Calibration of Image Sequences for Model Visualisation

A. Broadhurst

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

R. Cipolla

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

Abstract

The object of this paper is to find a quick and accurate method for computing the projection matrices of an image sequence, so that the error is distributed evenly along the sequence. It assumes that a set of correspondences between points in the images is known, and that these points represent rigid points in the world. This paper extends the algebraic minimisation approach developed by Hartley so that it can be used for long image sequences. This is achieved by initially computing a trifocal tensor using the three most extreme views. The intermediate views are then computed linearly using the trifocal tensor. An iterative algorithm is presented which perturbs the twelve entries of one camera matrix so that the algebraic error along the whole sequence is minimised.

1 Introduction

In recent years, the availability of texture mapping graphics hardware has increased, so that most desktop computers are able to display exceedingly realistic three-dimensional scenes. To achieve these new levels of realism it is important that textured models are obtained from photographic images.

Established geometric techniques exist for computing the motion between two or three views. The fundamental matrix describes the projective geometry between two views and the trifocal tensor describes the geometry of three views. In the past, sequences have been computed by linking together many fundamental matrices or tensors into long chains.

The trifocal tensor [1] expresses the geometry of three views and can be computed from six points in three views, up to one of three projective ambiguities, as was shown by Quan [2]. Recently research into higher order tensors by Heyden [3] has shown that the quadrifocal tensor (four views) is the highest order possible. For longer sequences the only improvement is the ability to average over a larger number of images.

Since closely spaced cameras have a small motion, it is important to use widely spaced views to improve the accuracy of the results. In this paper the trifocal tensor will be used to develop a method for computing the projection matrices of a sequence of images, so that all the projection matrices are in the same projective framework. This will be one of a family of solutions that all describe the projective geometry. The self-calibration technique by Pollefeys et al.[4] is used to select one metric solution from the family of projective solutions.

2 Background

A point $\mathbf{X} = [X\ Y\ Z\ 1]^T$ is projected onto an image plane by a 3×4 projection matrix P, which can be expressed in homogeneous co-ordinates as

$$\mathbf{p} = \begin{bmatrix} sx \\ sy \\ s \end{bmatrix} \cong \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = P\mathbf{X} \quad (1)$$

The matrix P has 11 degrees of freedom and can be computed from six known points.

$$P' = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} = [A\ a_4] \quad (2)$$

If a second camera P' is introduced and the first camera is assumed to be the normalised camera $P = [I \mid 0]$ then the fundamental matrix is given by:

$$F = [a_4]_\times A \quad (3)$$

Points on the two images are related by equation 4 which can be expressed geometrically as the epipolar constraint (see figure 1).

$$p'^T F p = 0 \quad (4)$$

$$F p = 0 \quad (5)$$

$$p'^T F = 0 \quad (6)$$

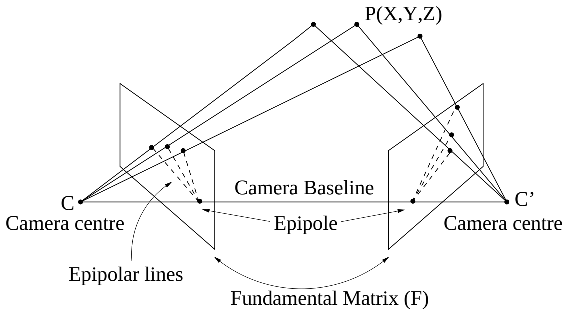

Figure 1: Epipolar geometry.

The ray projected from a point in the first image will be viewed as a line in the second image. Alternatively, a point in the world defines an epipolar plane which intersects both image planes at an epipolar line, and the projection of the point in the world must lie along the epipolar line. All epipolar lines intersect at the epipole, so the fundamental matrix must have a zero determinant, and this can be seen in equation 5. Since the fundamental matrix has 9 entries, can be computed up to scale, and has a zero determinant, it must have 7 degrees of freedom.

$$P'' = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \end{bmatrix} \tag{7}$$

If a third camera P'' is introduced, it defines a plane containing the three camera centres (the trifocal plane) and this plane introduces three constraints on the epipoles (equation 8..10) which are can be seen in figure 2.

$$e_{12}^T F_{31} e_{32} = 0 \tag{8}$$

$$e_{31}^T F_{23} e_{21} = 0 \tag{9}$$

$$e_{23}^T F_{12} e_{13} = 0 \tag{10}$$

The $3\times\ 3\ \times 3$ trifocal tensor [1] is a more general way of representing the geometry of three views and is expressed in terms of lines (equation 12). A line λ in the first view defines a plane through its camera centre. If a line λ' or point p' in the second image is projected onto this plane, it defines a line or point in the world. This line or point in the world can then be projected into the third view. Alternatively contracting the tensor with λ defines a homography for lines from the second to third images (equation 13).

