

A FAST MLE-BASED METHOD FOR ESTIMATING THE FUNDAMENTAL MATRIX

W. Chojnacki, M. Brooks, A. van den Hengel, D. Gawley

Department of Computer Science
University of Adelaide
Adelaide, SA 5005, Australia
{wojtek | mjb | hengel | dg}@cs.adelaide.edu.au

ABSTRACT

We present a novel method for estimating the fundamental matrix, a key problem arising in stereo vision. The method aims to minimise a cost function that is derived from maximum likelihood considerations. The respective minimiser turns out to be significantly more accurate than the familiar algebraic least squares technique. Furthermore, the method is identical in accuracy to a Levenberg-Marquardt minimiser, while proving simpler and faster.

1. INTRODUCTION

Many problems in computer vision may be couched in terms of parameter estimation. Accordingly, much effort has gone into the development of sophisticated techniques for generating estimates of parameters. Some of these techniques utilise covariance information characterising uncertainty in the data [1, 2]. This paper is concerned with the application of a recently introduced covariance-based method [3] to the problem of estimating the fundamental matrix (see also [1, 4–10]). However, we assume here that, as is often the case, covariance information is unavailable.

A 3D point in a scene perspectively projected onto the image plane of a camera gives rise to an *image point* represented by a pair (m_1, m_2) of coordinates, or equivalently, by the vector $\mathbf{m} = [m_1, m_2, 1]^T$. A 3D point projected onto the image planes of two cameras endowed with separate coordinate systems gives rise to a pair of *corresponding points*. When represented by $(\mathbf{m}, \mathbf{m}')$, this pair satisfies the *epipolar equation*

$$\mathbf{m}'^T \mathbf{F} \mathbf{m} = 0, \quad (1)$$

where $\mathbf{F} = [f_{ij}]$ is a 3×3 *fundamental matrix* that incorporates information about the relative orientation and internal geometry of the cameras [11]. The matrix \mathbf{F} is subject to the *rank-2 constraint* $\det \mathbf{F} = 0$.

If we let

$$\boldsymbol{\theta} = [f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33}]^T$$

be the vector of *parameters*, $\mathbf{x} = [m_1, m_2, m'_1, m'_2]^T$ be the vector of *variables*, and

$$\mathbf{u}(\mathbf{x}) = [m_1 m'_1, m_2 m'_1, m'_1, m_1 m'_2, m_2 m'_2, m'_2, m_1, m_2, 1]^T$$

be the vector of *transformed variables*, then (1) can be rewritten as $\boldsymbol{\theta}^T \mathbf{u}(\mathbf{x}) = 0$. It is this latter form of the epipolar equation that we exploit to design a fast, accurate method for estimating the fundamental matrix given a set of corresponding points. In fashioning the technique, we shall not be concerned with the issues of robustness and parameterisation. Accordingly, the performance of the new method will be gauged in isolation.

2. COST FUNCTIONS AND ESTIMATES

Estimating the fundamental matrix will rest upon the use of *cost functions* measuring the extent to which the data and candidate estimates fail to satisfy the epipolar equation. If—for simplicity—the rank-2 constraint is set aside, then, given a set of data $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ and a cost function $J = J(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n)$, a corresponding *estimate* $\hat{\boldsymbol{\theta}}$ is defined as the parameter which minimises J :

$$J(\hat{\boldsymbol{\theta}}) = \min_{\boldsymbol{\theta} \neq 0} J(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n). \quad (2)$$

Since (1) does not change if $\boldsymbol{\theta}$ is multiplied by a non-zero scalar, we consider only cost functions satisfying $J(t\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) = J(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n)$ for any non-zero scalar t . For such functions, $\hat{\boldsymbol{\theta}}$ satisfies (2) alongside $t\hat{\boldsymbol{\theta}}$ for any non-zero scalar t , and so the corresponding estimate is defined only to within a scalar factor.

