

# A Statistical Approach to Structure and Motion from Image Features

Kalle Åström, Fredrik Kahl, Anders Heyden, Rikard Berthilsson

Dept of Mathematics, Lund University, Box 118, S-221 00 Lund, Sweden

**Abstract.** The estimation of structure and motion from image sequences using corresponding points, lines, conics and structured patches is treated. Recent research has provided good tools for obtaining good initial estimates of structure and motion using point, line, conic and curve correspondences. These estimates are, however, not so accurate. In this paper it is shown how to obtain statistically optimal estimates of structure and motion using a combination of such image feature correspondences. The question of using proper weighting is important when different types of features are combined. We show how weights can be chosen in a statistical optimal sense. Experiments with real data are used to evaluate every step of the algorithm.

**Keywords:** Bundle adjustment, Points, Lines, Conics, Patches.

## 1 Introduction

This paper deals with the problem of estimating structure (3D-reconstruction) and motion (camera movement) from images taken by uncalibrated cameras. All kinds of features are used in a statistically unified manner; points, lines, conics and textured, planar patches. 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. The paper will focus on the accurate estimation of structure and motion with emphasis on feature extraction and structure and motion estimation. The correspondence problem will not be treated here.

In [17] the problem of estimating structure and motion from line correspondences is analysed. There, it is discussed both the question of obtaining an initial estimate and the question of improving these estimates. Several authors discuss the problem of obtaining initial estimates for lines and points [6, 8, 7] and 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, [11]. However, in [10] the use of conic correspondences to estimate the epipolar geometry is treated. Some work has also been done on the correspondence problem for conics,

[13, 12]. Finally, the problem of computing structure and motion by using general curves is treated in [4, 5].

The problem of obtaining a statistically optimal solution to the structure from motion problem, have traditionally been solved using the technique of so called bundle adjustments, see [3]. However, it is only possible to use points in the traditional approach to bundle adjustments. An attempt to use general curves has been made in [5].

Two major contributions are made in this paper: the novel algorithms for bundle adjustment of configurations of points, lines, conics and textured patches 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 Background

We will use extended or extended coordinates throughout the paper for geometrical entities. Thus, a point  $(x, y) \in \mathbb{R}^2$  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 points and directions in  $\mathbb{R}^3$ . We will assume that the camera is an ideal pinhole camera. The projection of an object point  $\mathbf{X} \in \mathbb{R}^3$  onto an image point  $\mathbf{x} \in \mathbb{R}^2$  is conveniently represented by a  $3 \times 4$  projection matrix  $\mathbf{P}$ , such that

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

where  $\lambda \in \mathbb{R}$ . A line in the image will be represented by a vector  $\mathbf{l} = [a \ b \ c]^T$ . The points of the line fulfil  $ax + by + c = \mathbf{l}^T \mathbf{x} = 0$ . A conic in the image is represented by a symmetric non-singular  $3 \times 3$  matrix  $\mathbf{C}$  and a quadric in space by a symmetric non-singular  $4 \times 4$  matrix  $\mathbf{Q}$ . The points of the conic fulfill  $\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} = 0$  and the points of the quadric  $\mathbf{X}^T \mathbf{Q}^{-1} \mathbf{X} = 0$ . A quadric  $\mathbf{Q}$  is projected to a conic  $\mathbf{C}$  according to

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

A general curve will be represented as a cubic b-spline containing. Using the control points, the curve has a natural parametrisation  $t \mapsto c(t)$ . A textured patch will be represented by four points, defining a quadrangle and a matrix defining the gray-level values inside the quadrangle.

## 3 Image Features

In this section we will describe how to estimate the different types of image features that have been used. The representation of these primitives and their uncertainty is presented.