$$T_i^{jk} \cong (a_i^j b_4^k - a_4^j b_i^k) \tag{11}$$

$$\lambda_i'' \cong \lambda_j \lambda_k' T_i^{jk} \tag{12}$$

$$\lambda'' \cong H_\lambda \lambda' \tag{13}$$

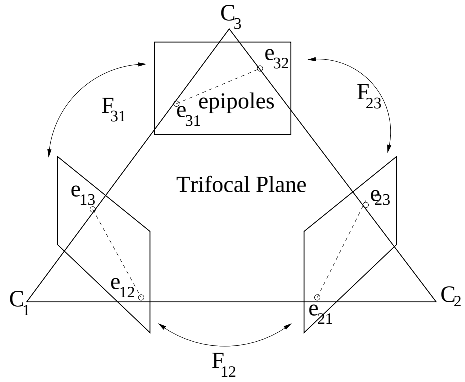

Figure 2: The trifocal plane.

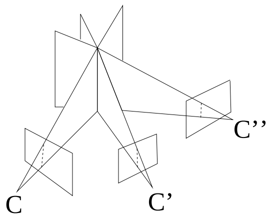

Figure 3: A line projected onto three views.

The trifocal tensor has 18 $(3 \times 7 - 3)$ degrees of freedom and can be computed linearly from 7 point matches. An alternative computation by Quan [2] using a canonical co-ordinate system, requires only 6 points matched across the three views, but produces one or three solutions. This approach was implemented and analysed in detail by Torr and Zisserman [5] who used a combination of robust sampling and optimisation to accurately compute the trifocal tensor. An alternate approach by Hartley [6] adjusts the epipoles to minimise the algebraic error of the linear solution. This paper describes an extension of this work, so that it can be used to compute the geometry of long image sequences.

3 Method

In this section a new method of computing the projection matrices of an image sequence will be presented.

A video camera can be used to obtain a long sequence of images. The motion between successive frames will be small but the overall motion of the viewer will be large. The trifocal tensor of the three most extreme views will be much more accurate than the tensor of three successive views. Once this outer tensor is known, the intermediate projection matrices can be computed using the linear algorithm which will now be presented.

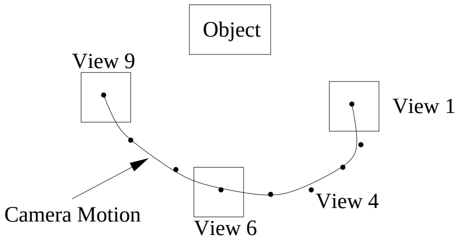

Figure 4: The initial tensor is computed from the three most extreme views

The projection matrices of two views can be obtained from the initial tensor. These two matrices are represented by a normalised camera matrix $P = [I \mid 0]$ and a 3×4 projection matrix P', then the object is to compute for each view in turn along the sequence the 3×4 camera matrix P''.

The equation for obtaining the trifocal tensor from three camera matrices where the first camera is normalised [1] is:

$$T_{ijk} = P'_{ji} P''_{k4} - P'_{i4} P''_{jk} \tag{14}$$

Since the entries of P'_{rc} are known, the vector $\mathbf{t}$, containing the entries of T_{ijk} written as a 27×1 column vector, can be written as a linear function of $\mathbf{b}$, which is a 12×1 column vector containing the entries of P''.

$$\mathbf{t} = H\mathbf{b} \tag{15}$$

In this equation H is a sparse 27×12 matrix containing the entries of P'.

The linear solution to the trifocal tensor can be written as a measurement matrix M multiplied by the tensor vector $\mathbf{t}$, and by combining equations 15 and 16, a linear solution to the third camera matrix is obtained.

$$M\mathbf{t} = 0 \tag{16}$$

$$MH\mathbf{b} = 0 \tag{17}$$

The linear solution to this equation gives the entries of the new camera matrix $\mathbf{b}$.

Algorithm 1 The linear algorithm

Compute initial tensor.
Extract P and P' (P'' is ignored).
for each intermediate image i **do**
  Compute the linear solution for P_i (equation 17)
end for

4 Non-linear Approach

The linear solution will not find an optimal solution for the projection matrices as it is dependent on the initial projection matrix P'. The non-linear approach solves this problem by optimising the values of P' so that the algebraic error [6] along the whole image sequence is minimised.

