Robust Detection of Degenerate Configurations for the Fundamental Matrix

P H S Torr, A Zisserman and S J Maybank*
Robotics Research Group
Department of Engineering Science
Oxford University, Parks Road, Oxford, OX1 3PJ, UK.

Abstract

*New methods are reported for the detection of multiple solutions (*degeneracy*) when estimating the fundamental matrix, with specific emphasis on robustness in the presence of data contamination (outliers). The fundamental matrix can be used as a first step in the recovery of structure from motion. If the set of correspondences is degenerate then this structure cannot be accurately recovered and many solutions will explain the data equally well. It is essential that we are alerted to such eventualities. However, current feature matchers are very prone to mismatching, giving a high rate of contamination within the data. Such contamination can make a degenerate data set appear non-degenerate, thus the need for robust methods becomes apparent. This paper presents such methods with a particular emphasis on providing a method that will work on real imagery and with an automated (non-perfect) feature detector and matcher. It is demonstrated that proper modelling of degeneracy in the presence of outliers enables the detection of outliers which would otherwise be missed. Results using real image sequences are presented. All processing, point matching, degeneracy detection and outlier detection is automatic.*

1 Introduction

Robotic vision has its basis in geometric modelling of the world. Vision algorithms attempt to estimate these geometric models from perceived data. The data are said to be degenerate if they are insufficient to determine a unique solution. It is important to know when degeneracy has occurred. This has been achieved in the past for data with no outliers by the use of covariance matrices to measure uncertainty; unfortunately such methods break down when the data contains any significant level of contamination. In this paper a statistically based estimator for the *fundamental matrix* [**F**] [Fau92, Har92] is presented that robustly detects the presence of outliers and degeneracy. The fundamental matrix encapsulates the epipolar geometry. It contains all the information on camera motion and intrinsic parameters available from image feature correspondences alone, and is often used as a first stage in structure estimation. Frequently it is of interest to know something about the sensitivity and accuracy of the estimated fundamental matrix. The ideal or Platonic fundamental matrix is determined by the intrinsic (camera) and extrinsic (motion) parameters alone. It does not depend on the depths of the scene points (structure). The estimated fundamental matrix is calculated from measured image correspondences however, not ideal ones, and its accuracy and sensitivity *do* vary with structure. When the estimated fundamental matrix alters disproportionately as a result of small changes in the data then the data are termed degenerate. This is usually equivalent to many fundamental matrices fitting the data equally well. This has obvious disadvantages from the point of view of reconstruction and segmentation.

A simple example illustrates the effects of outliers in the presence of degeneracy. Figures 1 (a), (b) and (c) show three cases of line-fitting to two dimensional data sets. A line is considered a good fit if the points all lie within the threshold indicated by the dotted lines. This threshold is typically related to the variance of the features. Figure 1 (a) shows a set which might be considered non-degenerate, and for which a line model is appropriate. Figure 1 (b) demonstrates degenerate data, in that there are an infinite number of acceptable fitted lines, all passing through the centroid of the data. A noise model is essential if this type of degeneracy is to be detected; in the absence of a noise model there is no criterion for determining whether a fit is good and hence no criterion for determining whether there are multiple solutions (good fits). Note that in Figure 1

*Financial support was provided by EU Esprit Project 6448 'VIVA' and UK EPSRC grant GR/J65372. The authors would like to thank David Murray for comments on the draft.

0-8186-7042-8/95 $4.00 © 1995 IEEE

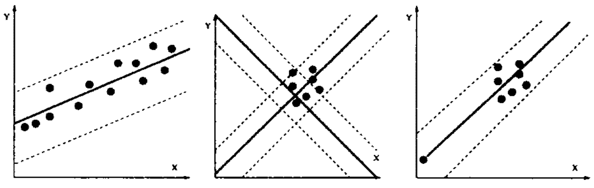

Figure 1: *Line fitting to 2D data sets. In all cases the statistical threshold for points to be consistent with the line are indicated by dashed lines. Lines that fit all the data are shown. (a) A non-degenerate data set, with no ambiguity in determining the best line fit. (b) A degenerate data set. (c) A single outlier renders a degenerate data set apparently non-degenerate.*

(a) if the noise is very high relative to the dispersion of the points then this might indeed be a degenerate set. The need for methods which can flag degeneracy in the presence of outliers is demonstrated by Figure 1 (c) where even one outlier can mask the degeneracy. It might be thought that a point rather than a line would better model the data in (b) or (c).

