Reprint from SSBA'04 proceedings. Minor errors correced.

Blobs in Epipolar Geometry

Per-Erik Forssén and Anders Moe
Computer Vision Laboratory
Department of Electrical Engineering
Linköping University, SE-581 83 Linköping, Sweden

Abstract

Epipolar geometry is the geometry situation of two cameras depicting the same scene. For un-calibrated cameras epipolar geometry is compactly described by the fundamental matrix. Estimation of the fundamental matrix is trivial if we have a set of corresponding points in the two images. Corresponding points are often found using e.g. the Harris interest point detector, but there are several advantages with using richer features instead. In this paper we will use blob features. Blobs are homogeneous regions which are compactly described by their colour, area, centroid and inertia matrix. Using blobs to establish correspondences is fast, and the extra information besides position, allows us to reject false matches more accurately.

1 Introduction

Epipolar geometry is the geometry situation of two cameras depicting the same scene. For a thorough description of epipolar geometry, see [3]. For uncalibrated cameras epipolar geometry is compactly described by the *fundamental matrix* $\mathbf{F}$. A point $\mathbf{x} = \begin{pmatrix} x_1 & x_2 & 1 \end{pmatrix}^T$ in image 1, and the corresponding point $\mathbf{x}' = \begin{pmatrix} x'_1 & x'_2 & 1 \end{pmatrix}^T$ in image 2 are related through the fundamental matrix as

$$\mathbf{x}'^T \mathbf{F} \mathbf{x} = 0\,. \tag{1}$$

If we know $\mathbf{F}$, we can search for points $\mathbf{x}'$, corresponding to $\mathbf{x}$ along its *epipolar line* $\mathbf{l}$

$$\mathbf{x}'^T \mathbf{l} = 0 \quad \text{where} \quad \mathbf{l} = \mathbf{F}\mathbf{x}\,. \tag{2}$$

Thus $\mathbf{F}$ can be used as a constraint on the correspondences. Once we have a set of correspondences, we can use triangulation [2] to compute disparities. With more knowledge about the cameras, we could then use these to compute distances to objects.

Computation of the fundamental matrix is possible using point correspondences alone, provided that not one of the following two *degenerate cases* is present

1. all scene points lie on a plane
2. The camera motion between the two cameras is a pure rotation.

Whenever either of these situations occur, $\mathbf{F}$ is not well defined, and we must instead estimate a *homography* $\mathbf{H}$, that relates corresponding points directly

$$h\mathbf{x}' = \mathbf{H}\mathbf{x} \quad \text{and} \quad g\mathbf{x} = \mathbf{H}^{-1}\mathbf{x}'\,. \tag{3}$$

Most work on epipolar geometry has involved correspondences between points, lines or conics [3], but recently work has started on using richer features, such as *affine invariant features* [4, 5].

2 Blobs

We will make use of blob features extracted using a clustering pyramid built using robust estimation in local image regions [1]. The current implementation processes 360×288 video frames at a rate of 1 sec/frame on a Intel P3 CPU at 697 MHz, and produces relatively robust and repeatable features. Each extracted blob is represented by its colour $\mathbf{p}_k$, area a_k, centroid $\mathbf{m}_k$, and inertia matrix $\mathbf{I}_k$. I.e. each blob is a 4-tuple

$$\mathcal{B}_k = \langle \mathbf{p}_k, a_k, \mathbf{m}_k, \mathbf{I}_k \rangle\ .$$

Since an inertia matrix is symmetric, it has 3 degrees of freedom, and we have a total of $3 + 1 + 2 + 3 = 9$ parameters for each blob.

The blob estimation has two main parameters: a colour distance threshold $d_{\max}$, and a propagation threshold $c_{\min}$ for the clustering pyramid. For all experiments in this paper we have used $d_{\max} = 0.16$ (RGB values in interval $[0, 1]$) and $c_{\min} = 0.5$.

The number of blobs in an image depends heavily on image content, we typically get between 25 and 500 blobs in each image. Figure 1 shows an example where 91 blobs have been found.

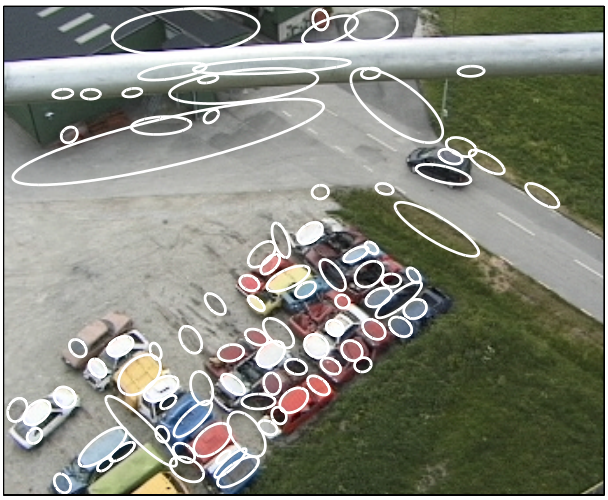

Figure 1: Detected blobs in an aerial image.

