delta Z}.\frac{\Delta Z}{Z}}{\cos\phi_p}\right) \quad (11)$$

The physical solution may actually be either ϕ_p as given in equation (9), or $\phi_p + \pi$. The correct solution is the one for which $\tan\theta_p \geq$ zero. The choice between equation (10) and (11) depends upon the denominator being non-zero.

3. SYSTEM OPERATION

The object of the processing is to carry out two types of correspondence matching on the stereo image sequence - temporal matching of corners along an image sequence from one particular camera, and stereo matching of corners between left and right images at one particular time. The purpose of integrating the temporal and stereo matching is to increase reliability and to reduce the processing required for a non-integrated approach. A fundamental aspect of the integration is the decision to use temporal matching as support for stereo matching - justifications for this were given in the opening section.

The first part of this section gives an overview of the work, and the second part gives a description of typical system operation.

3.1. Overview of Stereo-Motion Integration

This subsection does not attempt to cover the operation of the whole system - instead it concentrates on areas relevant to the stereo-motion integration.

Temporal matching is carried out just as it would be in a stand-alone motion system with no stereo. Corners are matched between consecutive images along a sequence, using corner strength as the matching attribute, and using an assumption of small displacement to determine the search area for potential matches. Each time a match is found, $\Delta Z/Z$ is computed (section 2.2) and stored with the corner.

The stereo matching makes use of the temporal matching in three ways, and these are described below.

(a) When comparing a corner in one stereo image with potential matches in the second stereo image, the matching attributes used are corner strength and the $\Delta Z/Z$ value. $\Delta Z/Z$ is identical at each corner in a stereo pair (in actual fact, $\Delta Z/Z$ is identical for **all** corners which lie at the same depth from the camera, so the use of this attribute is only a partial aid to disambiguation during matching).

(b) Instead of carrying out stereo matching from scratch on each new stereo pair, stereo matches are 'cascaded' forward through the image sequence whenever possible. This is carried out in the following way (using the notation 'Ln' and 'Rn' to indicate the left and right images respectively of frame n of the sequence). If corner A in Ln has a stereo match with corner B in Rn, and A has a motion match with A' in $L(n+1)$, and B has a motion match with B' in $R(n+1)$, then A'-B' is regarded as a potentially acceptable stereo match. The corner strength and $\Delta Z/Z$ attributes are tested to see whether the information available in frame $(n+1)$ supports the match and, if so, it is accepted. In this way, frame n stereo matches give rise to an initial set of frame $(n+1)$ stereo matches without any need for extensive search.

(c) The third technique used in stereo matching is the use of the stereo disparity prediction mechanism. The starting point for this processing is to generate a small set of reliable 'seed' stereo matches. This is done in one of two ways, depending on the current state of the processing. The easiest way to generate the seed set is by cascading forward stereo matches from the previous stereo pair to the current stereo pair - this was described above in (b). The cascade method will not be available at system startup. In this case, the seed set is generated by the usual techniques of static stereo matching - however, instead of attempting to find as many matches as possible, only those matches which have a very high confidence rating are accepted into the seed set.

The dominant criterion in creating the seed set is that virtually all the stereo matches which it contains should be correct. It is of less importance if the seed set represents only a small percentage of the total number of stereo matches.

Given the seed set, it is possible to compute $B/\Delta Z$ (section 2.3). In an error-free system, the generation of $B/\Delta Z$ would be straightforward - each one of the seed stereo matches would give rise to exactly the same value of $B/\Delta Z$, so the value could be found from a single stereo match. In a real system the generated values vary. $B/\Delta Z$ is taken to be the average of the values produced by the seed stereo matches.

A further problem in the generation of $B/\Delta Z$ arises from the fact that the computation uses the $\Delta Z/Z$ value of just one of the corners in a stereo match. In an error-free system, each corner in a stereo match would have the same $\Delta Z/Z$. In a real system, the values vary so some decision must be taken on which value is used in the generation of $B/\Delta Z$. The current approach is to generate one value of $B/\Delta Z$ using only the left corners from the seed stereo matches, and one value using only the right corners. Call these $(\frac{B}{\Delta Z})_L$ and $(\frac{B}{\Delta Z})_R$ respectively.

