

Planar Region Detection and Motion Recovery

D. Sinclair, A. Blake, S. Smith and C. Rothwell

Robotics Research Group
Department of Engineering Science
Oxford University OX1 3PJ

Abstract

This paper presents a means of segmenting planar regions from two views of a scene using point correspondences. The initial selection of groups of coplanar points is performed on the basis of conservation of two five point projective invariants (groups for which this invariant is conserved are assumed to be coplanar). The correspondences for four of the five points are used to define a projectivity which is used to predict the change in position of other points assuming they lie on the same plane as the original four. A distance threshold between actual and predicted position is used to find extended planar regions. If two distinct planar regions can be found then a novel motion direction estimator suggests itself.

1 Introduction

Classically the structure from motion problem has been seen as obtaining the distance from the camera optical center to points in the world from their motions in a sequence of images. This approach stems from the fact that if the position of a point within the field of view is known then the only piece of information left to recover is the point's 'depth'. Individual depth estimates are noisy and sparse. This makes looking for qualitative scene structure difficult or impossible.

The problems of recovering camera motion and scene structure are inextricably linked [6], [4], [7]. If the motion of the camera (or equivalently stereo geometry) is accurately known then it is straight forward to recover scene depth, once the correspondence problem has been solved. Alternatively if the depths of scene points are known then motion direction recovery is easy. Longuet-Higgins [5] derived a scheme for recovering camera motion if the scene was planar or points within it were coplanar. The solution suffered from a two fold ambiguity and was not demonstrated on real data. No a priori method was presented for determining whether or not points were coplanar. A mathematically equivalent scheme was independently derived by Tsai [14] with equivalence being shown by Faugeras in [2]. Most general methods, though, return the depths of a series of points. By themselves these depths convey little useful qualitative information about the scene. Authors have performed Delaunay triangulation on the points to provide a kind of surface. The surface does not reflect the underlying surface of the scene and for the most part any depth estimate derived from it will be wrong.

It is well known that the mapping between two projected views of a plane is completely specified by a 3 by 3 transformation matrix [11], [12]. The group of these

matrices is called the projective group $PGL(2)$. Members of this group have 8 rather than 9 degrees of freedom. This means that a projectivity is completely specified by four point correspondences if no three of the projected points are co-linear. If a fifth point correspondence is available then 2 projectively invariant quantities are defined. If these two quantities are not conserved between the two views then the five points do not lie on a plane. Conic invariants have been used in a similar manner in [3].

In this paper a simple test for planarity of sets of five points is derived. Implicit in the test is an estimate of the variance in the position of points in an image. This variance is used to provide an estimate of the variance in the values of two 'projective invariants'. Conservation of the values of the two invariants is taken to mean that a set of five points lies in a plane. The accepted difference in the values of the invariant between views is the linearised variance of the invariants themselves.

The above process provides groups of five points lying on possibly different planes. Four of the five points are used to generate the projectivity associated with the transformation of any point on the same plane as the set of five points between views. This transformation allows the new position of any point on the plane to be predicted. The predicted position and the actual new position are compared. A simple distance threshold, based on an assumed variance in the position of the point undergoing prediction, is used to decide whether or not a given point lies on the plane of the five points. This allows planar regions to be grown.

If two or more distinct planar regions have been found in an image then the two projectivities associated with them may be used to recover unambiguously the motion direction and the line of intersection of the planes in the image. Once the motion direction has been recovered it is then possible to solve for the camera's rotation and the normal of the planes. This does however require a calibrated camera.

Section 2 defines the notation to be used in this paper. Section 3 details the projective invariants used and the method of selecting groups of five points. Section 5 shows how the projectivity associated with the motion of one of the planes may be derived and section 6 how it may be used to look for other points on the same plane. Section 7 covers the recovery of camera motion.