3 Colour constraint

We will use a voting scheme to find an initial set of correspondences between the two images. A potential correspondence $\mathcal{B}_i \leftrightarrow \mathcal{B}'_j$ can quickly be discarded by making use of the colour parameter of the blobs. Thus we compute the colour distances of all blobs in image 1 and all blobs in image 2, and use them to define a correspondence matrix $\mathbf{M}$. We set $M_{ij} = 1$ whenever $||\mathbf{p}_i - \mathbf{p}'_j|| < d_{\max}$ and 0 otherwise. Note that $d_{\max}$ is the same parameter as used in the blob estimation stage. If this is the optimal choice is at present unclear, but clearly the two should be related. Typically $\mathbf{M}$ will have a density of about 15%, and thus this simple operation does a good job at reducing the correspondence search space.

4 Initial correspondences

Any pair of points in image 1 can be mapped to any pair in image 2 using a *similarity transform*. In homogeneous coordinates, a similarity transform looks like this

$$\mathbf{x}' = \begin{pmatrix} s\mathbf{R} & t \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{x}\,. \tag{4}$$

We now generate blob pairs in both images, by joining together spatially adjacent blobs. Each blob gets to form ordered pairs with its three nearest neighbours. Thus, if we had N_1 and N_2 blobs in the two images, we now have $3N_1$ and $3N_2$ blob pairs. We will now try to find correspondences of such blob pairs, i.e. $\langle \mathcal{B}_i, \mathcal{B}_k \rangle \leftrightarrow \langle \mathcal{B}'_j, \mathcal{B}'_l \rangle$. For each such correspondence, we first check if the colours match, using $\mathbf{M}$ (see section 3). This excludes most candidates. For the correspondences that match, we then calculate the similarity mapping (4). We then project both blobs in the pair through the mapping, and compute their shape distance

$$d_{ij} = \frac{||\mathbf{I}_i - s^2\mathbf{R}\mathbf{I}'_j\mathbf{R}^T||}{||\mathbf{I}_i + s^2\mathbf{R}\mathbf{I}'_j\mathbf{R}^T||}\,. \tag{5}$$

Both distances are summed and added in a new correspondence matrix $\tilde{\mathbf{M}}$ according to

$$\tilde{M}_{ij} + e^{-(d_{ij}^2 + d_{kl}^2)/\sigma_s^2} \mapsto \tilde{M}_{ij} \tag{6}$$

$$\tilde{M}_{kl} + e^{-(d_{ij}^2 + d_{kl}^2)/\sigma_s^2} \mapsto \tilde{M}_{kl} \tag{7}$$

where σ_s is a soft shape distance threshold.

This implements a soft voting scheme, where very few constraints on the image structure have been imposed. A set of potential candidate correspondences $\mathcal{B}_i \leftrightarrow \mathcal{B}'_j$ are now extracted from $\tilde{\mathbf{M}}$ by requiring that the position $\tilde{M}_{ij}$ should be a maximum along both row i and column j. The result of this operation for a pair of aerial images is shown in figure 2. Note that most correspondences are correct, but that there are also a few outliers.

5 RANSAC outlier rejection

We now improve the quality of correspondences using outlier rejection with RANSAC [3]. We draw a random subset of the correspondences (4 for homography estimation, and 8 for fundamental matrix estimation), and estimate the applicable mapping ($\mathbf{F}$ or $\mathbf{H}$). We then verify all correspondences that are valid according to $\mathbf{M}$ (see section 3) with respect to spatial projection error. For a homography $\mathbf{H}$, the spatial projection error is determined by

$$D_{ij} = ||\mathbf{m}_i - \tilde{\mathbf{m}}_j|| + ||\tilde{\mathbf{m}}_i - \mathbf{m}_j|| \tag{8}$$

where

$$h\begin{pmatrix}\tilde{\mathbf{m}}_i \\ 1\end{pmatrix} = \mathbf{H}\begin{pmatrix}\mathbf{m}_i \\ 1\end{pmatrix}, \quad h\begin{pmatrix}\tilde{\mathbf{m}}_j \\ 1\end{pmatrix} = \mathbf{H}^{-1}\begin{pmatrix}\mathbf{m}_j \\ 1\end{pmatrix}. \tag{9}$$