It is clear then that the existence of degeneracy depends upon three elements: the geometric model of the world, the error criterion and the statistical model. In the previous example, the geometric model is a line, the error criteria are the perpendicular distances of the points to the line, and the statistical model is zero mean isotropic Gaussian noise with a standard deviation less than the average point separation. It is also clear that the detection of outliers and the detection of degeneracy are inextricably linked.

The structure of the paper is as follows: In Section 2 previous work on robust estimators is reviewed. The most robust estimators are random sampling type estimators. A deficiency of this class of estimators is their disregard for the issues relating to degeneracy. Little successful work has appeared concerning this problem to date, which is surprising given the detrimental effect that ambiguous results may have on a working system. Section 3 sets out a random sampling estimator, acronym PLUNDER[1], that detects degeneracy even in the presence of outliers. Results are presented in Section 4.

2 Robust Fitting for [F]

Suppose that the viewed features arise from a 3D object which has undergone a rotation and non-zero translation. After the motion, the set of homogeneous image points $\{\mathbf{x}_i\}, i = 1, \ldots n$, is transformed to the set $\{\mathbf{x}_i'\}$, where $\mathbf{x}_i = (x_i, y_i, \zeta)^\top$, and $\mathbf{x}_i' = (x_i', y_i', \zeta)^\top$. The two sets of features are related by $\mathbf{x}_i'^\top[\mathbf{F}]\mathbf{x}_i = 0$ where $[\mathbf{F}]$ is the rank 2, 3×3 fundamental matrix [Fau92, Har92]. When estimating $[\mathbf{F}]$, outliers are caused by feature mismatches, which occur frequently in most automated applications. In least squares fitting outliers can have a severe effect on the final result, making it essential that a robust method is used for the estimation.

In Torr [Tor95] a comprehensive survey of robust estimators is reported, with comparisons of the results made on large scale synthetic tests as well as on real imagery. A highly robust example of a fitting algorithm is the random sample consensus paradigm (RANSAC) [FB81], which was used by Torr and Murray [TM93, TBM94] to estimate [F]. Given that a large proportion the data may be useless the approach is the opposite to conventional smoothing techniques. Rather than using as much data as is possible to obtain an initial solution and then attempting to identify outliers, as small a subset of the data as is feasible to estimate the parameters is used (e.g. two point subsets for a line, seven correspondences for a fundamental matrix), and this process is repeated enough times on different subsets to ensure that there is a 95% chance that one of the subsets will contain only good data points. The best solution is that which maximizes the number of point consistent correspondences. In order to determine whether or not a feature pair is consistent with a given fundamental matrix, the distance to epipolar line of each correspondence in the image is compared to a threshold based on the presumed probability distribution of the error [Tor95]. Once outliers are removed the set of points identified as non-outliers may be combined to give a final least squares solution.

A major deficiency with RANSAC is that there is no mechanism for detecting degeneracy. Consider Figures 2 (a) and (b) of a rotating toy truck, from a sequence generated by Harris [Har90]. Figure (b) has 27 matches superimposed. Figures 2 (c) and (d) show two epipolar geometries that are both consistent with the matches in (b), in that all the distances to epipolar lines of each correspondence lie below the threshold. If RANSAC is run taking the correspondences as input, either solution might result. If degeneracy is ignored then over fitting might occur; in this instance the estimated parameters are highly unstable and acutely influenced by noise and outliers.

[1] In keeping with the original RANSAC acronym, PLUNDER stands for Pick Least UNDEgenerate Randomly.