This fits in with the operation of the stereo matcher, which uses the technique of matching left-to-right image, then right-to-left image, and then selecting stereo matches which have been found at both stages. The $(\frac{B}{\Delta Z})_L$ value is used to predict stereo disparity when matching left-to-right, and the $(\frac{B}{\Delta Z})_R$ value is used when matching right-to-left. (The use of $B/\Delta Z$ to predict disparity is described in section 2.4.)

Having produced a disparity prediction for a corner, it is only necessary to search a small area around the predicted point to find the corner's stereo match. The search area is defined in the following way. Given a predicted position for a stereo match at (θ_p,ϕ_p), and a potential candidate for the stereo match at (θ_c,ϕ_c), the candidate match is acceptable (for further checking) only if

$$\sqrt{(\theta_p - \theta_c)^2 + (\phi_p - \phi_c)^2 \sin^2 \theta_p} \leq 0.03 \qquad (12)$$

The left-hand side of equation (12) represents arc length between two points on a unit sphere in a spherical co-ordinate frame. The '0.03' threshold was found by trial and error.

This approach contrasts with static stereo in which search is along an epipolar line - in this case, the search area is larger and matching ambiguities are more likely to arise.

3.2. Typical System Operation

Every image in the stereo sequence is subject to the following operations before other processing begins - (a) corners are identified in the images; (b) the transformation determined in section 1 is applied to the corners to produce spherical co-ordinates in the required co-ordinate frame. In the description below, each frame number refers to a stereo image **pair** from the image sequence.

Frame 1. No stereo matching is attempted at this stage, because no optical flow information is available.

Frame 2. Temporal matching is carried out between frame 1 and frame 2. The quantity $\Delta Z/Z$ is generated for each corner for which a temporal match is found.

Stereo matching is carried out, using corner strength and $\Delta Z/Z$ as matching attributes, but without the use of the disparity prediction mechanism. The criteria for accepting a stereo match are demanding, in order to minimise bad matches (even if this means that good matches are also missed). The accepted stereo matches form the seed set, and this is used to generate the quantity $B/\Delta Z$. Given $B/\Delta Z$, it is possible to carry out stereo matching with the disparity prediction mechanism on all unmatched corners for which $\Delta Z/Z$ is known.

Frame 3. Temporal matching is carried out between frame 2 and frame 3. The quantity $\Delta Z/Z$ is generated for each corner for which a temporal match is found.

Stereo matches are cascaded forward from frame 2 using the method described in section 3.1 (b). These matches form the seed set, and this is used to generate the quantity $B/\Delta Z$. Given $B/\Delta Z$, it is possible to carry out stereo matching with the disparity prediction mechanism on all unmatched corners for which $\Delta Z/Z$ is known.

Frame 4...Frame n. Processing for frame 4 and all subsequent frames is the same as for frame 3.

A special check ensures that if only a small number of stereo matches are cascaded forward to a new frame, then processing is carried out as at frame 2.

4. RESULTS

This section describes the results of processing an image sequence. The sequence was taken during straight translational motion at a speed of about 30 cm/s with image capture at 8Hz. The scene was an indoor environment with depth variations up to about 15m. Processing was carried out offline.

The initialisation process described in section 1 required about 15 frames of the sequence. On completion, the main processing was begun anew at the first frame.

A typical stereo pair is shown in figure 5. Table 1 shows data for the four stereo matches which are marked by lines in figure 5. The column headed **Error in (θ, ϕ) of predicted match** shows the difference (in radians) between the co-ordinates of the predicted and true match for the corner (the true match was found by human in-

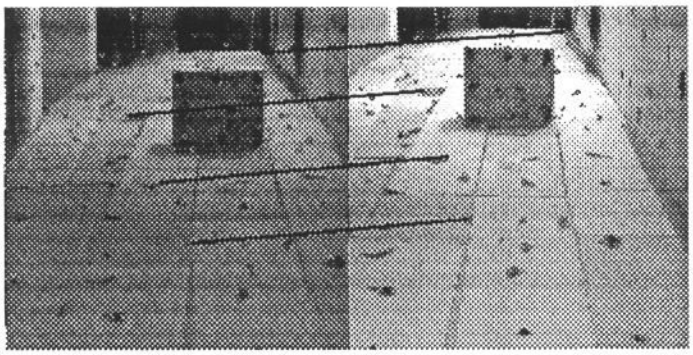

Figure 5: *A stereo pair with example correspondences found during the processing. Artificial 'texture' has been placed in the scene to ensure an abundance of corners. The difference in brightness between the images is due to the different characteristics of the two cameras.*