2 Notation

In this chapter the following notation is adopted, \mathbf{x} a vector in the projective plane, \mathbf{x}_i the i th vector of a set of vectors, \mathbf{x}' a transformed vector, \mathcal{P} any invertible 3 by 3 projective transformation matrix. In projective space the following identification is made;

$$\mathbf{x} = \lambda \mathbf{x}, \quad (1)$$

where λ is any constant, hence 3 dimensional vectors in the projective plane only have 2 degrees of freedom. \mathbf{x}'_i is therefore given by

$$\mathbf{x}'_i = \lambda_i \mathcal{P} \mathbf{x}_i. \quad (2)$$

The vectors \mathbf{x} are then the homogeneous co-ordinates of the image positions of corners. The matrix whose columns are the vectors $\mathbf{x}_i, \mathbf{x}_j$ and \mathbf{x}_k is written

$$M_{ijk} = (\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k). \quad (3)$$

3 The Two Five Point Planar Invariants

Projective invariants are quantities which do not change under projective transformations. A full review of the uses of invariants is given in [10]. There are two convenient invariants that may be defined for groups of five points. They correspond to the two degrees of freedom of the projective position of the fifth point with respect to the first four. The two invariants may conveniently be written as the ratios of determinants of matrices of the form M_{ijk} .

$$I_1 = \frac{|M_{124}| |M_{135}|}{|M_{134}| |M_{125}|} \quad (4)$$

and

$$I_2 = \frac{|M_{241}| |M_{235}|}{|M_{234}| |M_{215}|}. \quad (5)$$

these two quantities may be seen to be conserved under a projective transformation if \mathbf{x}' is substituted for \mathbf{x} ,

$$\frac{|M'_{124}| |M'_{135}|}{|M'_{134}| |M'_{125}|} = \frac{|\lambda_1 \mathcal{P}_{\mathbf{x}_1}, \lambda_2 \mathcal{P}_{\mathbf{x}_2}, \lambda_4 \mathcal{P}_{\mathbf{x}_4}| |\lambda_1 \mathcal{P}_{\mathbf{x}_1}, \lambda_3 \mathcal{P}_{\mathbf{x}_3}, \lambda_5 \mathcal{P}_{\mathbf{x}_5}|}{|\lambda_1 \mathcal{P}_{\mathbf{x}_1}, \lambda_3 \mathcal{P}_{\mathbf{x}_3}, \lambda_4 \mathcal{P}_{\mathbf{x}_4}| |\lambda_1 \mathcal{P}_{\mathbf{x}_1}, \lambda_2 \mathcal{P}_{\mathbf{x}_2}, \lambda_5 \mathcal{P}_{\mathbf{x}_5}|}, \quad (6)$$

which gives,

$$\frac{|M'_{124}| |M'_{135}|}{|M'_{134}| |M'_{125}|} = \frac{\lambda_1^2 \lambda_2 \lambda_3 \lambda_4 \lambda_5 |\mathcal{P}| |M_{124}| |M_{135}|}{\lambda_1^2 \lambda_2 \lambda_3 \lambda_4 \lambda_5 |\mathcal{P}| |M_{134}| |M_{125}|}. \quad (7)$$

From a combinatorial point of view it might be thought that there were 10 independent invariants. The determinant $|M_{ijk}|$ does not change under cyclic permutations of the indices and only changes sign under acyclic permutations. However there are only two. Generally four points have eight degrees of freedom all of which are required to define the transformation into the canonical frame. The fifth point has two degrees of freedom which form the values of the two invariants [10].

In selecting the groups of five points it is important that they be sufficiently far apart that image measurement noise does not swamp the invariant. To this end the nearest four points outside a circle of radius 25 pixels are selected for the invariant. Other selection strategies are under investigation as the invariants degenerate when any three of the five points are collinear.

4 Covariance Matrices of the 2 Invariants

There is a measurement uncertainty associated with the estimated position of corner features in any image. These errors are assumed to be normally distributed with variances σ_x and σ_y which are additionally assumed to be isotropic and equal to σ . This value will vary with the type of corner detector used. The linearised variance of the invariants may be computed as follows,

$$\delta I_1^2 = \sum_{i=1}^5 \left(\frac{\partial I_1}{\partial \mathbf{x}_i} \cdot d\mathbf{x}_i \right) \left(\frac{\partial I_1}{\partial \mathbf{x}_i} \cdot d\mathbf{x}_i \right) \quad (8)$$

or as noise is uncorrelated [1] (by the law of propagation of error),

$$\text{Var}(I_1) = \sum_{i=1}^5 \frac{\partial I_1^T}{\partial \mathbf{x}_i} \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \frac{\partial I_1}{\partial \mathbf{x}_i}. \quad (9)$$