2.1. Algebraic least squares estimator

A straightforward estimator is derived from the cost function

$$J_{\text{ALS}}(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \|\boldsymbol{\theta}\|^{-2} \sum_{i=1}^n \boldsymbol{\theta}^T \mathbf{A}_i \boldsymbol{\theta},$$

where $\mathbf{A}_i = \mathbf{u}(\mathbf{x}_i)\mathbf{u}(\mathbf{x}_i)^T$. Here each summand $\theta^T \mathbf{A}_i \theta$ is the square of the *algebraic distance* $|\theta^T \mathbf{u}(\mathbf{x}_i)|$. When the individual datum \mathbf{x}_i is represented by $(\mathbf{m}_i, \mathbf{m}'_i)$, the algebraic distance between the datum and a candidate fundamental matrix \mathbf{F} can be written as $|\mathbf{m}'_i{}^T \mathbf{F} \mathbf{m}_i|$. The number $J_{\text{AML}}(\theta; \mathbf{x}_1, \dots, \mathbf{x}_n)$ measures the degree to which the ray $\{t\theta: t \text{ a non-zero scalar}\}$ of equivalent candidate parameters fits the data. The value of θ for which J_{ALS} is minimal is termed the *algebraic least squares* (ALS) *estimate* and is denoted $\hat{\theta}_{\text{ALS}}$. It is uniquely determined, up to a scalar factor, by an eigenvector of $\sum_{i=1}^n \mathbf{A}_i$ associated with the smallest eigenvalue, and can be found in practice by performing singular-value decomposition (SVD) of $[\mathbf{u}(\mathbf{x}_1), \dots, \mathbf{u}(\mathbf{x}_n)]$.

3. APPROXIMATED MAXIMUM LIKELIHOOD ESTIMATOR

Adopting a maximum likelihood approach and making some necessary concessions to tractability, a strong case may be mounted for adoption of the cost function given by

$$J_{\text{AML}}(\theta; \mathbf{x}_1, \dots, \mathbf{x}_n) = \sum_{i=1}^n \frac{\theta^T \mathbf{A}_i \theta}{\theta^T \mathbf{B}_i \theta},$$

where $\mathbf{B}_i = \partial_{\mathbf{x}} \mathbf{u}(\mathbf{x}_i) \partial_{\mathbf{x}} \mathbf{u}(\mathbf{x}_i)^T$ and

$$\begin{aligned} \partial_{\mathbf{x}} \mathbf{u}(\mathbf{x})^T &= [(\partial u_j / \partial x_i)(\mathbf{x})]_{1 \leq i \leq 4, 1 \leq j \leq 9} \\ &= \begin{bmatrix} m'_1 & 0 & 0 & m'_2 & 0 & 0 & 1 & 0 & 0 \\ 0 & m'_1 & 0 & 0 & m'_2 & 0 & 0 & 1 & 0 \\ m_1 & m_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & m_1 & m_2 & 1 & 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

(see [3, 12]). The J_{AML} -based estimate of θ will be called the *approximated maximum likelihood* (AML) *estimate* and will be denoted $\hat{\theta}_{\text{AML}}$.

With the parameters to be estimated in matrix rather than vector form, the function J_{AML} underlying fundamental matrix estimation reduces to

$$J(\mathbf{F}) = \sum_{i=1}^n \frac{(\mathbf{m}'_i{}^T \mathbf{F} \mathbf{m}_i)^2}{\mathbf{m}_i{}^T \mathbf{F} \mathbf{I}^* \mathbf{F}^T \mathbf{m}_i + \mathbf{m}'_i{}^T \mathbf{F}^T \mathbf{I}^* \mathbf{F} \mathbf{m}'_i},$$

where $\mathbf{I}^* = \text{diag}(1, 1, 0)$.