3 The Degeneracy Algorithm: PLUNDER

How might the detection of degeneracy be incorporated into RANSAC? One approach might be to establish whether there are two or more distinct solutions, as in the example given in the previous section. When trying to determine how much of the data are consistent with two solutions, a new question immediately arises: how different must these two solutions be in order to be statistically distinct. Without knowledge of the true covariance matrix of the parameters this question cannot be answered. An alternative approach is adopted, in which degenerate data are described by a model with fewer parameters. The problem is then to select the most appropriate model to describe the data: degenerate or non-degenerate. To do this the number of features consistent with each model is estimated using RANSAC, and a decision process selects the most appropriate model.

A list of the causes of degeneracy and the reduced models that fit them is given in Table 1. It encompasses all the models used traditionally, including the affine camera model [MZ92], image affinities (as opposed to the affine camera model), and image projectivities. Although features that are consistent with a camera translation are not degenerate with respect to the estimation of the fundamental matrix, it is useful to test explicitly for the existence of pure translation, as only two feature correspondences need be used to instantiate an estimate.

Having described the models to be employed, their error terms and the variance of their error terms must be established. It transpires that such knowledge is essential when attempting to devise an algorithm using multiple models, in order to make a meaningful comparison between models. The error criteria are as follows:

1. For the fundamental matrix and affine fundamental matrix: root mean square of the distances from the epipolar lines to the points.

2. For the image-image projectivities and affinities: root mean square distance of projected features to their predicted locations over the two images.

The standard deviation of the error criterion is a conflation of the standard deviation of the error in feature location, the scene structure, the camera motion, and choice of estimator. Analytic derivation of the standard deviations is intractable, thus empirical means are used based on Monte Carlo simulations [Tor95].

The algorithm comprises two distinct phases: A sampling phase in which the maximum number of correspondences consistent with each of the models is estimated, and a model selection phase in which the most appropriate model is selected to describe the data.

3.1 Sampling Phase

To estimate the best fitting constraint using RANSAC for each model type in Table 1 would be computationally inefficient. The computation time would be needlessly increased by attempting to estimate model parameters on samples that are degenerate (with respect to that model). Instead, for each random sample principal component analysis affords a quick test for degeneracy, if the sample is degenerate then a lower order model is fitted to that sample, as described in [Tor95]. Thus the sample selected at each iteration determines the model that will gain support from that sample. The best fit for each model is stored.

3.2 Model selection phase

At the termination of the random sampling phase of PLUNDER estimates are provided for the largest set consistent with each model: affinity, projectivity, fundamental matrix etc. The size of each of these (largest) sets is termed its cardinality. A mechanism is now necessary to decide which of the models is most descriptive of the data. A naïve approach would select only the model consistent with the most correspondences. Generally, in the presence of noise and outliers, this scheme would always lead to the selection of the non-degenerate model, the fundamental matrix. This can be understood with reference to Figure 1, where the data are two dimensional. The model fitted is a line (non-degenerate) and the degenerate model a point. In Figure 1 (a) the data are non-degenerate hence the number of features consistent with a line is always greater than the number consistent with a point. In Figure 1 (b), with no outliers, the two models have the same features consistent with both. In Figure 1 (c) all the data are degenerate bar one outlying point, and consequently more points are consistent with the non-degenerate than the degenerate model. In each case the size of the largest degenerate sets will typically be smaller than the largest non-degenerate set.

If the data are degenerate with added outliers, then the largest set of correspondences consistent with a fundamental matrix will (usually) include the degenerate set plus some of the outliers. In Figure 2 (a) there are 27 correspondences, out of 36, consistent with an image affinity, and 9 correspondences that are outliers. The random sampling algorithm duly returns 31 correspondences consistent with a fundamental matrix and 27 consistent with an image affinity. To determine which model gives the best description of the data a rough confidence interval is developed on the number of features consistent with the fundamental matrix. Should the number of features consistent with a degenerate model lie within this interval, then the degenerate model is deemed just as good a description. Details of the derivation of this confidence interval are given in [Tor95]. From this a threshold is derived t_s being the 95% lower bound of the this interval. Hence if a degenerate model has more consistent correspondences than t_s it is accepted.