The vectors \mathbf{x}_i have components,

$$\mathbf{x}_i = \begin{pmatrix} x_i \\ y_i \\ f \end{pmatrix}, \quad (10)$$

where f is the focal length of the camera. The derivatives of I_1 are then given by,

$$\frac{\partial I_1}{\partial x_1} = I_1 \left(\frac{|(\mathbf{e}_1, \mathbf{x}_2, \mathbf{x}_4)|}{|(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_4)|} + \frac{|(\mathbf{e}_1, \mathbf{x}_3, \mathbf{x}_5)|}{|(\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_5)|} - \frac{|(\mathbf{e}_1, \mathbf{x}_3, \mathbf{x}_4)|}{|(\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4)|} - \frac{|(\mathbf{e}_1, \mathbf{x}_2, \mathbf{x}_5)|}{|(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5)|} \right) \quad (11)$$

where,

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad (12)$$

$$\frac{\partial I_1}{\partial x_2} = I_1 \left(\frac{|(\mathbf{x}_1, \mathbf{e}_1, \mathbf{x}_4)|}{|(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_4)|} - \frac{|(\mathbf{x}_1, \mathbf{e}_1, \mathbf{x}_5)|}{|(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5)|} \right) \quad (13)$$

etc.

$$\frac{\partial I_1}{\partial y_1} = I_1 \left(\frac{|(\mathbf{e}_2, \mathbf{x}_2, \mathbf{x}_4)|}{|(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_4)|} + \frac{|(\mathbf{e}_2, \mathbf{x}_3, \mathbf{x}_5)|}{|(\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_5)|} - \frac{|(\mathbf{e}_2, \mathbf{x}_3, \mathbf{x}_4)|}{|(\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4)|} - \frac{|(\mathbf{e}_2, \mathbf{x}_2, \mathbf{x}_5)|}{|(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5)|} \right) \quad (14)$$

where,

$$\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad (15)$$

$$\frac{\partial I_1}{\partial y_2} = I_1 \left(\frac{|(\mathbf{x}_1, \mathbf{e}_2, \mathbf{x}_4)|}{|(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_4)|} - \frac{|(\mathbf{x}_1, \mathbf{e}_2, \mathbf{x}_5)|}{|(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5)|} \right), \quad (16)$$

and so forth. The same analysis may be performed in order to derive a similar expression for the derivatives of I_2 . These expressions permit the computation of the variance of the two invariants. To test whether it is possible that five points visible in two views lie on a plane the difference in each of the two invariants must obey the condition

$$|\mathbf{I}' - \mathbf{I}| < 2\sqrt{\text{Var}(I)}. \quad (17)$$

Figure 1 shows an image from a motion sequence with motion vectors from tracked corners superimposed on it. Figure 2 shows all the starting positions of the flow vectors as square boxes. Two of the groups of five points found using the invariant planarity test are marked by darker symbols. From this sequence a total of 13 groups of five planar points were found. In this case the value of σ was taken to be 0.2. Points at extreme distance will tend to behave as if they were on a plane as well as those on flat surfaces.

5 Finding the Projectivity between four Points in two Views

If four coplanar points are available in two views then a projectivity may be defined which will transform the first set of points into the second (provided no three of the



Figure 1: An image from a motion sequence with flow vectors of tracked corners superimposed on it.

points are collinear). The same projectivity will predict the position of any point on the plane in the first image in the second image [11].

If \mathbf{x}_i are the initial four points and \mathbf{x}'_i the transformed four points then,

$$\mathbf{x}'_i = \lambda_i \mathcal{P} \mathbf{x}_i. \quad (18)$$

The easiest way of finding \mathcal{P} is first to transform both \mathbf{x}_i and \mathbf{x}'_i to the canonical frame \mathbf{e}_i where \mathbf{e}_1 and \mathbf{e}_2 are as before and

$$\mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad (19)$$

$$\mathbf{e}_4 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}. \quad (20)$$

$$\mathbf{x}_i = \lambda_i \mathcal{M}_1 \mathbf{e}_i, \quad (21)$$

where using i from 1 to 3 gives,

$$\mathcal{M}_1 = (\alpha \mathbf{x}_1, \beta \mathbf{x}_2, \gamma \mathbf{x}_3) \quad (22)$$

and using $i = 4$ gives

$$\mathbf{x}_4 = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \quad (23)$$