3.1. Fundamental numerical scheme

Following previous work in a general setting [3], we now develop a method to calculate $\hat{\theta}_{\text{AML}}$. As a minimiser of J_{AML} , $\hat{\theta}_{\text{AML}}$ satisfies

$$[\partial_{\theta} J_{\text{AML}}(\theta; \mathbf{x}_1, \dots, \mathbf{x}_n)]_{\theta=\hat{\theta}_{\text{AML}}} = \mathbf{0}^T,$$

where $\partial_{\theta} J_{\text{AML}}$ denotes the row vector of the partial derivatives of J_{AML} with respect to θ . We term this the *variational equation*. Direct computation shows that

$$[\partial_{\theta} J_{\text{AML}}(\theta; \mathbf{x}_1, \dots, \mathbf{x}_n)]^T = 2\mathbf{X}_{\theta}\theta, \quad (3)$$

where \mathbf{X}_{θ} is the 9×9 symmetric matrix

$$\mathbf{X}_{\theta} = \sum_{i=1}^n \frac{\mathbf{A}_i}{\theta^T \mathbf{B}_i \theta} - \sum_{i=1}^n \frac{\theta^T \mathbf{A}_i \theta}{(\theta^T \mathbf{B}_i \theta)^2} \mathbf{B}_i.$$

Thus the variational equation can be rephrased as

$$[\mathbf{X}_{\theta}\theta]_{\theta=\hat{\theta}_{\text{AML}}} = \mathbf{0}. \quad (4)$$

Closed-form solutions of the variational equation may be infeasible, so in practice $\hat{\theta}_{\text{AML}}$ has to be found numerically. Throughout we shall assume that $\hat{\theta}_{\text{AML}}$ lies close to $\hat{\theta}_{\text{ALS}}$. This assumption is meant to increase the chances that any candidate minimiser obtained via a numerical method seeded with $\hat{\theta}_{\text{ALS}}$ coincides with $\hat{\theta}_{\text{AML}}$.

A vector θ satisfies (4) if and only if it falls into the null space of the matrix \mathbf{X}_{θ} . Thus, if θ_{k-1} is a tentative guess, then an improved guess can be obtained by picking a vector θ_k from that eigenspace of $\mathbf{X}_{\theta_{k-1}}$ which most closely approximates the null space of \mathbf{X}_{θ} ; this eigenspace is, of course, the one corresponding to the eigenvalue closest to zero. The *fundamental numerical scheme* implementing this idea is presented in Figure 1. The algorithm can be regarded as a variant of the Newton-Raphson method (see [3]).

1. Set $\theta_0 = \hat{\theta}_{\text{ALS}}$.
2. Assuming that θ_{k-1} is known, compute the matrix $\mathbf{X}_{\theta_{k-1}}$.
3. Compute a normalised eigenvector of $\mathbf{X}_{\theta_{k-1}}$ corresponding to the eigenvalue closest to zero and take this eigenvector for θ_k .
4. If θ_k is sufficiently close to θ_{k-1} , then terminate the procedure; otherwise increment k and return to Step 2.

Fig. 1. Fundamental numerical scheme.

4. EXPERIMENTS

The fundamental numerical scheme (FNS) was tested against the algebraic least squares estimator (ALS) and a Levenberg-Marquardt minimiser (LM).

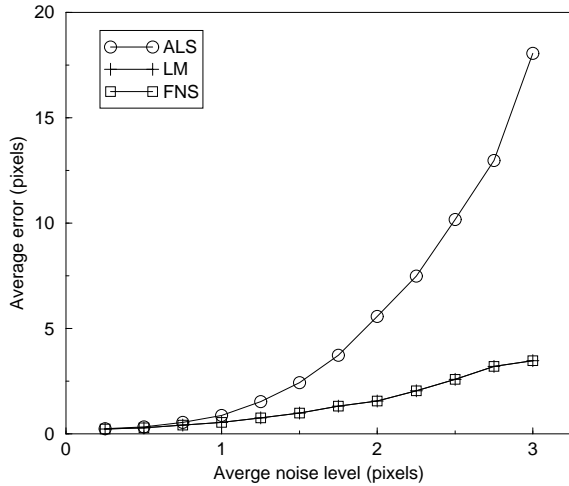


Fig. 2. Fundamental matrix estimation errors vs. average noise level.

FNS was implemented using the EISPACK routine RS for computation of eigenvalues and associated eigenvectors of symmetric matrices. The ALS method uses the LINPACK routine DSVDC to perform SVD. For the LM scheme, the MINPACK routine LMDER was used to directly minimise J_{AML} , with the analytical derivatives of J_{AML} , as in (3), supplied so as to improve the execution time.

