

Using Combinations of Points, Lines and Conics to Estimate Structure and Motion *

Kalle Åström

Dept of Mathematics, Lund University

Box 118, S-221 00 Lund, Sweden

email: kalle@maths.lth.se

Abstract

The estimation of structure and motion from image sequences using point, line and conic correspondences is treated. Recent research has provided tools for obtaining good initial estimates of structure and motion using images of these features. These estimates are, however, not so accurate. In this paper it is shown how to obtain statistically optimal estimates of structure and motion using such image feature correspondences. The question of using proper weighting is important when different types of features are combined. Experiments with real data are used to evaluate every step of the algorithm.

1 Introduction

Using monocular image sequences taken by a camera that moves relative to a rigid scene, one can generally determine the parameters of structure and motion up to an unknown projective transformation, cf. [3]. If additional information is given, e.g. if the skew is known or if the internal parameters are known to be constant, then it is possible to determine the parameters of structure and motion up to a change of Euclidean coordinate system and scale, [7]. In order to understand this process it is convenient to think of this computation in three steps. Firstly, features are extracted in each image. Secondly, the interframe correspondences between the selected features are established. Thirdly, the structure and motion is calculated from these feature correspondences. Each step is important. The correspondence problem will not be treated here. Instead the paper will focus on the accurate estimation of structure and motion with emphasis on feature extraction and structure and motion estimation.

In [17] the problem of estimating structure and motion from line correspondences is discussed. They discuss both the question of obtaining an initial estimate of camera motion using the trilinear constraint for lines and the question of improving these estimates. Several authors discuss the problem of obtaining initial estimates for lines and points [4, 6, 5], and more

recently using points, lines and conics [9].

Previous work on conic based vision is mostly devoted to calculating structure from known motion and calculating motion from known structure, [10]. Some work has also been done on the correspondence problem for conics, [13, 12].

Two major contributions are made in this paper: the novel algorithms for bundle adjustment of configurations of points, lines and conics and the consistent treatment of uncertainty throughout the whole process. We would like to point out the importance of treating not only the geometric features but also their uncertainty in every step of the computation.

2 Notations

We will use extended or homogeneous coordinates throughout the paper for geometrical entities. Thus, a two-dimensional point (x, y) will be represented as a three-vector $\mathbf{x} = [x \ y \ 1]^T$. A direction $(\cos(\alpha), \sin(\alpha))$ in the two-dimensional plane will be represented by a three vector $\mathbf{n} = [\cos(\alpha) \ \sin(\alpha) \ 0]^T$. Similar notations will be used for three-dimensional points and directions. We will assume that the camera is an ideal pinhole camera. The projection of a three-dimensional point \mathbf{X} is conveniently represented by a 3×4 projection matrix \mathbf{P} , such that

$$\lambda \mathbf{x} = \mathbf{P} \mathbf{X} .$$

A line in the image will be represented by a vector $\mathbf{l} = [a \ b \ c]^T$. The points of the line fulfill $ax + by + c = \mathbf{l}^T \mathbf{x} = 0$. A conic in the image is represented by a symmetric 3×3 matrix \mathbf{C} and a quadric in space by a symmetric 4×4 matrix \mathbf{Q} . The points of the conic fulfill $\mathbf{x}^T \mathbf{C} \mathbf{x} = 0$ and the points of the quadric $\mathbf{X}^T \mathbf{Q} \mathbf{X} = 0$. A quadric \mathbf{Q} is projected to a conic \mathbf{C} according to

$$\lambda \mathbf{C} = \mathbf{P} \mathbf{Q} \mathbf{P}^T .$$

3 Images and Image Features

In this section we will describe how to estimate the three types of **image features** that have been used: image points, lines and conics. The representation of a point and its uncertainty is straightforward and

*This work has been done within then ESPRIT Reactive LTR project 21914, CUMULI and the Swedish Research Council for Engineering Sciences (TFR), project 95-64-222

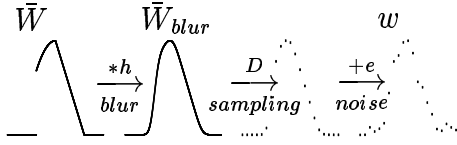


Figure 1: Illustration of the image acquisition model. The original intensity distribution \bar{W}_{ideal} is blurred by the kernel h . The blurred signal \bar{W} is then sampled to form a discrete signal w_0 . Finally noise is added. The result is the measured discrete signal $v_0 = w_0 + e_0$.

has been omitted. The representation of the other primitives and their uncertainty is presented. A brief discussion on how to estimate these features is also given.