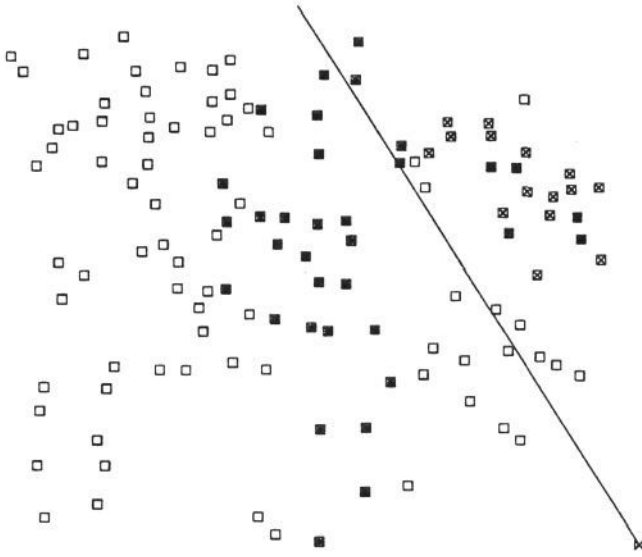


Figure 2: The starting positions of the flow vectors in figure 1 are marked. Two of the groups of five planar points are marked by darker boxes. The boxes marked with stars were found to be on the same plane as the left hand group of five. The boxes marked with crosses were found to be on the same plane as the right hand group of five. The line represents the line of intersection of the two planes.

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)^{-1} \mathbf{x}_4, \quad (24)$$

this may be solved for α , β , γ by inverting the matrix whose columns are \mathbf{x}_1 , \mathbf{x}_2 , \mathbf{x}_3 . The corresponding matrix \mathcal{M}_2 that maps \mathbf{x}'_i to \mathbf{e}_i must then be found and then,

$$\mathcal{P} = \mathcal{M}_2 \mathcal{M}_1^{-1}. \quad (25)$$

Three matrix inversions must hence be performed to find \mathcal{P} . Analytical expressions are available for the inverse of 3 by 3 matrices therefore, with the aid of Mathematica the variance of \mathcal{P} with respect to either set of four points can be determined.

Alternative methods of calculating this projectivity have been explored. A pseudo inverse technique [9] [13] using all five of the points was tried. A minimum eigenvalue technique similar to that used by Kanatani [8] was also used. The most 'useful' technique in terms of predicting the motion of planar points was the four point method detailed above.

6 Finding Additional Coplanar Points

Once it has been established that four points lie on a plane then the projectivity that maps the four points to their images may be used to predict the new image

position of any point on the same plane [11]. If the projectivity between the two frames is \mathcal{P} then,

$$\hat{\mathbf{x}} = \lambda \mathcal{P} \mathbf{x}, \quad (26)$$

with λ constrained so that $\hat{\mathbf{x}} \cdot \mathbf{e}_3 = f$. If a rapid answer is required a simple distance threshold between predicted and actual image position may be used. We derive here an expression for the x and y variances of the predicted position $\hat{\mathbf{x}}$ of a point \mathbf{x} . The variance of the x component of the predicted position is given by,

$$\delta \hat{x}^2 = \left(\frac{\partial \hat{x}}{\partial x} \right)^2 dx^2 + \frac{\partial \hat{x}}{\partial y} dy \frac{\partial \hat{x}}{\partial x} dx + \left(\frac{\partial \hat{x}}{\partial y} \right)^2 dy^2 \quad (27)$$

or as dx and dy are uncorrelated,

$$\text{Var}(\hat{x}) = \begin{pmatrix} \frac{\partial \hat{x}}{\partial x} & \frac{\partial \hat{x}}{\partial y} \end{pmatrix} \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \begin{pmatrix} \frac{\partial \hat{x}}{\partial x} \\ \frac{\partial \hat{x}}{\partial y} \end{pmatrix} \quad (28)$$

where

$$\frac{\partial \hat{x}}{\partial x} = \frac{f}{\mathcal{P}_{31}x + \mathcal{P}_{32}y + \mathcal{P}_{33}f} \mathcal{P}_{11} - (\mathcal{P}_{11}x + \mathcal{P}_{12}y + \mathcal{P}_{13}f) \frac{f \mathcal{P}_{31}}{(\mathcal{P}_{31}x + \mathcal{P}_{32}y + \mathcal{P}_{33}f)^2} \quad (29)$$

$$\frac{\partial \hat{x}}{\partial y} = \frac{f}{\mathcal{P}_{31}x + \mathcal{P}_{32}y + \mathcal{P}_{33}f} \mathcal{P}_{12} - (\mathcal{P}_{11}x + \mathcal{P}_{12}y + \mathcal{P}_{13}f) \frac{f \mathcal{P}_{32}}{(\mathcal{P}_{31}x + \mathcal{P}_{32}y + \mathcal{P}_{33}f)^2} \quad (30)$$

