Structure and motion estimation from complex features in three views

Björn Johansson, Magnus Oskarsson, Kalle Åström
Centre for Mathematical Sciences, Lund University
P.O. Box 118, SE -221 00 Lund, Sweden
{bjornj,magnuso,kalle}@maths.lth.se

Abstract

In this paper we introduce the notion of a quiver, which is a feature based on a point and a number of directions. We investigate the constraints different kinds of quivers pose on the camera geometry. In particular we study three types of quivers, having one, two and three directions respectively. For these quivers we investigate structure and motion estimation for two minimal cases. For 1-quivers, three such features seen in three affine views yields a (unique) linear solution. For three 3-quivers in three uncalibrated projective views we get up to twelve solutions, but simulations show that in most cases we get a unique solution. We also study two-quivers seen in three views, where simulations show that such features give more stable estimation of the trifocal tensor, as opposed to only using line or point information.

1. Introduction

When one considers an urban scene there are in general many buildings and structures. These structures often contain corners. In this paper we propose the use of features consisting of a point and a number of directions from this point. We will call such features *quivers*. A quiver with three directions is essentially what we would call a corner. Quivers with one or two directions can be used if not the whole of the corner is visible in all images or if the corner is located on a planar surface.

Often when one extract points from an image one gets gradient information in a neighborhood of the points. This information can be used to extend a point feature to a quiver.

An accurate way of estimating the position of a point is to intersect a number of lines that cross in that point in the image. If these lines correspond over images, one might as well use the line information and consider the point and the lines as a quiver.

In this paper we will study structure and motion estimation from only image quiver features. One difficult step in the reconstruction process is finding correspondences between features in different views. We will assume that these correspondences are known and concentrate on the estimation of the geometry. Correspondences between quivers is somewhat more difficult than for just points or lines. When we have correspondence between two quivers we assume that the correspondences between the directions in the quivers are also known. This means, for example, that for two quivers with three directions, there are six possible ways of correspondence. However in many instances one can eliminate many of these. In the case of a corner on a solid building these six possible ways reduce to three.

In this paper we investigate two minimal cases for reconstruction using quiver features. By a minimal case we mean that omission of some data would give an infinite number of solutions. Solving minimal cases to perform 3D reconstruction is of both theoretical and practical importance. Algebraic solutions obtained from the minimal cases can be used to bootstrap robust estimation algorithms such as RANSAC or LMS schema [5, 8, 10].

2. Preliminaries

A point in space, represented in homogeneous coordinates by $\mathbf{X}$, is projected onto the image plane according to the equation

$$\lambda \mathbf{x} = P\mathbf{X}. \tag{1}$$

Here $\mathbf{x}$ are the homogeneous coordinates of the image point, P is the camera matrix and λ is a scalar. The camera matrix is defined up to scale and can be decomposed as $P = KR[I \,|\, -\mathbf{t}]$, where K is a right triangular 3×3 matrix representing the calibration of the camera, R is a orthogonal matrix representing the orientation of the camera and $\mathbf{t}$ is the focal point. In the projective camera model, the camera matrix P may be any 3×4 matrix. If K is known the camera is said to be calibrated. If P has the form

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & p_{34} \end{bmatrix},$$

then the affine camera model is used. This corresponds to a projective camera model in the case when the focal length is infinite.

A common problem in computer vision is to estimate the structure of a scene and the motion of the cameras using only image information. Consider a line in 3D space. The line is projected to an image line. A line in a plane may be represented in dual form in homogeneous coordinates as $\mathbf{l} = [a\ b\ c]^T$. The equation of the line is $\mathbf{l}^T\mathbf{u} = 0$, where $\mathbf{u} = [x\ y\ 1]^T$ denote points on the line.

The problem of determining the 3D line from its image lines $\mathbf{l}^i$, with known camera matrices P_i can be solved by intersecting the planes $P_i^T\mathbf{l}^i$. A necessary condition for the planes to meet in a line is that

$$\mathrm{rank}[P_1^T\mathbf{l}^1 \ldots P_m^T\mathbf{l}^m] = 2, \tag{2}$$

i.e. each 3×3 minor of the $4 \times m$ matrix above vanishes. These constraints may be formulated in terms of elements of the trifocal and quadrifocal tensors. Consider three cameras $\{P_1, P_2, P_3\}$ and three corresponding lines $\{\mathbf{l}^1, \mathbf{l}^2, \mathbf{l}^3\}$. Using Einsteins' summation convention the constraints from (2) may be written as

$$\mathbf{l}^3 \sim T_i^{jk} l_j^1 l_k^2 \Leftrightarrow \mathbf{l}^3 \times T_i^{jk} l_j^1 l_k^2 = 0.$$

Here $\sim$ denotes equality up to scale and T_i^{jk} denotes elements in the trifocal tensor, defined as

$$T_i^{jk} = \epsilon_{ii'i''} \det \begin{bmatrix} P_1^{i'} \\ P_1^{i''} \\ P_2^j \\ P_3^k \end{bmatrix},$$

where ϵ_{ijk} denotes the permutation symbol, i.e. $\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1$, $\epsilon_{321} = \epsilon_{213} = \epsilon_{132} = -1$ and $\epsilon_{ijk} = 0$ if two indices are equal.