Pixels

Each image feature is estimated using the information in the digital images. In a digital camera, the blurred image intensity distribution is typically measured by a CCD array. One can think of each pixel intensity as the weighted mean of the intensity distribution in a window around the ideal pixel position. Taking the weighted mean around a position is equivalent to convolution followed by ideal sampling. Due to quantisation and other errors, stochastic errors are introduced. The original intensity distribution \bar{W} is first blurred or convoluted with a low-pass kernel h , $\bar{W}_{\text{blur}} = h * \bar{W}$. Each pixel $w(i, j)$ is an estimate of this distribution at position (i, j) , i.e. $w(i, j) = \bar{W}_{\text{blur}}(i, j) + e(i, j)$, where the error $e(i, j)$ are independent stochastic variables with zero mean and standard deviation σ . This is illustrated in Figure 1.

The digital image of an almost ideal edge is also shown in Figure 2b. The object in the scene was a very sharp step edge between black and white. Notice that the image is less sharp and discrete. The particular camera used in Figure 1 measures the intensity with 8-bits. A number between 0 and 255 is obtained. Experiments have shown that the intensity errors $e(i, j)$ in each pixel for this particular camera can be modeled as independent noise with standard deviation $\sigma \approx 3$.

Edge detection

How accurate can the image position of the edge be estimated in good conditions such as in Figure 2b? In [2] it is shown that by proper interpolation of the discrete signal w it is possible to obtain a continuous signal $W = \bar{W} * g + E$, where the error E is a stationary random field. Furthermore it is possible to calculate the covariance of E . What this means in practice is that the smoothed intensity W and its derivatives can be estimated at arbitrary interpolated positions. This can be used to find the sub-pixel edge position in the following way: Define a search line

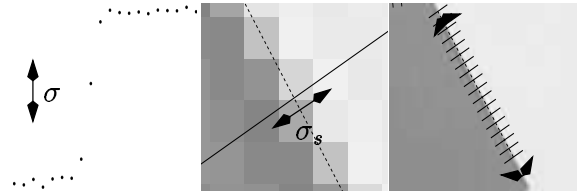


Figure 2: The uncertainty in intensity measurement causes uncertainty in each edge point position. These uncertainties in turn cause uncertainty in the extracted feature.

roughly perpendicular to the edge. Find the position along that line where the directional derivative of the interpolated intensity W is the highest. In practice one first calculates the directional derivative at a discrete number of positions along the search line, then finds the point where the directional derivative is the highest. This position is used as an initial estimate in a Newton Raphson algorithm to find the point with the highest directional derivative. This idea was used in [1] and it was shown how to estimate the standard deviation of the edge point estimate. Thus, the edge point \mathbf{p} and its uncertainty (the standard deviation σ_s in the direction \mathbf{n} of the search line) is calculated from the image w and its uncertainty (the variance in each intensity measurements σ^2).

Lines

The process of fitting lines to extracted edge points is well known. Figure 2c illustrates several search lines and the extracted edge point positions. Using these edge points and their uncertainty, the edge is estimated using the maximum likelihood method. Let \mathbf{x}_i be point number i , \mathbf{n}_i the direction of the search line and $\sigma_{s,i}$ the estimated standard deviation. The line parameters $\mathbf{l}_m = [a \ b \ c]^T$ are estimated so that $\mathbf{l}_m^T \mathbf{l}_m = 1$, $\mathbf{l}_m^T (\mathbf{x}_i + s_i \mathbf{n}_i) = 0$ and such that

$$\sum_i \frac{s_i^2}{\sigma_{s,i}^2}$$

is minimised.