and σ is as before. The corresponding variance in \hat{y} is derived analogously. Points are accepted as being on the plane associated with the projectivity if the difference between actual and predicted position in the second frame is less than two standard deviations in the x or y directions. That is,

$$|\hat{x} - x'| \leq 2\sqrt{\text{Var}(\hat{x})} \quad (31)$$

and

$$|\hat{y} - y'| \leq 2\sqrt{\text{Var}(\hat{y})}. \quad (32)$$

The permissible difference in position may be anisotropic. Figure 2 shows the additional groups of points found to be on the same plane as the initial two groups of five.

It would have been possible to insert successive points into the five point planar invariant keeping four of them constant and using the analysis of section 4. This however would take more time and as will be seen in the following sections the projectivity associated with a plane in motion has additional uses. For simply detecting planar regions no camera calibration at all is required. It is only if information like plane normal or motion direction are required that calibration is necessary.

7 Finding the Motion Direction

As is well known if four planar point correspondences are available then the motion direction may be recovered up to a two fold ambiguity [5] [14]. If two non-parallel planes are visible then this ambiguity may be resolved. A more direct method of recovering an observer's motion direction is presented here. Let the two projectivities

associated with the motion of points on the two planes be \mathcal{P}_1 and \mathcal{P}_2 . Projectivity \mathcal{P}_1 may be used to predict the new position of any point x ,

$$\hat{x} = \lambda \mathcal{P}_1 x \quad (33)$$

and likewise \mathcal{P}_2 may also be used to predict a new position of any point,

$$\hat{x}' = \lambda \mathcal{P}_2 x. \quad (34)$$

The two predicted positions will only coincide for certain points. The predictions will agree for the points on the line of intersection of the two planes and the motion direction, (the two predictions will agree on the epipole). Why this is so is shown in figure 3. The line of intersection (in the image) of the two planes and the motion

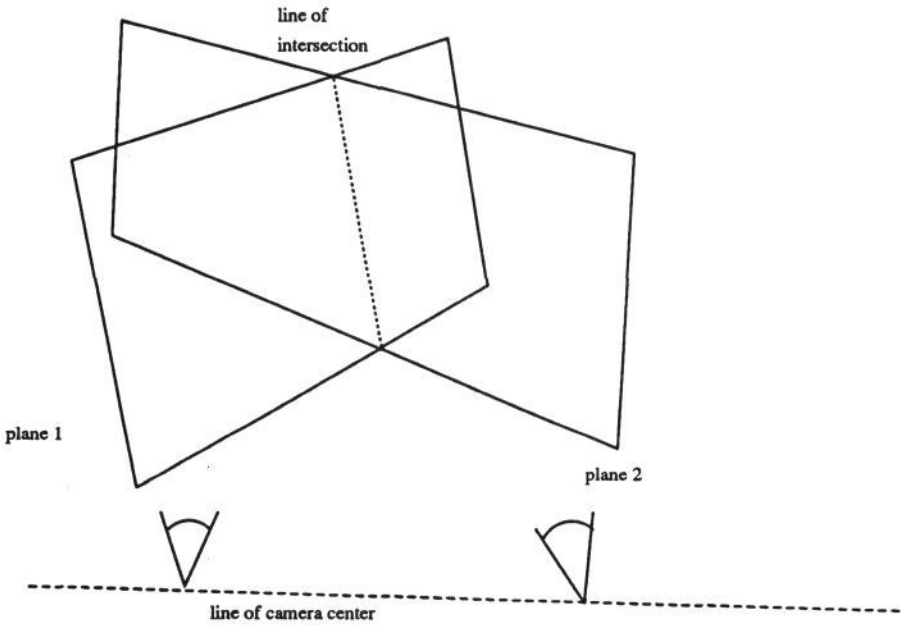


Figure 3: This figure show two camera positions relative to two planes. The line of intersection is preserved between views of the planes. The only other image direction which both projectivities predict go to the same position is the motion direction of the camera.

direction may be found by solving the following eigenvector equation,

$$\mathcal{P}_1 x = \lambda \mathcal{P}_2 x. \quad (35)$$

If the camera undergoes any rotation at all the real eigenvalue will have eigen direction parallel to the motion direction and the degenerate pair will have eigen directions that span the line of intersection of the two planes. The method will fail when the system has three degenerate roots. This corresponds to the case when the motion direction points towards a point on the line of intersection of the two planes. Figure 2 shows the line of intersection of the two planes in figure 1.