In a similar way the four constraints from three corresponding points, $\{p_1, p_2, p_3\}$ may be expressed as

$$p_3^i T_i^{jk} l_j^1 l_k^2 = 0,$$

where l_1, l_2 are chosen as lines passing through the corresponding point in views one and two. Different choices of these lines give in total four linearly independent constraints on the tensor.

The necessary and sufficient conditions for the trifocal tensor for projective cameras are given in e.g. [4, 3, 7] and for an affine camera in [1]. The same analysis can be performed on four images.

We define a new feature, n-quiver, as a point and n directions from this point, cf. Figure 1. Counting the degrees of freedom and the image information, we get that corresponding points in three images give three constraints on the geometry and five constraints in four images. Corresponding lines give two and four constraints in three and four views respectively. However when the point is on the line the constraints are dependent.

In Table 2 the number of constraints different corresponding features give on three and four images is given. In order to understand how much information is needed in order to solve the structure and motion problem for different camera models using different kinds of features, the total number of degrees of freedom for different camera model systems is shown in table 2. The degrees of freedom are calculated as the number of cameras multiplied with the number of unknowns in the current camera model minus the degrees of freedom in the solution parameter space, cf. [6]. Comparing the number of constraints and degrees of freedom a number of minimal cases were found, i.e. cases where the image data exactly constrains the geometry. We have concentrated on two of these problems, that are of most practical use. The first case is three 1-quivers in three affine views and the second is three 3-quivers in three projective views.

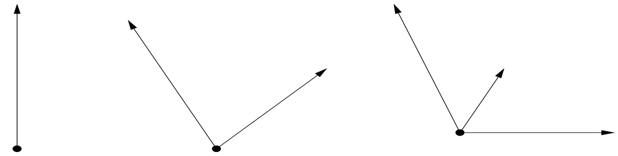

Figure 1: Quivers with one, two and three directions.

feature	3 images	4 images
line	2	4
point	3	5
point + direction	4	7
point + 2 directions	5	9
point + 3 directions	6	11

Table 1: The number of constraints different corresponding features give.

3. A point with one direction

In this section the feature defined by a point and one direction is considered. A correspondence between such features gives four and seven constraints on the geometry for three and four images respectively according to Table 2. One case of such features is particularly interesting, namely three such quivers seen in three affine views. This is a minimal case, and it turns out that it can be solved linearly with

camera model	3 images	4 images
projective	3*11-15=18	4*11-15=29
affine	3*8-12=12	4*8-12=20
calibrated	3*6-7=11	4*6-7=17

Table 2: The total number of degrees of freedom for different camera model systems.

a unique solution.

3.1. Problem formulation and solution

A quiver with one direction seen in three affine views gives four constraints on the camera geometry. And since three affine cameras have twelve degrees of freedom according to Table 2 this is a minimal case. A point in three views gives essentially three constraints on the camera geometry but gives four constraints on the trifocal tensor, cf. [6]. Similarly it turns out that a 1-quiver gives five constraints on the affine trifocal tensor.

The necessary and sufficient conditions for the trifocal tensor for projective cameras are given in [4, 3, 7]. For a trifocal tensor to be affine there are 11 linear of these constraints turn linear,

$$\begin{gathered} T_1^{13}=0,\ T_1^{23}=0,\ T_1^{33}=0,\ T_1^{32}=0,\\ T_1^{31}=0,\ T_2^{13}=0,\ T_2^{23}=0,\ T_2^{33}=0,\\ T_2^{32}=0,\ T_2^{31}=0,\ T_3^{33}=0. \end{gathered} \tag{3}$$

These 11 constraints plus the 15 linear constraints given by the quivers is enough to linearly estimate the trifocal tensor, since the trifocal tensor has 27 entries and is determined up to scale.

3.2. An experiment with real data

In Figure 3.2 three views with three 1-quivers are shown. The quivers were extracted from the images manually. The affine trifocal tensor was then estimated linearly according to the description in the previous section. In Figure 3.2 one can also see a number of points that were extracted from the images to test the estimated solution. The trifocal tensor was used to intersect the extracted points to estimate the structure of the points. These points were then projected in the images using the estimated tensor. The original points are shown as asterisks and the reprojected points as circles. Most errors are on the order of a pixel. Points further away from the used quivers are more poorly estimated.

