Robust Egomotion Estimation from Affine Motion Parallax

Jonathan M. Lawn and Roberto Cipolla

Department of Engineering, University of Cambridge, CB2 1PZ, England

Abstract. A method of determining the motion of a camera from its image velocities is described that is insensitive to noise and intrinsic camera parameters. This algorithm is based on a novel extension of motion parallax which does not require the instantaneous alignment of features, but uses sparse visual motion estimates to extract the direction of translation of the camera directly, after which determination of the camera rotation and the depths of the image features follows easily. A method for calculating the expected uncertainty in the estimates is also described which allows optimal estimation and can also detect and reject independent motion and false correspondences. Experiments using small perturbation analysis show a favourable comparison with existing methods, and specifically the Fundamental Matrix method.

1 Introduction

The visual motion of a scene from a camera provides a cue for the determination of the scene structure and the viewer motion (*egomotion*), and many systems would benefit from using a freely moving camera with no other motion sensors. An important criterion for the judging of algorithms is *robustness*: insensitivity to, and graceful failure from, image motion measurement noise, independent motion and erroneous measurements, sparse image motion data, narrow field of view and camera calibration.

Motion parallax methods [10, 13, 5] try to find the *epipole*, the intersection of the direction of motion with the imaging surface, independent of *intrinsic camera calibration* (knowledge of the projection plane centre, the aspect ratio and the focal length, assuming no other distortions are significant) by using the fact that the relative visual motions of points instantaneously coincident in the image can be used to cancel the rotational part of the visual motion, but unfortunately this requires dense velocity field measurements at sudden depth changes [5, 13].

The *Fundamental Matrix* method [3] also attempts to find the epipole independent of intrinsic camera parameters by estimating a 3×3 singular homogeneous matrix (7 independent variables), as does Heeger and Jeepson's Subspace method [4] using a two-dimensional search of a non-linear function. There are many other methods for estimating the camera motion from an image pair or sequence, but many are inexact, involve very high dimensional searches, or use higher derivatives of the image motion than just the velocities which makes them susceptible to measurement noise [12].

Lecture Notes in Computer Science, Vol. 800
Jan-Olof Eklundh (Ed.)
Computer Vision – ECCV '94

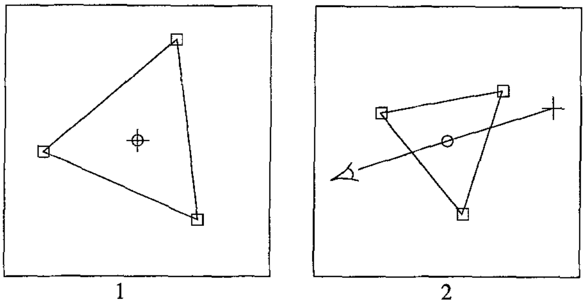


Fig. 1. The Affine Motion Parallax Algorithm
In view one, there are four real feature points, the *parallax* point (the cross), and the three *basis* points (the squares). The circle is an imaginary *virtual* point that is defined to be coincident in the image with the parallax point, but on the plane in space defined by the basis points. In view two, the motion of the parallax point is observed, and that of the virtual point is deduced from the basis points, assuming the plane that they define is deforming affinely in the image. We have therefore recovered exact motion parallax, the relative motion of the virtual point and the parallax point, from four sparse points [1, 7]. The epipole must lie on the line given by the relative motion [10].

2 Theoretical Framework

Using the *affine* (linear projection) assumption globally will result in parallel motion parallax vectors [7], as it is only valid for small fields of view. However weak perspective assumptions can be used in a number of small regions of a wide, full perspective view. We propose using small regions of four points[1] to produce 1D constraints on the direction of motion (Figure 1). This is similar to motion parallax [10] but does not require the alignment of features. Two or more of these constraints can then be combined to find the epipole.[2]

The camera rotation can be found by a least-squares estimate from the component of visual motion perpendicular to the direction of camera translation [4]. This requires a minimum of three points. The feature depths can then be found.[3] Both are dependent on the intrinsic camera parameters, but the scene structure can still be extracted to within a 3D projective transformation without calibration once the epipole is found [2].