Corner	Position (θ, ϕ)	$\Delta Z/Z$	Error in (θ, ϕ) of predicted match
1 left	0.1966, 1.9425	0.0166	-0.0036, -0.0532
1 right	0.2070, 2.0751	0.0163	0.0027, 0.0551
2 left	0.1712, 2.2344	0.0155	-0.0009, -0.0234
2 right	0.1823, 2.3279	0.0154	0.0003, 0.0228
3 left	0.1637, 2.6086	0.0125	0.0044, -0.0013
3 right	0.1705, 2.6479	0.0112	-0.0036, 0.0033
4 left	0.0380, 1.3535	0.0097	0, 0.2247
4 right	0.0372, 1.4020	0.0067	0.0011, -0.1367

Table 1: *Corners belonging to the example correspondences shown in figure 5. The bottom stereo match in figure 5 is described by the 'corner 1 left'/'corner 1 right' entries in the table, the match above that to the next table entry, and so on.*

spection). Typically, results are found to degrade with increasing depth and increasing proximity to the FOE. This is because both of the latter conditions are associated with small optical flow vectors, and hence cause greater susceptibility to error.

5. FUTURE WORK

• The current system works for straight translational motion only, and its extension to handle rotation will be investigated.

• When using disparity prediction in the current system, the search for a corner's stereo match is confined to an area of fixed variation around the predicted co-ordinates. A more sophisticated approach would compute the uncertainty in the prediction, and use this to define the search area.

• An alternative approach to the generation of the seed stereo matches is to find them at coarse image resolution, and to use the coarse stereo matches to predict disparities for corners at normal resolution.

• An alternative approach to the stereo matching is to use an initial set of seed stereo matches for disparity prediction on unmatched corners, then to find new stereo matches and add them to the seed set, and to repeat the whole process. This could be carried out over a number of iterations.

6. CONCLUSION

Solving the correspondence problem for a single stereo pair is a difficult task. This paper has described a system in which the motion information available in a stereo sequence is used to ameliorate the process. An important motivation for the chosen approach is the desire to avoid the use of explicit 3D information. A key issue has been the use of an internal parameter ($B/\Delta Z$) which is transient and varies with vehicle motion, but which has global significance and usefulness at any particular time.

Acknowledgements

Thanks to Andrew Zisserman and Guy Scott for their help, and to Dr Chris Harris for access to the DROID system.

References

[1] **D. Marr.** *Vision.* Freeman, San Francisco, 1982.

[2] **C.G. Harris.** "Determination of ego - motion from matched points." In *3rd Alvey Vision Conference*, pages 189–192, 1987.

[3] **C.G. Harris and J.M. Pike.** "3D positional integration from image sequences." In *3rd Alvey Vision Conference*, pages 233–236, 1987.

[4] **D. Charnley and R.J. Blissett.** "Surface reconstruction from outdoor image sequences." In *4th Alvey Vision Conference*, pages 153–158, 1988.

[5] **E. Sparks and M. Stephens.** "Integration of stereo and motion." In *British Machine Vision Conference*, 1990.

[6] **R.A. Brooks, A.M. Flynn, and T. Marill.** "Self calibration of motion and stereo vision for mobile robots." In *4th International Symposium on Robotics Research*, pages 277–286, 1988.

[7] **C.G. Harris and M. Stephens.** "A combined corner and edge detector." In *4th Alvey Vision Conference*, pages 147–151, 1988.

[8] **S. Ganapathy.** "Decomposition of transformation matrices for robot vision." In *Proc. of IEEE Conference on Robotics*, pages 130–139, 1984.

[9] **G.L. Scott.** *Local and Global Interpretation of Moving Images.* Pitman Publishing, London, 1988.