The resulting line parameter \mathbf{l}_m lies on the surface of the sphere $\mathbf{l}_m^T \mathbf{l}_m = 1$. It is convenient to measure small deviations from \mathbf{l}_m in the tangent plane T_p to \mathbf{l}_m on the sphere. Thus, for lines 'close' to \mathbf{l}_m we use the following parametrisation

$$\mathbf{l}_{\text{local}} = P(\mathbf{l}) = \frac{\mathbf{l}}{\mathbf{l}_m^T \mathbf{l}}.$$

In this parametrisation $P(\mathbf{l}_m)$ can be regarded as a stochastic variable with covariance matrix C_1 . This matrix is estimated from the line fitting algorithm. Since the domain of P is two-dimensional, C_1 is (a 3×3 matrix) of rank 2. Using the Cholesky factorisation of the pseudo-inverse C_1^\dagger of C_1 we obtain the following factorisation, $L^T L = C_1^\dagger$, where L is a 2×3 matrix.

Then, the stochastic variable $LP(\mathbf{l}_m)$ has covariance matrix $LC_1L^T = I$. The two variables in $LP(\mathbf{l}_m)$ are uncorrelated and have unit standard deviation.

Conics

The process of fitting conics to extracted edge points is similar. The goal is to calculate the conic and its uncertainty from the edge points and their uncertainty. Figure 2c illustrates several search lines, the extracted edge point positions. Let \mathbf{x}_i be point number i , \mathbf{n}_i the direction of the search line and $\sigma_{s,i}$ the estimated standard deviation. The conic parameters $u_m = [a \ b \ c \ d \ e \ f]^T$ are estimated so that $u_m^T u_m = 1$,

$$(\mathbf{x}_i + s_i \mathbf{n}_i) \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}^{-1} (\mathbf{x}_i + s_i \mathbf{n}_i)^T = 0$$

and such that

$$\sum_i \frac{s_i^2}{\sigma_{s,i}^2}$$

is minimised. From the final result we obtain the measured conic parameters u_m and the covariance matrix C_u . Again since the conic parameters lie on the surface of a 6 dimensional sphere $u_m^T u_m = 1$, the covariance matrix has less than full rank. Small deviations are measured in the tangent plane T_p to the sphere at u_m . Similar to the case of lines, the uncertainty can be represented by a matrix L (a 6×6 matrix in this case).

4 Initial estimate

A rough estimate of the structure and motion parameters is obtained using either projective methods for points and lines using the trilinear tensor, cf. [16], shape based factorisation methods for points, cf. [15], or affine method for points, lines and conics, cf. [8].

The estimate is reasonably correct, but the errors between the reprojected features and the estimated features are severe (several pixels).

5 Bundle Adjustment

The idea of the bundle adjustment method is to optimize the structure and motion parameters so that the residual between the measured features and the reprojected features is small in some sense. For each feature type it will be shown: how to parametrise local changes in structure, how to reproject, and how to calculate the weighted residuals and their derivatives with respect to changes in structure and motion.

Points

Three dimensional points are parametrised in extended coordinates as $\mathbf{X} = [X \ Y \ Z \ 1]^T$.

Changes in \mathbf{X} are parametrised as $\mathbf{X}(\Delta x) = [X + \Delta x(1) \ Y + \Delta x(2) \ Z + \Delta x(3) \ 1]^T$. The projection \mathbf{x}_p and its derivatives are given by $\mathbf{x}_p = \mathbf{P}\mathbf{X}$, $\partial \mathbf{x}_p = \partial \mathbf{P}\mathbf{X} + \mathbf{P}\partial \mathbf{X}$. The weighted residual is measured as