4 Results

In all the examples in this paper, the corners are obtained by using the detector described in [Har88], the matching procedure uses cross correlation in a square search window. Table 2 summarises the sizes of various degenerate sets found for the figures, using the robust algorithm outlined in Section 3. The tests were conducted under the assumption that the standard deviation of the feature locations was (1) $\sigma = 0.7$ and (2) $\sigma = 1.6$. We found that the results are largely unaffected by the choice of σ in this range.

Truck data. In the truck data, shown in Figure 2 (b), the model selected is an affinity requiring only three points to establish; 27 points are consistent with the affinity and 31 with a full fundamental matrix which has $7 - 3 = 4$ more degrees of freedom, thus the data are clearly degenerate. This is in accordance with what we would expect, as this data was original created to demonstrate the existence of degeneracy [Har90].

Gun Fighter data. Figure 3 (a) (b) shows a mobile gun fighter viewed by a tracking camera. The gunfighter is moving to the right as the camera moves in parallel to keep him in the centre of the image. The outliers and inliers are shown in Figure 3 (c) and (d). Features in independent motion on the gun fighter are identified if they are inconsistent with the camera epipolar constraint i.e. background motion. The model accommodating the largest number of features is a translation (the correct model) which is also consistent with $[\mathbf{F}]_A$.

Model house data. Figure 4 (a) (b) shows a scene in which a camera rotates and translates whilst fixating on a model house. The scene is correctly flagged as non-degenerate, as the translational and rotational components of the camera motion were both significant.

5 Conclusion

We have presented a method for the detection of degeneracy in a set of correspondences when estimating the fundamental matrix. The method is highly robust and simultaneously flushes outliers and selects the type of model that best fits the data. The methodology is general and could be applied to many different parameter estimation problems where degeneracy might occur. Within this paper we have discussed the motion models typically used in robotic vision applications and integrated them into a common structure. The extension of PLUNDER to motion segmentation is discussed in [TZM95].

References

[Fau92] O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proc. 2nd European Conference on Computer Vision , LNCS 588*, pages 563–578, 1992.

[FB81] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, vol. 24:381–95, 1981.

[Har88] C. Harris. The droid 3d vision system. Technical Report 72/88/N488U, Plessey Research Roke Manor, 1988.

[Har90] C. Harris. Structure-from-motion under orthographic projection. In O. Faugeras, editor, *Proc. 1st European Conference on Computer Vision , LNCS 427*, pages 118–128. Springer–Verlag, 1990.

[Har92] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In G. Sandini, editor, *Proc. 2nd European Conference on Computer Vision , LNCS 588*, pages 579–87. Springer–Verlag, 1992.

[MZ92] J. Mundy and A. Zisserman. *Geometric Invariance in Computer Vision.* MIT press, 1992.

[TM93] P. H. S. Torr and D. W. Murray. Outlier detection and motion segmentation. In P. S. Schenker, editor, *Sensor Fusion VI*, pages 432–443. SPIE volume 2059, 1993. Boston.

[TBM94] P. H. S. Torr, P. A. Beardsley, and D. W. Murray. Robust vision. In J. Illingworth, editor, *bmvc94*, pages 145–155. BMVA Press, 1994.

[Tor95] P. H. S. Torr. *Outlier Detection and Motion Segmentation.* PhD thesis, University of Oxford, 1995.

[TZM95] P. H. S. Torr, A. Zisserman, and D. W. Murray. Motion clustering using the trilinear constraint over three views. In R. Mohr and C. Wu, editors, *accepted for Europe-China Workshop on Geometrical Modelling and Invariants for Computer Vision*, 1995.