Our experiments proceeded as follows. A realistic stereo camera configuration was first selected with non coplanar optical axes, and slightly differing left and right camera intrinsic parameters. 3D points were then projected onto the images so as to generate many pairs of corresponding points. A range of tests was then conducted, each carried out with respect to an *average level of noise*, σ .

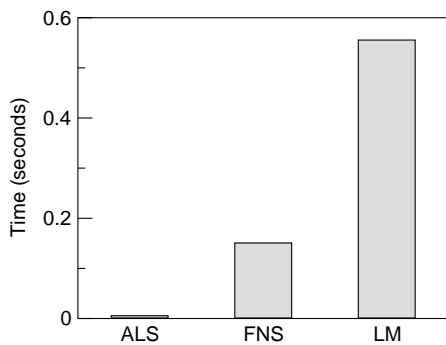


Fig. 3. Timing results for various methods.

For a given test, each image point was perturbed by adding zero-mean Gaussian noise of standard deviation σ

independently to each of the two coordinates. Each method was then supplied with these noisy matching points and challenged to compute the fundamental matrix.

For each σ , the fundamental matrix was computed 50 times from a specific set of 96 corresponding points, with new perturbations being added each time. For each fundamental matrix obtained, an error measure was computed as the sum of the distances of the underlying *true points* to the epipolar lines derived from the estimated fundamental matrix, in both the left and right images. A composite error measure was then obtained by averaging this error over all 50 trials. This entire process was repeated for different average levels of noise (with σ varying from 0.25 to 3 pixels in steps of 0.25).

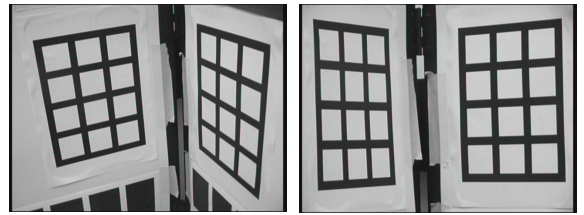


Fig. 4. Images used in the reconstructions.

Figure 2 shows the average epipolar-distance pixel errors obtained for each method. The tests reveal that LM and FNS perform identically. In our experiments, therefore, FNS succeeded in minimising J_{AML} as well as LM did. In contrast, ALS is seen to be far less successful.

Timing tests were conducted on the various schemes. Stopping conditions were devised so as to place similar demands upon the iterative methods, FNS and LM. None of the schemes' timings were affected significantly by a change in noise level. Figure 3 shows histogram timing data for the methods. In each case, the bar denotes the time taken to complete a single test, averaged over the complete suite of experiments.

Finally, an indicative reconstruction test was carried out. Corresponding points were extracted from a stereo image pair of a calibration book opened at 90 deg (see Figure 4). These points were fed to the ALS and FNS estimators and the associated fundamental matrices were computed. A self-calibration procedure was then used to determine the relative orientation of the cameras, with the camera intrinsics having been pre-calibrated in the laboratory. Figures 5 and 6 show various views of the calibration book. The FNS reconstruction exhibits better depth estimates, with, for example, a far more accurate angle of opening of the calibration book.

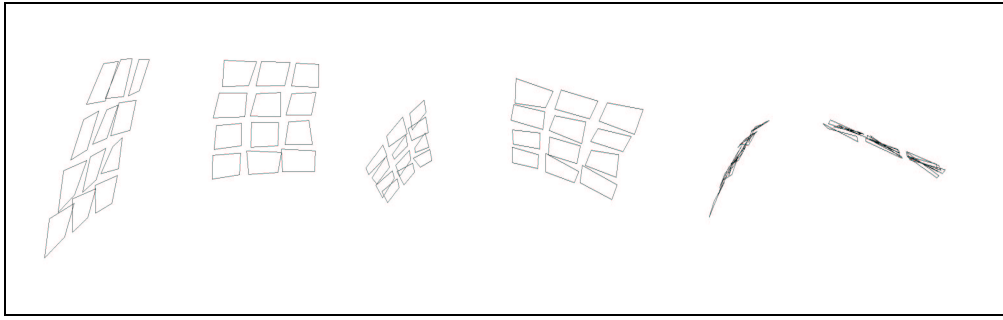


Fig. 5. Reconstructions of the noisy points using ALS.