### Edge detection:

How accurate can the image position of the edge be estimated in good conditions

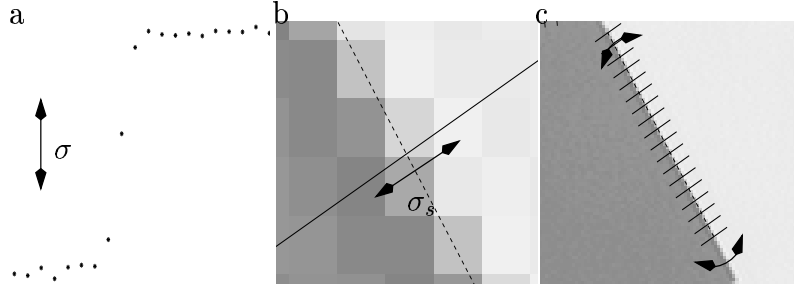
such as in Figure 3a? In [2, 1] and it was shown how to estimate the standard deviation of the edge point estimate. Thus, the edge point  $\mathbf{p}$  and its uncertainty (the standard deviation  $\sigma_s$  in the direction  $\mathbf{n}$  of the search line) is calculated from the image  $w$  and its uncertainty (the variance in each intensity measurements  $\sigma^2$ ). In practice, it is assumed that the standard deviation  $\sigma$  for the error in each pixel intensity is constant.

**Lines:**

The process of fitting lines to extracted edge points is well known. Figure 1c illustrates several search lines and the extracted edge point positions. Using these edge points and assuming their uncertainty is normally distributed, the line is estimated according to 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.



**Fig. 1.** The uncertainty in intensity measurement causes uncertainty in each edge point position. These uncertainties in turn cause uncertainty in the extracted feature.

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

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

In this parametrisation  $P(\mathbf{l})$  is 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  $3 \times 3$  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  $2 \times 3$  matrix. Then, the stochastic variable  $LP(\mathbf{l})$ , has covariance matrix

$$LC_1L^T = I .$$

The two variables in  $LP(\mathbf{l})$  are uncorrelated and have unit standard deviation. In the bundle adjustment it is necessary to measure differences between the measured line  $\mathbf{l}_m$  and the reprojected line  $\mathbf{l}_p$ . The following weighted residual

$$r(\mathbf{l}_p) = L(P(\mathbf{l}_p) - \mathbf{l}_m)$$

is used. Thus, the line and its uncertainty is calculated from the edge points and their uncertainty.

**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. The conic 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 conic parameters  $u_m = [a b c d e f]^T$  are estimated such 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  $u_m$ . Thus for conics  $u$  'close' to  $u_m$  we use the following parametrisation instead

$$u_{local} = P(u) = \frac{u}{u_m^T u} .$$

In this parametrisation  $P(u)$  is a stochastic variable with covariance matrix  $C_u$  which is estimated from the conic fitting algorithm. Since the domain of  $P$  is five-dimensional,  $C_u$  is (a  $6 \times 6$  matrix) of rank 5. Using the Cholesky factorisation of the pseudo-inverse of  $C_u$  it is possible to factorise  $C_u$  according to,  $L^T L = C_u^\dagger$ , where  $L$  is a  $5 \times 6$  matrix. Then the stochastic variable  $LP(u)$  has the covariance matrix

$$LC_uL^T = I .$$

Thus the conic and its uncertainty, is calculated from the edge points and their uncertainties.

**Patches:**

The process of extracting textured patches is somewhat different from the other features. It is assumed that a patch is planar, but it is only the relative positions of patches that can be measured in the images. A patch and its uncertainty is extracted as follows. First, a textured area is chosen in the first image and

identified in the other images by correlation. From these matches, a source patch is defined as a quadrangle together with the gray-levels inside the quadrangle. Note that since the patch is assumed to be planar, there is a homography from the source patch to the image patch. This homography can be represented by a  $3 \times 3$  matrix  $H$ , defined up to scale. The gray-levels of the source patch and the homographies between the source patch and each image are chosen to minimise the residuals in intensity.. Similar to the case of points, lines and conics the covariance is estimated and used as weights in the optimization.

## 4 Initial estimates of structure and motion

An initial 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], shape based factorisation methods for curves, cf. [4, 5], or affine method for points, lines and conics, cf. [9]. The estimate is reasonably correct, but the errors between the reprojected features and the estimated features are sometimes large (several pixels).

## 5 Bundle Adjustment

The idea of the bundle adjustment method is to optimise 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, i.e. how to calculate the appearance of the feature.
- how to calculate the weighted residuals and their derivatives with respect to changes in structure and motion.