For the fundamental matrix $\mathbf{F}$ the projection error is the distance between each point and the epipolar line $\mathbf{l}$, generated by the corresponding point

$$D_{ij} = \frac{|\mathbf{m}_i'^T\mathbf{l}'_j|}{\sqrt{l'^2_{j,1} + l'^2_{j,2}}} + \frac{|\mathbf{m}_j^T\mathbf{l}_i|}{\sqrt{l^2_{i,1} + l^2_{i,2}}}\,. \tag{10}$$

A correspondence is considered valid if the spatial projection error D_{ij} is a minimum along both row i and column j in the projection error matrix

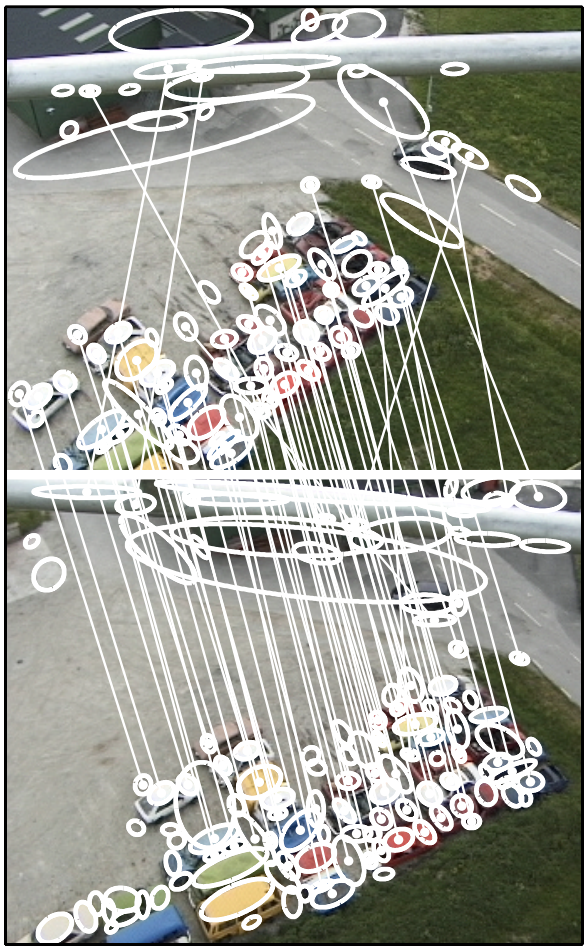

Figure 2: Raw correspondences found using voting.

D, and furthermore is smaller than some threshold s_{max}.

For each random subset of correspondences, we use the number of correspondences as a measure of how good the generated mapping was, and choose the one with most correspondences as the correct one.

The decision on how many random samples to draw is made according to

$$N = \frac{\log(1-m)}{\log(1-(1-\epsilon)^K)} \tag{11}$$

where K is the number of samples needed for the estimation of **F** or **H**, and m is the required probability of picking an inlier only sample (set to 0.99). ϵ is the probability of picking an outlier correspondence [3]. For the method in section 4 it is safe to set $\epsilon = 0.2$. For estimation of **F** we need $K = 8$ samples, which gives us $N = 25$, and for **H** we need $K = 4$ samples, and get $N = 9$.

The result of applying this RANSAC scheme on the initial correspondences in figure 2 is shown in figure 3 for a homography estimation. The blob at the top of the upper figure is still to be considered an unfortunate correspondence. It has the right colour, and the right position, but the shape of the corresponding blob is clearly wrong.

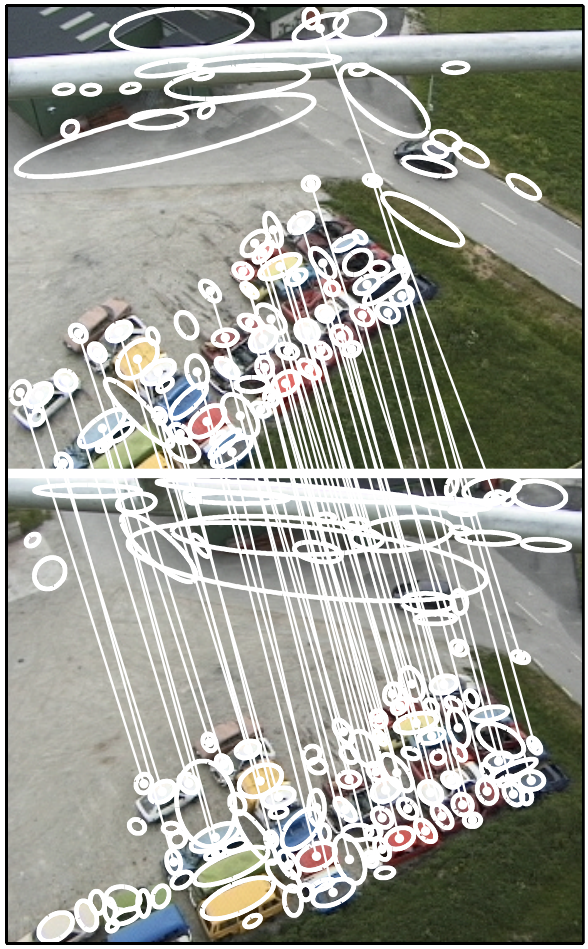

Figure 3: Refined correspondences found using RANSAC and a homography constraint.