4. A point plus two directions

In this section the 2-quiver is considered. From Table 2 we see that corresponding 2-quivers give five and nine constraints respectively on the geometry of three and four images. There are no minimal cases for any of the camera models in Table 2. However, the 2-quiver case is particularly interesting since many algorithms for finding point correspondences in images also give information about two directions in every point. For example, to find corner points with high precision the intersection of two lines corresponding to the edges around the corner is estimated.

Figure 2: Three images with three extracted quivers shown in each image. Also shown are a number of points used to test the estimated solution. Original points are shown as '*' and reprojected points as 'o'.

4.1. An almost minimal case

The information in four 2-quivers viewed by three projective cameras gives a system that is slightly over determined. There are 20 equations and 18 unknowns. One 2-quiver gives six linear constraints on the trifocal tensor, so we will have a $26 - 6 * 4 = 2$ parameter solution. We can use the necessary constraints on the tensor given in [7] to solve the parameters. These constraints are all of total degree three in the two parameters. If two of these non-linear constraints are used, we get $3^2 = 9$ solutions. These solutions can then be verified using the other non-linear constraints to find the unique solution.

4.2. Comparing reprojection errors

A simulation was conducted to compare the accuracy in the trifocal tensor estimation using point-, line- and 2-quiver-correspondences in three images. In this simulation we envision a scenario where points are extracted from the images as intersection of lines.

A number of lines that intersects pairwise were randomly placed in 3D and projected into three images. Then noise was added to the image lines. First the trifocal tensor was computed using only point correspondences. The points were given as the intersections of the pair of lines. Then the trifocal tensor was computed using only line correspondences, and finally the trifocal tensor was computed using the 2-quivers given by the point of intersection and two directions on the line. To compare the different tensors the reprojection error of a lattice was calculated using the true motion and the motions derived from the estimated tensors. The reprojection error for the tensors with different amount of noise on the image lines is shown in Figure 3 for 15 and 25 features. From the figure it is clear that the performance is best for the 2-quivers. This is perhaps not so surprising since we have more information in this case, but it shows that if both point and line information is available one should use quiver features to stabilize the estimation. It can also be seen from the two figures that 15 quivers seem to give better results than 25 points or lines (that give about the same number of constraints on tensor as 15 quivers do).

5. A point with three directions

In this section we concentrate our efforts on the case of features consisting of a point with three directions. We will in particular study the minimal case of three such features viewed in three projective images.

A feature with one point and three directions gives in three images six constraints on the camera geometry, cf. Table 2. Given three such features we have 18 constraints, which is exactly the same as the freedom in three uncalibrated projective cameras. One could try to solve the problem using the trifocal tensor as in the previous section. Each

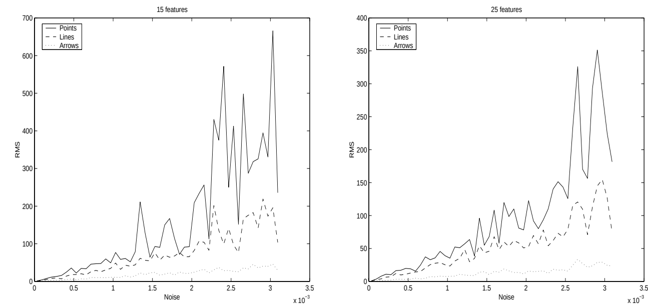

Figure 3: Reprojection error for motion estimated using points (solid), lines (dashed) and 2-quivers (dotted) for 15 features to the left and 25 features to the right.

feature would then give seven linear constraints on the tensor which would give the tensor up to a five parameter family of solutions. The necessary constraints on the tensor may then be used to solve for the parameters. However, choosing a specific parameterization of the cameras leads to a polynomial system of lower degree than the one given by the necessary constraints on the tensor.

5.1. The problem statement

We will solve for the camera geometry by a special parameterization of the cameras. We will fix the coordinate system in space by the three quivers. Introduce a projective coordinate system such that the 3 points in space are assigned to the projective coordinates

$$\begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \tag{4}$$

and the corresponding 3 image points in each image are assigned to

$$\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{5}$$

We have fixed nine degrees of the projective structure using the three points. The six remaining degrees of freedom in space are determined by specifying one of the directions in each quiver. These are given by specifying the following three points on each line,

$$\begin{bmatrix} \mathbf{X}'_1 & \mathbf{X}'_2 & \mathbf{X}'_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{bmatrix}. \tag{6}$$

Only two degrees of freedom remain in each image, so we can only specify two of the three corresponding lines. The three corresponding lines in each image are chosen as:

$$\begin{bmatrix} \mathbf{l}^1 & \mathbf{l}^2 & \mathbf{l}^3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & l_x \\ 1 & 0 & l_y \\ -1 & -1 & 0 \end{bmatrix}. \tag{7}$$

Using this choice of coordinates we get a special form of camera matrices

$$\mathbf{P} = \begin{bmatrix} l_y/l_x - a & -l_y/l_x & a & 0 \\ 0 & b & -1 & 1-b \\ 0 & b & a & 0 \end{bmatrix}. \tag{8}$$

with (a, b) unknown.