### Points:

Points in  $\mathbb{R}^3$  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:

Lines in  $\mathbb{R}^3$  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

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

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

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

**Conics:**

Quadrics in  $\mathbb{R}^3$  are parametrised using a  $4 \times 4$  symmetric matrix  $\mathbf{Q}$ . Changes in  $\mathbf{Q}$  are parametrised as  $\mathbf{Q}(\Delta x) = \mathbf{Q} + \Delta x$  where  $\Delta x$  is a symmetric matrix containing 10 different elements. The projected conic  $u_p$  and its derivatives are given by

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

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

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

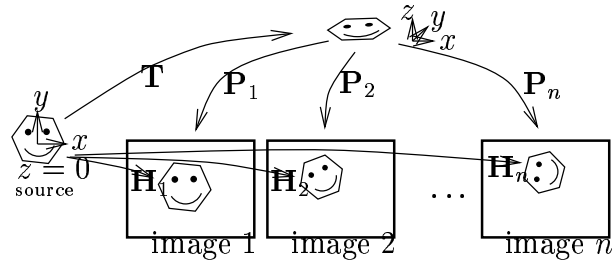
**Patches:**

Three dimensional patches are parametrised by a homography from the source patch to a 3D-plane and the position of this plane in space. In total, there are 11 degrees of freedom which can be represented by a  $4 \times 3$  matrix  $\mathbf{T}$  defined up to scale, see Figure 2. The homography of the projected patch  $H_p$  and its derivatives are given by

$$\mathbf{H}_p = \mathbf{PT}, \quad \partial \mathbf{H}_p = \mathbf{P}\partial \mathbf{T} + \partial \mathbf{PT} .$$

From the above equation, the homography vector  $h_p$  and its derivatives  $\partial h_p$  are calculated. The weighted residuals are measured as

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

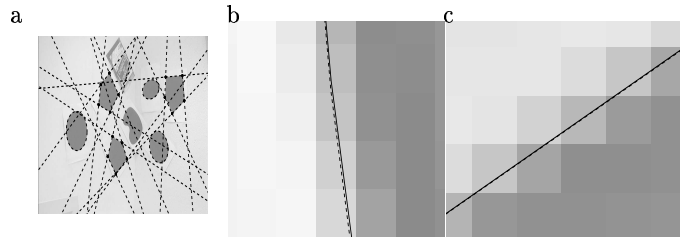


**Fig. 2.** Textured patches

### 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 such 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 approximation.

## 6 Experimental Validation



**Fig. 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.

Figure 3a shows one of 5 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

So far, most research in computer vision has concentrated on point features. But there are other important image features, like lines, conics and patches. They contain more information than points. It is often easier to detect and track these features. They are more robust and stable features. However, when using these features and in particular in combination, 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, conics and even textured patches. 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.

## References

1. K. Åström and A. Heyden. Stochastic analysis of sub-pixel edge detection. In *Proc. International Conference on Pattern Recognition, Vienna, Austria*, pages 86–90, 1996.
2. K. Åström and A. Heyden. Stochastic analysis of image acquisition and scale-space smoothing. In J. Sparring, M. Nielsen, L. Florack, and P. Johansen, editors, *Gaussian Scale-Space Theory*. Kluwer Academic Publishers, 1997.
3. K. B. Atkinson. *Close Range Photogrammetry and Machine Vision*. Whittles Publishing, 1996.
4. R. Berthilsson and K. Åström. Reconstruction of 3d-curves from 2d-images using affine shape methods for curves. In *Proc. Conf. Computer Vision and Pattern Recognition*, 1997.
5. R. Berthilsson, K. Åström, and A. Heyden. Projective reconstruction of 3d-curves from its 2d-images using error models and bundle adjustments. In *Proc. 10th Scandinavian Conference on Image Analysis*, pages 581–588, 1997.
6. O. D. Faugeras and B. Murrain. 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.
7. 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.
8. R. I. Hartley. Lines and points in three views and the trifocal tensor. *Int. Journal of Computer Vision*, 22(2):125–140, 1997.
9. 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.
10. 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*, pages 761–766, 1998.
11. S.D. Ma. Conic-based stereo, motion estimation, and pose determination. *Int. Journal of Computer Vision*, 10(1):7–25, 1993.
12. L. Quan. Conic reconstruction and correspondence from two views. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(2):151–160, February 1996.
13. L. Quan and T. Kanade. Affine structure from line correspondences with uncalibrated affine cameras. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(8), August 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*, pages 328 – 333, 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.S. 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.