$$r = L \left(\frac{\mathbf{x}_p}{\mathbf{x}_p(3)} - \mathbf{x}_m \right) .$$

Lines

Three dimensional lines are parametrised using two points \mathbf{X}_1 and \mathbf{X}_2 in extended coordinates. Changes in \mathbf{X}_1 and \mathbf{X}_2 are parametrised as $\mathbf{X}_1(\Delta x) = \mathbf{X}_1 + \Delta x(1)V_1 + \Delta x(2)V_2$ and $\mathbf{X}_2(\Delta x) = \mathbf{X}_2 + \Delta x(3)V_1 + \Delta x(4)V_2$, where V_1 and V_2 are orthogonal directions orthogonal to the direction between \mathbf{X}_1 and \mathbf{X}_2 . The projected line \mathbf{l}_p and its derivatives are given by

$$\mathbf{l}_p = (\mathbf{P}\mathbf{X}_1) \times (\mathbf{P}\mathbf{X}_2) ,$$

$$\partial \mathbf{l}_p = (\partial \mathbf{P}\mathbf{X}_1 + \mathbf{P}\partial \mathbf{X}_1) \times (\mathbf{P}\mathbf{X}_2) + (\mathbf{P}\mathbf{X}_1) \times (\partial \mathbf{P}\mathbf{X}_2 + \mathbf{P}\partial \mathbf{X}_2)$$

where \times denotes vector product. The weighted residual is measured as

$$r = L \left(\frac{\mathbf{l}_p}{\mathbf{l}_p^T \mathbf{l}_p} - \mathbf{l}_m \right) .$$

Conics

Three dimensional quadrics are parametrised using a 4×4 symmetric matrix \mathbf{Q} . Changes in \mathbf{Q} are parametrised as

$$\mathbf{Q}(\Delta x) = \mathbf{Q} + \begin{bmatrix} \Delta x(1) & \Delta x(2) & \Delta x(3) & \Delta x(4) \\ \Delta x(2) & \Delta x(5) & \Delta x(6) & \Delta x(7) \\ \Delta x(3) & \Delta x(6) & \Delta x(8) & \Delta x(9) \\ \Delta x(4) & \Delta x(7) & \Delta x(9) & \Delta x(10) \end{bmatrix} .$$

The projected conic u_p and its derivatives are given by

$$\mathbf{C}_p = \mathbf{P}\mathbf{Q}\mathbf{P}^T, \quad \partial \mathbf{C}_p = \mathbf{P}\partial \mathbf{Q}\mathbf{P}^T + \partial \mathbf{P}\mathbf{Q}\mathbf{P}^T + \mathbf{P}\mathbf{Q}\partial \mathbf{P}^T .$$

From the conic \mathbf{C}_p and its derivatives the conic vector u_p and its derivatives are extracted. The weighted residual is measured as

$$r = L \left(\frac{u_p}{u_m^T u_p} - u_m \right) .$$

Adjustment

The idea of the bundle adjustment is straightforward. Calculate a vector r containing all weighted residuals and its derivatives with respect to changes in all structure parameters and all motion parameters. The weighted residuals have been normalised so that they are approximately uncorrelated with zero mean and unit variance, using the true values of the motion parameters. Minimising r in a least squares sense is therefore statistically optimal in the first order.

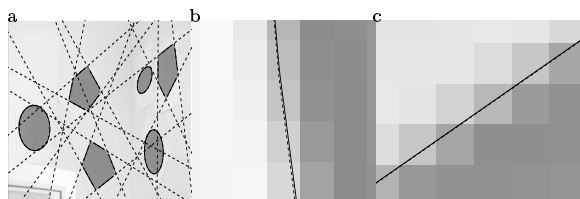


Figure 3: Figure 3a shows one image in the experiments. Solid lines represent measured features and dashed line represent reprojected features. A close-up is shown of part of a conic in Figure 3b and part of a line in Figure 3c.

Outlier detection

Since the residuals are normalised, outliers will result in residuals significantly larger than one. This can be used to detect and remove outliers.

6 Experimental Validation

Figure 3a shows one of 7 images that were used in the experiment with overlaid measured lines and conics and reprojected lines and conics. Figure 3b shows a close-up centered at a curved part of a conic. The solid curve illustrates the measured conic and the dashed line the reprojected conic. The residuals are of the order of 0.1 pixels. Similarly Figure 3c shows a close-up centered at a detected line. The solid curve illustrates the measured line and the dashed line the reprojected line. Again the residuals are small.