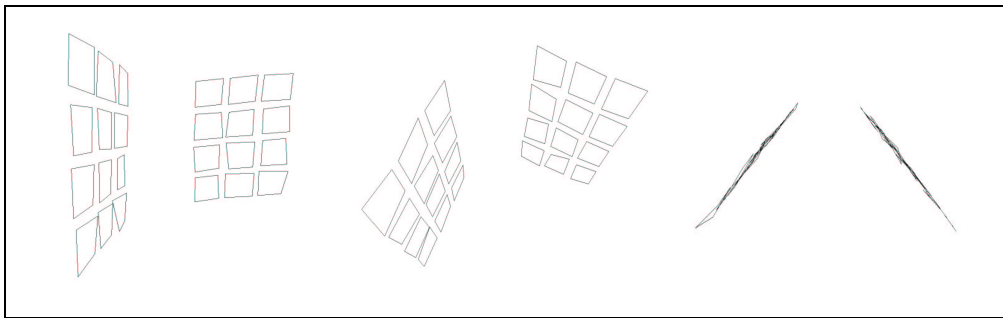


Fig. 6. Reconstructions of the noisy points using FNS.

5. ACKNOWLEDGEMENTS

The authors are grateful for the insightful comments of Kenichi Kanatani, Garry Newsam, Naoya Ohta and Robyn Owens. This work was in part funded by the Australian Research Council and the Cooperative Research Centre for Sensor Signal and Information Processing.

6. REFERENCES

- [1] M. Mühlich and R. Mester, "The role of total least squares in motion analysis," in *Proc. 5th European Conference on Computer Vision, Freiburg, Germany*. 1998, vol. 1407 of *Lecture Notes in Computer Science*, pp. 305–321, Springer-Verlag, Berlin.
- [2] B. Triggs, "Optimal estimation of matching constraints," in *3D Structure from Multiple Images of Large-Scale Environments, European Workshop, Freiburg, Germany*. 1998, vol. 1506 of *Lecture Notes in Computer Science*, pp. 63–77, Springer-Verlag, Berlin.
- [3] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley, "On the fitting of surfaces to data with covariances," *IEEE Trans. Pattern Analysis & Machine Intelligence*, vol. 22, no. 11, pp. 1294–1303, 2000.
- [4] R. Hartley, "In defense of the eight-point algorithm," *IEEE Trans. Pattern Analysis & Machine Intelligence*, vol. 19, no. 6, pp. 580–593, 1997.
- [5] Y. Leedan and P. Meer, "Heteroscedastic regression in computer vision: problems with bilinear constraint," *Int. J. Computer Vision*, vol. 37, no. 2, pp. 127–150, 2000.
- [6] Q.-T. Luong and O. D. Faugeras, "The fundamental matrix: theory, algorithms, and stability analysis," *Int. J. Computer Vision*, vol. 17, no. 1, pp. 43–75, 1996.
- [7] K. Kanatani, "Unbiased estimation and statistical analysis of 3-D rigid motion from two views," *IEEE Trans. Pattern Analysis & Machine Intelligence*, vol. 15, no. 1, pp. 37–50, 1993.
- [8] P. H. S. Torr and D. W. Murray, "The development and comparison of robust methods for estimating the fundamental matrix," *Int. J. Computer Vision*, vol. 24, no. 3, pp. 271–300, 1997.
- [9] Z. Zhang, "Determining the epipolar geometry and its uncertainty: a review," *Int. J. Computer Vision*, vol. 27, no. 2, pp. 161–195, 1998.
- [10] Z. Zhang, "On the optimization criteria used in two-view motion analysis," *IEEE Trans. Pattern Analysis & Machine Intelligence*, vol. 20, no. 7, pp. 717–729, 1998.
- [11] O. D. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, The MIT Press, Cambridge, Mass., 1993.
- [12] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier, Amsterdam, 1996.