An initial solution is computed using the linear method, and the twelve entries of the last camera matrix are adjusted, such that the algebraic error of the whole sequence is minimised. To ensure the minimisation process does not converge on the trivial solution (where all cameras are at the same point) the constraint $\|\mathbf{t}\| = 1$ must be enforced. This is implemented using Levenberg-Marquardt minimisation.

Algorithm 2 The non-linear algorithm

Compute initial tensor.
Extract P and P'. Normalise P.
while converging **do**
  Compute the linear solution for the sequence.
  Update P' to minimise the algebraic error.
end while

This algorithm is an extension of the work by Hartley [6] from the three view trifocal tensor to long image sequences. Hartley solved for the trifocal tensor by optimising the epipoles so that the algebraic error of the camera matrix entries was minimised. In this work one matrix (P') is fixed and the projection matrices of the image sequence are computed.

5 Results

The stability of the linear and non-linear algorithms have been investigated to determine their robustness in the presence of Gaussian noise. A synthetic set of point data was generated by randomly selecting a number of world points with mean zero and with a standard deviation of 1.0 . The cameras were positioned 5 units away from the synthetic point data, and the rotations of the cameras and a small perturbation of their position was chosen randomly. Image points from P and P' were projected into each of the intermediate images, and the geometric distance between the projected points and the correct locations was measured.

The mean and standard deviation of the geometric point error for a 20 point dataset is shown for both the linear and non-linear algorithms in figures 5 and figure 6, respectively. An analysis of the effect of the size of the matched point set on the non-linear algorithm is shown in figure 7.

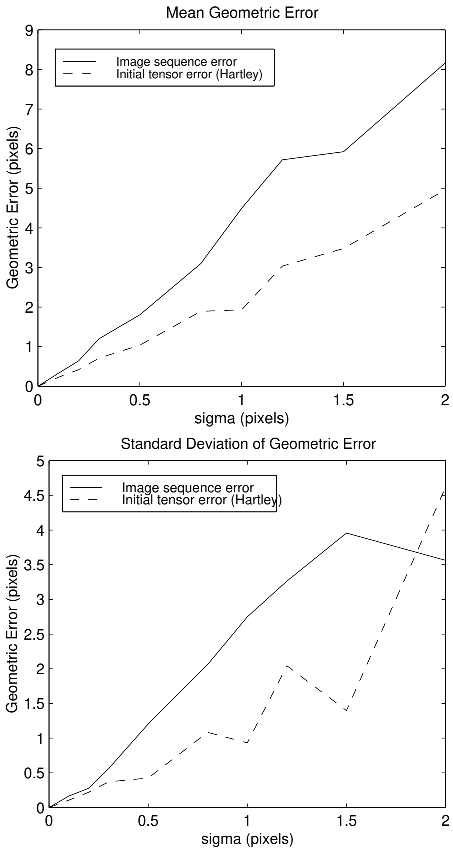

Figure 5: Stability analysis of the linear algorithm. The initial trifocal tensor (dashed) is computed using Hartley's algebraic minimisation. The upper curve is the mean geometric distance between a point projected from the reference views and its correct location.

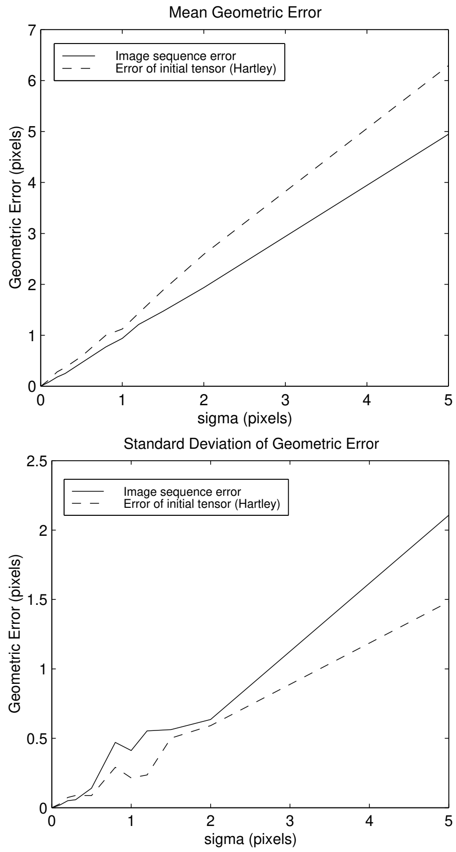

Figure 6: Stability analysis of the non-linear algorithm. The initial trifocal tensor (dashed) is computed using Hartley's algebraic minimisation. The upper curve is the mean geometric distance between a point projected from the reference views and its correct location.