7 Conclusions and Discussions

Conics and lines are important image features. They contain more information than points. Minimal cases for structure and motion using conics involve less features. This is an advantage in situations where less features are available. It is easier to detect and track conics than points. They are more robust and stable features. This has led to an upsurge in research on conic based vision. However, when using conics and in particular in combination with other features, it is important to treat the uncertainty in a proper manner. In this paper we have focused on the systematic treatment of features and their uncertainty all the way from image acquisition to edge and feature detection and finally to their proper weighting in a bundle adjustment method. This treatment makes it possible to estimate structure and motion in a statistically optimal way using a combination of points, lines and conics. The theory predicts that edge points can be extracted with high precision (approximately 0.1 pixels) and this is verified by the results from the bundle adjustment.

Acknowledgements

The author thanks Anders Heyden, Fredrik Kahl and Kent Stråhlén for valuable discussions.

References

- [1] K. Åström and A. Heyden. Stochastic analysis of sub-pixel edge detection. In *Proc. International Conference on Pattern Recognition, Vienna, Austria, 1996*.
- [2] K. Åström and A. Heyden. Stochastic analysis of image acquisition and scale-space smoothing. In J. Sporring, M. Nielsen, L. Florack, and P. Johansen, editors, *Gaussian Scale-Space Theory*. Kluwer Academic Publishers, 1997.
- [3] O. D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proc. 2nd European Conf. on Computer Vision, Santa Margherita Ligure, Italy*, pages 563–578. Springer-Verlag, 1992.
- [4] O. D. Faugeras and B. Mourrain. On the geometry and algebra on the point and line correspondences between n images. In *Proc. 5th Int. Conf. on Computer Vision, MIT, Boston, MA*, pages 951–956. IEEE Computer Society Press, 1995.
- [5] R. I. Hartley. A linear method for reconstruction from lines and points. In *Proc. 5th Int. Conf. on Computer Vision, MIT, Boston, MA*, pages 882–887, 1995.
- [6] R. I. Hartley. Lines and points in three views and the trifocal tensor. *Int. Journal of Computer Vision*, 22(2):125–140, 1997.
- [7] A. Heyden and K. Åström. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Proc. Conf. Computer Vision and Pattern Recognition, 1997*.
- [8] F. Kahl and A. Heyden. Structure and motion from points, lines and conics with affine cameras. In *Proc. 5th European Conf. on Computer Vision, Freiburg, Germany, 1997*. submitted to ECCV'98.
- [9] F. Kahl and A. Heyden. Using conic correspondences in two images to estimate the epipolar geometry. In *Proc. 6th Int. Conf. on Computer Vision, Mumbai, India, 1998*.
- [10] S. Ma. Conic-based stereo, motion estimation, and pose determination. *Int. Journal of Computer Vision*, 10(1):7–25, 1993.
- [11] L. Quan. Invariant of a pair of non-coplanar conics in space: Definition, geometric interpretation and computation. In *Proc. 5th Int. Conf. on Computer Vision, MIT, Boston, MA*, pages 926–931, 1995.
- [12] L. Quan. Conic reconstruction and correspondence from two views. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(2):151–160, Feb. 1996.
- [13] L. Quan and K. T. Affine structure from line correspondences with uncalibrated affine cameras. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(8), Aug. 1997.
- [14] C. R. Rao. *Linear Statistical Inference and Its Application*. John Wiley & Sons, Inc., 1965.
- [15] G. Sparr. Simultaneous reconstruction of scene structure and camera locations from uncalibrated image sequences. In *Proc. International Conference on Pattern Recognition, Vienna, Austria, 1996*.
- [16] M. E. Spetsakis and J. Y. Aloimonos. A unified theory of structure and motion. In *Proc. DARPA Image Understanding Workshop, 1990*.
- [17] J. Weng, T. Huang, and N. Ahuja. Motion and structure from line correspondences: Closed-form solution, uniqueness, and optimization. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(3), 1992.