8 Possible Extension

The central idea in the paper is that the transformation, associated with a change in viewing position, of a group of planar points may be written as 3 by 3 matrix. This fact permits the prediction of the new position of any point on the plane. The following section suggests a way of making use of this fact.

8.1 Using Lines in the Invariant

In a man made world planar regions are often bounded by lines (books, roads, sides of buses, houses, etc). If a line has been found in an image from a motion sequence then is it possible to use points lying on one side of the line to construct invariants, conservation of which will imply that the lines and points are coplanar. Line are the dual of points in $P(2)$ and hence a minimum of one line and four points are required to construct two invariants.

A line l through a set of points x is defined by the equation,

$$l \cdot x = 0 \quad (36)$$

this means that l transforms according as,

$$l' = \lambda_l \mathcal{P}^{-T} l. \quad (37)$$

Noting that,

$$l' \cdot x'_i = \lambda_l \lambda_i l \cdot x_i, \quad (38)$$

where x_i are points not on the line l , invariants may be constructed of the form,

$$I = \frac{|M_{123}| l \cdot x_4}{|M_{124}| l \cdot x_3}. \quad (39)$$

The covariance of these invariants may be derived in a similar manner to section 4. These may be used to look for groups of four points coplanar with the line. This remains to be implemented.

9 Conclusions

The mechanism proposed provides a robust and rapid means of isolating sets of planar points from two views of a scene. If two planar regions are available then it is straightforward to recover the camera's motion direction. The direction of the plane normal may be isolated from the projectivity associated with a particular plane and the motion direction used to resolve the two fold ambiguity. If estimates of plane normal and motion direction have been obtained then the camera's rotation may be estimated.

10 Acknowledgements

We acknowledge discussions with Professor J. M. Brady, Dr A. Zisserman, Dr. D. Murray, P. Beardsley and R. Cipolla of the Robotics Research Group. We are grateful for support from the SERC, the EEC Esprit program and DRA RARDE Chertsey.

References

- [1] Coehlo C., Heller A., Mundy J., and Forsyth D. An experimental evaluation of projective invariants. In *DARPA Image Understanding Workshop*, 273–294, 1991.
- [2] O.D. Faugeras and Maybank S.J. Motion from point matches: Multiplicity of solutions. *Int. Journal of Computer Vision*, 4:225–246, 1988.
- [3] D.A. Forsyth, J.L. Mundy, A.P. Zisserman, and C.M. Brown. Projectively invariant representations using implicit algebraic curves. In *Proc. 1st European Conf. on Computer Vision*, 427–436. Springer-Verlag, 1990.
- [4] C.G. Harris. Determination of ego - motion from matched points. In *3rd Alvey Vision Conference*, 189–192, 1987.
- [5] Longuet-Higgins H.C. The reconstruction of a plane surface from two projections. In *Proc. R. Soc. Lond.*, 399–410, 1986.
- [6] B.K.P. Horn and B.G. Schunk. Determining optical flow. *Artificial Intelligence*, vol.17:185–203, 1981.
- [7] B.K.P. Horn and E.J. Weldon. Direct methods for recovering motion. *Int. Journal of Computer Vision*, vol.2:51–76, 1988.
- [8] K. Kanatani. *Geometric Computation for Machine Vision*, volume 1st Edition. MIT Press, 1991.
- [9] Carlsson S. Projectively invariant decomposition of planar shapes. In J.L. Mundy and A. Zisserman, editors, *Geometric Invariance in Computer Vision*, 267–276. MIT Press, 1992.
- [10] J.L. S. Mundy and A. Zisserman. *Geometric Invariance in Computer Vision*, volume First Edition. MIT Press, 1992.
- [11] J.G. Semple and G.T. Kneebone. *Algebraic projective geometry*. Oxford University Press, 1952.
- [12] C.E. Springer. *Geometry and Analysis of Projective Spaces*, volume 1. Freeman, 1964.
- [13] G. Strang. *Linear Algebra and its Applications*, volume I. Academic Press, 1980.
- [14] R.Y. Tsai and T.S. Huang. Estimating three-dimensional motion parameters of a rigid planar patch. *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol.ASSP-29,no.6:1147–1152, 1981.