5.1 Results Using Real Images

A sequence of six images, of a calibration grid, was captured using a single PULNIX camera with an 8mm lens. The Harris [7] corner detector was used to locate features in each of the six images, and the features were manually matched. Since not all of the corners were accurately located, a robust computation of the trifocal tensor was used (Torr and Zisserman [5]). The mean geometric distances using the linear algorithm are listed in table 1. The final matrices are shown in figure 8. Epipolar lines and their corresponding point matches have been marked on the image. The mean geometric distance from each point to its epipolar lines is 0.6 pixels.

6 Discussion

The algebraic minimisation algorithm has a number of advantages over a conventional bundle adjustment. A bundle adjustment requires $11N$ parameters to be optimised. If the derivatives are approximated this requires $11N$ computations of the geometric point error for each iteration. By perturbing just one of the matrices, and using the linear solution to the intermediate images, the optimisation problem is reduced to 12 parameters.

It is interesting to compare the results of the linear and non-linear algorithms. It can be seen that the non-linear solution is significantly more robust to noise than the linear solution. The 5 pixels of noise produced only 2 pixels of error in the standard deviation. The reason for the poor stability of the linear solution is that small errors in P' cause large errors in the intermediate images (see figure 5). By varying P' the non-linear algorithm becomes significantly more robust to noise.

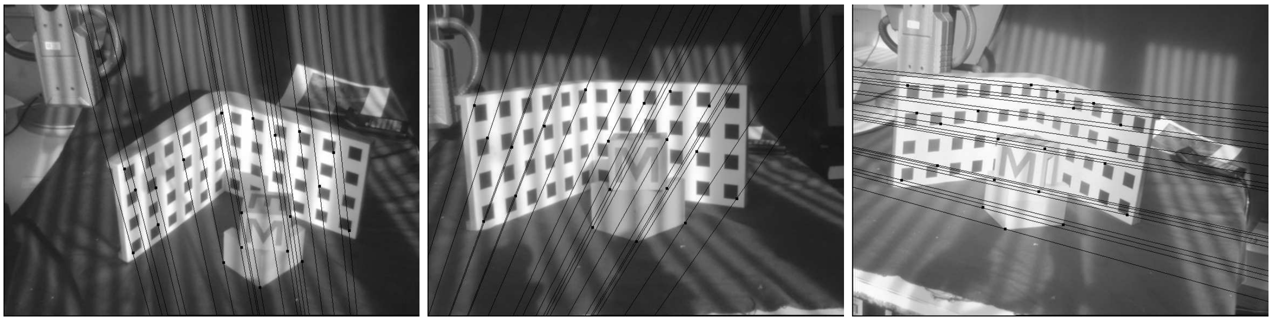

Figure 8: Calibration grid image sequence.

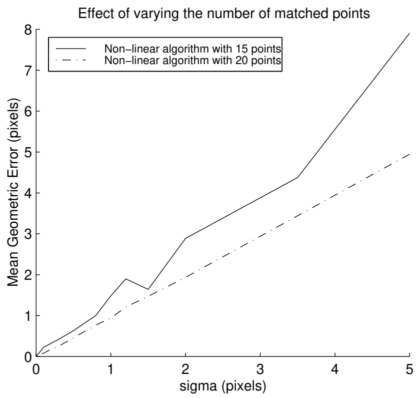

Figure 7: Effect of number of matched points on the stability of the non-linear algorithm.

Table 1: Projection Matrices

Image	Mean Distance (all 30 matches)	Mean Distance (25 inliers)
Reference	2.41	1.52
2	3.76	2.5
3	3.20	1.8
4	2.26	1.4
5	1.28	0.99

7 Application 1: Visualisation of Models from Image Sequences

As the reader may recall, the motivation for computing the projection matrices, was to visualise three-dimensional models. In this application the models are stored as a sequence of segmented images. Each segment is assigned to a planar surface and each surface is assigned a vector normal.

A new viewpoint is generated by first projecting the known segmentations into the new viewpoint. If the camera matrix of the reference image is $P = [I \mid 0]$, and the camera matrix of the new image $P' = [A \,|\, a_4]$, the plane $[\mathbf{d}\ n]$ is transfered using the homography [8] :

$$H = nA - a_4 d^T \tag{18}$$

The visualisation algorithm requires an image segmentation for each reference image, and the plane vector of each surface. This information is specified once for each surface. A closed contour is drawn using a mouse and three points along the contour are matched with corresponding points in any of the other reference images. The linear solution to the plane vector is computed:

$$p' \cong (nA - a_4 d^T)p \tag{19}$$
$$0 = [p']_\times (nA - a_4 d^T)p \tag{20}$$
$$0 = ([p']_\times A)n - [p']_\times (a_4 p^T)d \tag{21}$$
$$0 = M[d\ n]^T \tag{22}$$

This equation can be easily solved using the SVD from three point matches, provided that the image co-ordinates are correctly normalised. For stable solutions it, is essential that the corresponding points obey the epipolar constraint.