Uncertainty and Errors: For real, noisy data the epipole constraints will not intersect. A method is needed that will optimally integrate the constraints and reject the outliers due to independent motion and incorrect visual motion measurements. The uncertainty in the motion measurements and then constraint

[1] It can be shown that it is sufficient that all four are close in the image, and therefore weak perspective or affine assumptions are not necessary [9].

[2] This entire calculation can be performed by geometric construction on the image plane, and is therefore independent of intrinsic camera parameters.

[3] Full details can be found in [9].

estimates may be represented by covariance matrices [8]. Different methods have been investigated for optimally combining the constraints. Initially a closed form solution was used, weighting the sum of the epipole estimates given by every pair of constraints so as to give the smallest covariance trace. However, this was superseded by an iterative method that finds the epipole that is most perpendicular to the constraint plane normals.[4] 'Renormalisation' [6] was not found useful, due to the uncertainties of the constraints not being known accurately.

3 Experimental Comparisons

In this section, we test the algorithm described above and compare it with an existing technique, the Fundamental Matrix method. We shall assume that a number of 'corner' features can be tracked between frames. The features used are hand-picked corners with significant amounts of noise (including integer quantization) taken from low resolution video sequences from a hand-held camera (600×500 pixels). Some sample images from a sequence involving large arbitrary camera rotations but consistent camera translation (towards the doors) are displayed in Figures 2 and 3.

Affine Motion Parallax Method: Each detected corner and its three nearest neighbours form a group that provides a parallax measurement (Fig. 2b). The calculations are formulated using the image hemisphere and velocities rather than discrete displacements. Constraints with excessive errors are rejected [9]. The rotational and then translational components of the image motion are then extracted (Figs. 2c and 2d).

Fundamental Matrix Method: This implementation scheme uses a non-linear method (DIST-L [11]) which aims to minimise the distance from the points to their epipolar lines in both of the pair of images, given that the transformation is singular, by varying eight homogeneous coefficients.

First order perturbation analysis is used to determine the instability of the algorithms (Fig. 3).

Comparison: The emphasis of our method is on robustness and graceful failure, and these results (Figs. 2 and 3) demonstrate its success. The constraint rejection scheme does occasionally fail, but this is probably inevitable, and hopefully using temporal filtering along a sequence of frames will provide sufficient information to correct this. The accuracy will also be improved by a wider field of view.

The Fundamental Matrix method, as implemented, regularly fails to locate the correct minima (Figure 3): if a good estimate of the egomotion is used to derive an initial F matrix from which to start the minimization, then the method can improve the accuracy of the estimates (better than the Affine Motion Parallax method), but without a good initialization, a large number of false minima exist and minimisation becomes very difficult. Another obvious disadvantage of the Fundamental Matrix method is that it does not naturally produce an uncertainty measure, whereas the data from the Affine Motion Parallax algorithm does, and is therefore easily fused with independent information.

[4] in a least squares sense, weighted according to the covariances of the constraints in the direction of the current epipole estimate