6 Mapping an ellipse through a homography

For a homography it is possible to further constrain which correspondences are allowed as inliers by mapping the ellipse shape through the homography, and rejecting the correspondence if the shape distance (5) is above a threshold.

We will now derive a homography transformation of an ellipse. Note that even though an ellipse mapped through a homography is a new ellipse,

this mapping is merely approximately correct for regions which are not elliptical. An ellipse in image 1, is the set of points $\mathbf{x}$ in homogeneous coordinates fulfilling the relation

$$\mathbf{x}^T\mathbf{G}\mathbf{x} \leq 0\,, \quad \mathbf{G} = \begin{pmatrix} \mathbf{A} & -\mathbf{A}\mathbf{m} \\ -\mathbf{m}^T\mathbf{A} & \mathbf{m}^T\mathbf{A}\mathbf{m} - 1 \end{pmatrix} \tag{12}$$

where $\mathbf{A} = \frac{1}{4}\mathbf{I}^{-1}$ [1]. In the coordinate system of $\mathbf{x}'$, where $h\mathbf{x} = \mathbf{H}^{-1}\mathbf{x}'$, this becomes

$$\frac{1}{h^2}\mathbf{x}'^T\mathbf{H}^{-T}\mathbf{G}\mathbf{H}^{-1}\mathbf{x}' \leq 0 \tag{13}$$

which gives us a new matrix of the form

$$\frac{1}{h^2}\mathbf{H}^{-T}\mathbf{G}\mathbf{H}^{-1} = \frac{1}{h^2}\begin{pmatrix} \mathbf{C} & \mathbf{d} \\ \mathbf{d}^T & e \end{pmatrix} \tag{14}$$

We now identify the result of (13) as a new ellipse expression

$$\mathbf{x}'^T \begin{pmatrix} \mathbf{B} & -\mathbf{B}\mathbf{n} \\ -\mathbf{n}^T\mathbf{B} & \mathbf{n}^T\mathbf{B}\mathbf{n} - 1 \end{pmatrix} \mathbf{x}' \leq 0\,. \tag{15}$$

This allows us to identify $\mathbf{n}$ and $1/h^2$ as

$$\mathbf{n} = -\mathbf{C}^{-1}\mathbf{d} \quad , \quad \frac{1}{h^2} = \mathbf{n}^T\mathbf{C}\mathbf{n} - e \tag{16}$$

and $\mathbf{B}$ as

$$\mathbf{B} = \frac{1}{\mathbf{n}^T\mathbf{C}\mathbf{n} - e}\mathbf{C}\,. \tag{17}$$

Finally, the mapped ellipse shape can be found as $\tilde{\mathbf{I}} = \frac{1}{4}\mathbf{B}^{-1}$. This mapping of blob shapes through the homography is illustrated in figure 4.

7 Discussion

The results presented in this paper are quite preliminary. We have not yet tried the fundamental matrix estimation on sufficiently non-planar scenes. While the fundamental matrix constraint appears to help eliminating outliers in the scenes shown in figures 1-4, the resultant matrix $\mathbf{F}$ is quite unstable, indicating that the scene structure is too close to the degenerate plane case (see section 1).

8 Acknowledgements

The work presented in this paper has been performed within the VISCOS project (VISion in COgnitive Systems), and is funded by SSF, which is gratefully acknowledged.

Figure 4: Blob shapes projected through the homography. Original blob shapes are black ellipses, projected shapes are white.

References

[1] Per-Erik Forssén. *Low and Medium Level Vision using Channel Representations*. PhD thesis, Linköping University, Sweden, SE-581 83 Linköping, Sweden, March 2004. Dissertation No. 858, ISBN 91-7373-876-X.

[2] R. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, November 1997.

[3] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[4] S. Obdrzalek and J. Matas. Object recognition using local affine frames on distinguished regions. In *13th BMVC*, pages 113–122, September 2002.

[5] Tinne Tuytelaars and Luc Van Gool. Matching widely separated views based on affinely invariant neighbourhoods. *International Journal of Computer Vision*, 2003. To appear.