An example of surface matching is shown in figure 9 and the rendered output is shown in figure 10 .

Figure 9: A surface is added to the model by specifying it in one view. Three points must be matched to define the plane. The segment can be projected into all of the images.

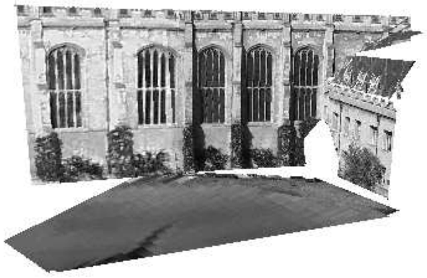

Figure 10: This image was rendered using a segmented sequence of images, from surface normal vectors.

8 Application 2: Image Editing

Another interesting application for calibrated image sequences is image editing. Consider the case where we want to take a photograph of a building which is obscured by a number of trees or a road sign. If a number of images were available from other viewpoints, and the model of the building was known, then the area obscured by the tree or road sign could be rendered using information from the other images. This requires both a knowledge of the camera geometry and the model of the obscured objects.

In the following example a calibration grid is obscured by some wooden blocks. The projection matrices of six images were recovered using the non-linear technique presented in this paper. The geometry of the image segments was entered manually using equation 22. The original image is shown in figure 11 and the rendered image with the block removed is shown in figure 12.

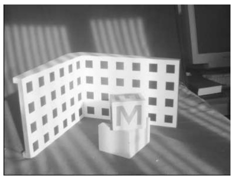

Figure 11: Image Editing: The original image.

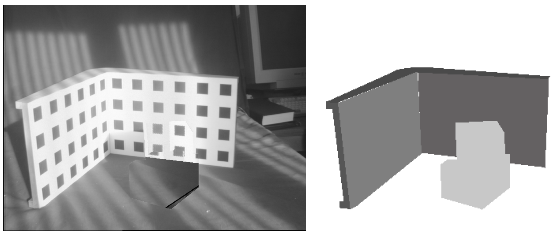

Figure 12: Image Editing: The block removed.

9 Future Work

The next stage of the project is to automatically compute the image segmentation. This will be achieved by using robust techniques to compute the homographies between the surfaces in the image.

10 Conclusion

In this paper a new method for computing the projection matrices of an image sequence, has been presented. This approach is an extension of Hartley's algebraic minimisation technique. It is an attractive technique as it requires only twelve parameters to be optimised, regardless of the length of the image sequence. The algorithm has been tested on synthetic data, in the presence of noise. The projection matrices computed in the first half of this paper were then used for visualising three-dimensional models of image sequences. A second application, image editing, was also presented which made use of the calibrated image sequences.

Acknowledgements

This research is funded by Trinity College Cambridge, the Cambridge Commonwealth Trust and the Overseas Research Students (ORS) Award Scheme.

References

[1] R.I. Hartley. A linear method for reconstruction from lines and points. *Proc. 5th Int. Conf. on Computer Vision*, pages 882–887, 1995.

[2] L. Quan. Invariants of 6 points from 3 uncalibrated images. *Proc. 3rd European Conference on Computer Vision*, pages 459–470, 1994.

[3] A. Heyden. A common framework for multiple view tensors. In *Proc. 5th European Conference on Computer Vision*, pages 3–10, 1998.

[4] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *Proc. 6th Int. Conf. on Computer Vision*, pages 90–95, 1998.

[5] P.H.S. Torr and A. Zisserman. Robust parameterisation and computation of the trifocal tensor. *Image and Vision Computing*, 15(8):591–605, 1997.

[6] R.I. Hartley. Minimizing algebraic error in geometric estimation problems. *Proc. 6th Int. Conf. on Computer Vision*, pages 469–476, 1998.

[7] C.G. Harris and M. Stephens. A combined corner and edge detector. In *4th Alvey Vision Conference*, pages 147–151, 1988.

[8] O. Faugeras. *Three-dimensional computer vision - A geometric viewpoint.* MIT Press ISBN 0-262-06158-9, 1993.