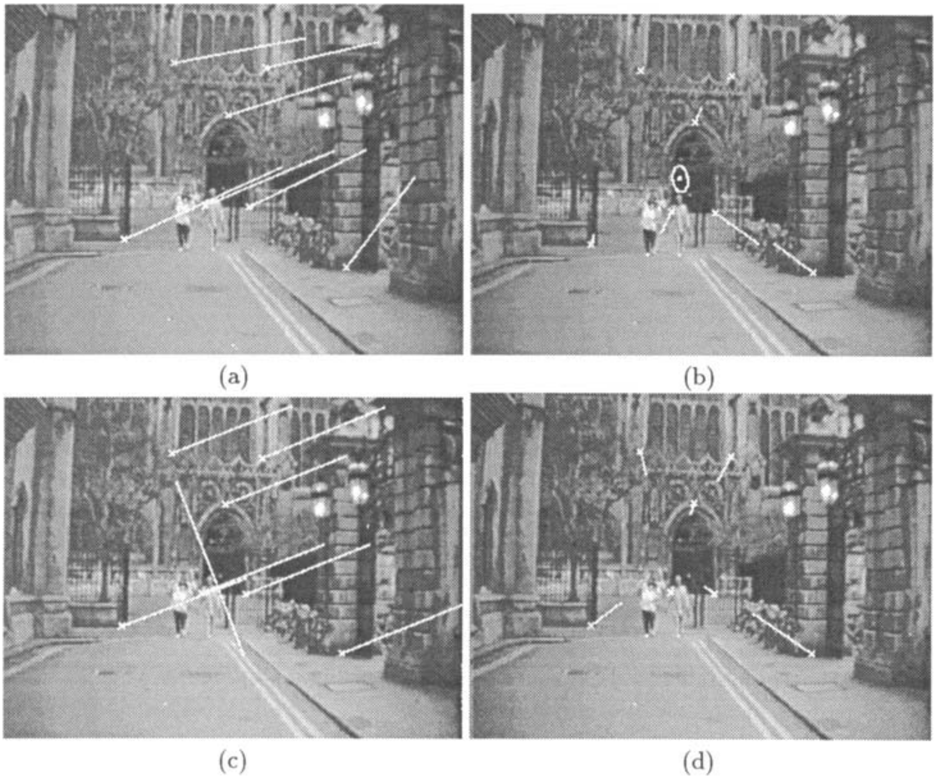
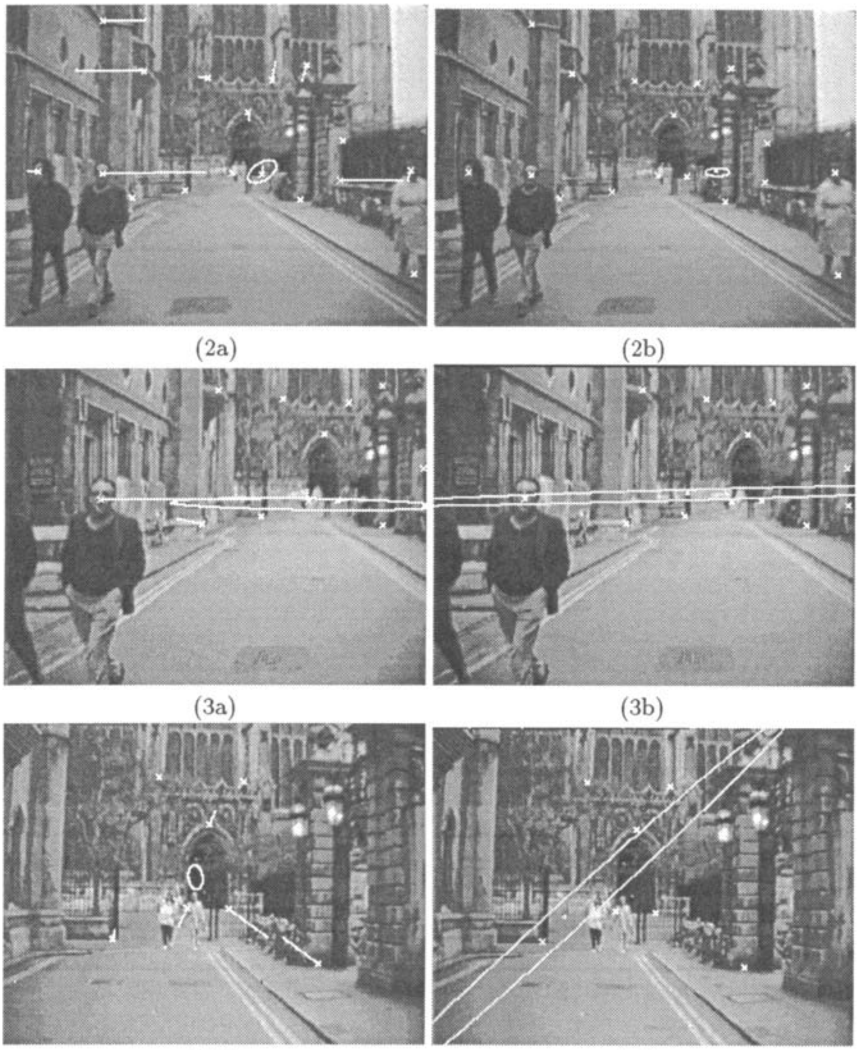


Fig. 2. The approach to Kings College Chapel North Gate, frames (3-)4 using Affine Motion Parallax. (a) shows the image motion measurements used (with the lines indicating the inter-frame displacement). Despite the sparseness of the data, the algorithm extracts five parallax constraints (length indicating certainty), and makes an accurate estimate of the epipole (b), shown by the ellipse indicating predicted uncertainty (1 standard deviation, assuming $\sqrt{2}$ pixel standard deviation image measurement noise) plotted around the estimated epipole. (c) shows the projection of the axis of rotation and the rotational components of the image motion, found from a least squares estimate of the image velocities perpendicular to the epipolar line, assuming approximate intrinsic camera parameters. Though the rotational components are large, they can still be modelled to a sufficient accuracy by the velocity model. (d) shows the translational component of the image motion, calculated by subtracting the rotational components. With accurate intrinsic calibration, the translational components will all point at the epipole and have a magnitude approximately inversely proportional to their depth and proportional to their distance from the epipole, and qualitatively this can be seen here.