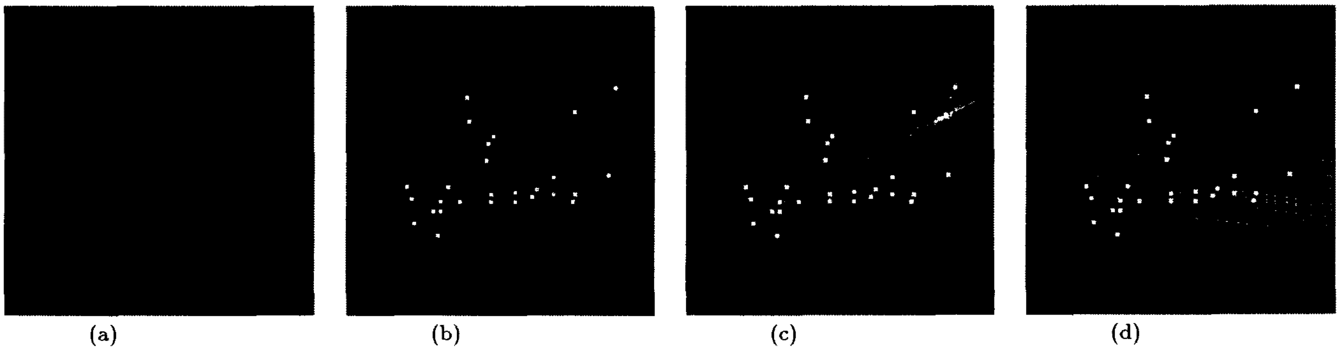

Figure 2: *Rotating toy truck sequence (a) first image (b) second image with correspondences superimposed, and outliers removed. (c) (d) Two epipolar geometries that fit the correspondences shown in (b).*

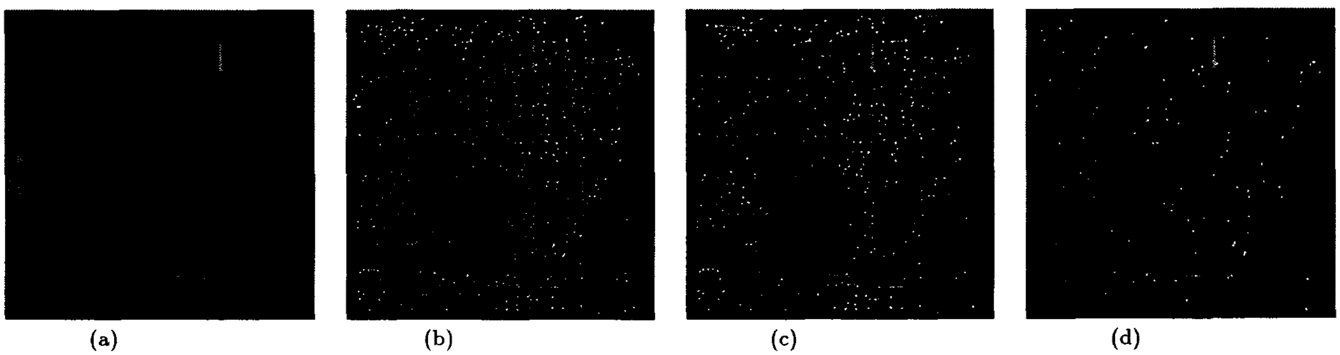

Figure 3: *Two images of the gun fighter sequence. The camera is tracking to the right and the gun fighter raises his gun. (a) first image, (b) second image and matches, (c) inliers, and (d) outliers. Note that the inliers correspond to the background motion, and the outliers to the mismatches. The model selected is a translation.*