(2a) (2b)

(3a) (3b)

(4a) (4b)

Fig. 3. The approach to Kings College Chapel North Gate (Frames 1–4) using Affine Motion Parallax (a) and the Fundamental matrix method (b), showing the epipole uncertainty ellipse found using small perturbation analysis (again assuming $\sqrt{2}$ pixel standard deviation image measurement noise). (a) also shows the affine motion parallax constraints found (length indicating certainty), some of which are rejected as inconsistent. The people are indistinguishable from stationary objects that are twice as close, and are therefore not rejected as independent motion. As the sequence progresses, the number of features drops, but though the Fundamental Matrix method becomes very unstable (3b and 4b), our algorithm degrades gracefully. Normally only constraints with low uncertainties are used, and knowing only one constraint on the epipole can still be very useful [1]. Here the rejection threshold has been relaxed to ensure an epipole estimate can be displayed even when the translational motion is small.

4 Conclusion

We have presented a method of extracting the epipole from a pair of frames. The camera rotation and feature depths can then be found. The method is robust because it uses first derivatives (or displacements) of the image positions only, and because the first stage involves a direct solution for only two variables (the epipole direction) and is independent of camera rotation and intrinsic parameters. A method for optimally combining the epipole constraints using the uncertainties of each estimate is also presented. There are a number of approximations due to statistical dependence, but results show the 'optimal' solution found to be good. Results from real scenes [9] show that this method compares favourably with the Fundamental Matrix method, especially when the epipole is not in view or the point set is sparse.

References

1. R. Cipolla, Y. Okamoto, and Y. Kuno. Robust structure from motion using motion parallax. In *Proc. 4th Int. Conf. on Computer Vision*, pages 374–382, 1993.
2. 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*, pages 563–578. Springer–Verlag, 1992.
3. O.D. Faugeras, Q.-T. Luong, and S.J. Maybank. Camera self-calibration: Theory and experiments. In G. Sandini, editor, *Proc. 2nd European Conference on Computer Vision*, pages 321–334. Springer–Verlag, 1992.
4. D.J. Heeger and A.D. Jepson. Subspace methods for recovering rigid motion I: Algorithm and implementation. *Int. J. of Comp. Vision*, 7(2):95–117, 1992.
5. E.C. Hildreth. Recovering heading for visually guided navigation in the presence of self-moving objects. *Phil. Trans. Roy. Soc. London, Ser. B*, 337:305–313, 1992.
6. K. Kanatani. Renormalization for unbiased estimation. In *Proc. 4th Int. Conf. on Computer Vision*, pages 599–606, 1993.
7. J.J. Koenderink and A.J. van Doorn. Affine structure from motion. *J. Opt. Soc. America*, 8:377–385, 1991.
8. J.M. Lawn and R. Cipolla. Epipole estimation using affine motion-parallax. In *Proc. 4th British Machine Vision Conference*, pages 379–388, 1993.
9. J.M. Lawn and R. Cipolla. Robust egomotion estimation from affine motion-parallax. Tech. Rep. CUED/F-INFENG/TR.160, University of Cambridge, 1994.
10. H.C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proc. of the Royal Society of London, Series B*, 208:385–397, 1980.
11. Q.-T. Luong, R. Deriche, O. Faugeras, and T. Papadopoulo. On determining the fundamental matrix: Analysis of different methods and experimental results. Technical Report 1894, INRIA, France, 1993.
12. S. Negahdaripour and S. Lee. Motion recovery from image sequences using only first order optical flow information. *Int. J. of Comp. Vision*, 9(3):163–184, 1992.
13. J.H. Rieger and D.L. Lawton. Processing differential image motion. *J. Opt. Soc. America*, A2(2):354–360, 1985.