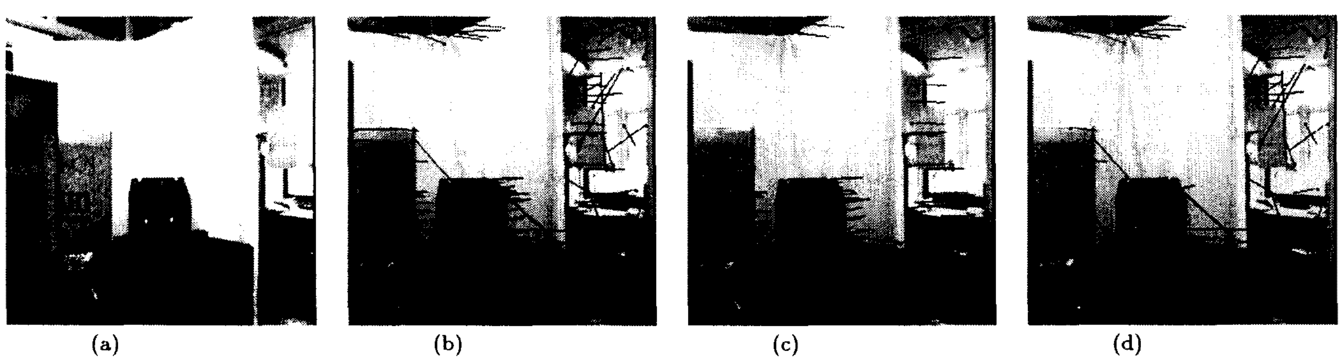

Figure 4: *Indoor sequence, camera translating and rotating to fixate on the house; (a) first, (b) second image with matches superimposed. (c) inliers and (d) outliers. The data are found to be non-degenerate. Again it can be seen that the outliers (d) correspond to mismatches.*

Model	p	Constraint	Parameters	Domain
Fundamental matrix	7	$\mathbf{x}'^{\top}[\mathbf{F}]\mathbf{x} = 0$	$[\mathbf{F}] = \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix}$	Most general.
Affine Fundamental matrix	4	$\mathbf{x}'^{\top}[\mathbf{F}]_A\mathbf{x} = 0$	$[\mathbf{F}]_A = \begin{bmatrix} 0 & 0 & a_1 \\ 0 & 0 & a_2 \\ a_3 & a_4 & a_5 \end{bmatrix}$	All features distant subtending a small field of view. The depth variation may be large, if imaging conditions are near orthographic.
Projective image transformation	4	$\mathbf{x}' = [\mathbf{Q}]\mathbf{x}$	$[\mathbf{Q}] = \begin{bmatrix} q_1 & q_2 & q_3 \\ q_4 & q_5 & q_6 \\ q_7 & q_8 & q_9 \end{bmatrix}$	Features arising from a rotation of the camera about its optic centre; or features arising from a plane in the scene.
Affine image transformation	3	$\mathbf{x}' = [\mathbf{K}]\mathbf{x}$	$[\mathbf{K}] = \begin{bmatrix} k_1 & k_2 & k_3 \\ k_4 & k_5 & k_6 \\ 0 & 0 & 1 \end{bmatrix}$	Features distant and arising from a rotation or a plane
Translation	2	$\mathbf{x}'^{\top}[\mathbf{F}]_T\mathbf{x} = 0$	$[\mathbf{F}]_T = \begin{bmatrix} 0 & g_3 & -g_2 \\ -g_3 & 0 & g_1 \\ g_2 & -g_1 & 0 \end{bmatrix}$	Relative motion between camera and object is pure translation.
Image Translation	1	$\mathbf{x}' = [\mathbf{L}]\mathbf{x}$	$[\mathbf{L}] = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$	Features are distant from the camera, or on a fronto-parallel plane that is translating parallel to the image plane etc.
No motion	0	$\mathbf{x}' = [\mathbf{I}]\mathbf{x}$	$[\mathbf{I}] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	No Motion between camera and object.

Table 1: *A description of the reduced models that are fitted to degenerate sets of correspondences. p is the minimum number of correspondences needed in a sample to estimate the constraint.*

Model	n	Number Consistent with [F]	[F]$_A$	[Q]	[K]	[F]$_T$	[L]	No Motion	t_s
Figure 2	36	31	14	22	<u>27</u>	26	15	2	22
Figure 3	406	305	304	216	125	<u>305</u>	217	16	276
Figure 4	186	<u>115</u>	24	54	53	69	13	0	85

Table 2: *The largest set consistent for each of the reduced models given in Table 1, and the value of t_s, the threshold for degeneracy, if the number of correspondences consistent with any degenerate model exceeds this then the data is deemed degenerate. The underline indicates the model selected.*