Each camera is now parameterized with two unknowns. We have used the point and one of the directions in each 3-quiver in our parameterization. The remaining $2 \cdot 3$ feature lines are used to solve for the six parameters (a, b, c, d, e, f). Using the rank constraints, cf. (2), we get one polynomial constraint from each line. Due to our special choice of coordinate systems in the images, these turn out to be of total degree three for the lines in quiver one and two, and of degree two for the lines in quiver three. The polynomials have the following structure:

$$\left\{ \begin{array}{l} *bde + *adf + *bcf + *af + *cf + *df + *de + \ldots \\ \ldots + *ad + *bf + *bc + *be + *bd = 0 \\ *bde + *adf + *bcf + *af + *cf + *df + *de + \ldots \\ \ldots + *ad + *bf + *bc + *be + *bd = 0 \\ *bce + *ade + *acf + *af + *cf + *ce + *de + \ldots \\ \ldots + *ad + *ae + *bc + *be + *ac = 0 \\ *bce + *ade + *acf + *af + *cf + *ce + *de + \ldots \\ \ldots + *ad + *ae + *bc + *be + *ac = 0 \\ *af + *cf + *de + *ad + *bc + *be + *a + *b + \ldots \\ \ldots + *c + *d + *e + *f = 0 \\ *af + *cf + *de + *ad + *bc + *be + *a + *b + \ldots \\ \ldots + *c + *d + *e + *f = 0 \end{array} \right. \tag{9}$$

Here $*$ denotes a known coefficient depending only on image data.

5.2. The solution

We have to solve a system of six polynomials in six variables of which four are of total degree three, and two are of total degree two. A system of this type may have up to $3^4 \cdot 2^2 = 324$ solutions according to the theorem of Bezout, cf. [2]. This bound is true in the case of a dense system with general coefficients. A better bound on the number of solutions is given by the so called *mixed volume* of a system, cf. [2]. Calculating this bound for the system in (9) gives a mixed volume equal to 42, so we have at most 42 solutions to our system.

In order to solve our equations we have used a polynomial solver called PHC, which is described in [9]. This program starts with calculating the mixed volume of the system. After the calculation of the mixed volume the solver proceeds by constructing a more easily solved system with the same structure as the original problem. It solves this system and the 42 solutions are propagated to the solution to the original system by a homotopy continuation method. Some of these solutions go out to infinity and are not solutions to the original problem.

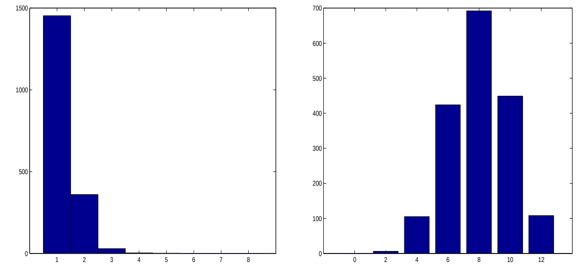

Figure 4: To the left a histogram over the number of solutions with the right orientationa and to the right a histogram over the number of real solutions for about 1800 simulated cases.

In order to study the number of solutions of our case we simulated around 1800 cases with three 3-quivers in three images. We then used PHC to solve the simulated cases. The solver found 12 solutions in each case, of which some were complex. We assume that we have the directions in each image. Up til now we have not used this information, but have only considered the lines. The information about the directions can be used to remove solutions with wrong orientation. It turns out that in most cases there was only one true solution to our problem! To the left in Figure 4 a histogram of the number of solutions with the right orientation is given and to the right of the number of real solutions.

5.3. A real image example

In Figure 5, three images of a scene with three quivers is shown. The quivers were manually extracted. We then used our parameterization and the solver to solve for the camera geometry. In this case there were ten real solutions but only one of them had the right orientation. To test this solution we extracted a number of points in the scene. Using our calculated cameras we then intersected the structure of the points and reprojected them in the images. The result is shown in Figure 5. One can see that the reprojection errors are quite small. The original points are shown as asterisks and the reprojected points as circles. The largest errors are for points furthest from the quivers as would be expected.

6. Conclusions

In this paper we have introduced the notion of a quiver.

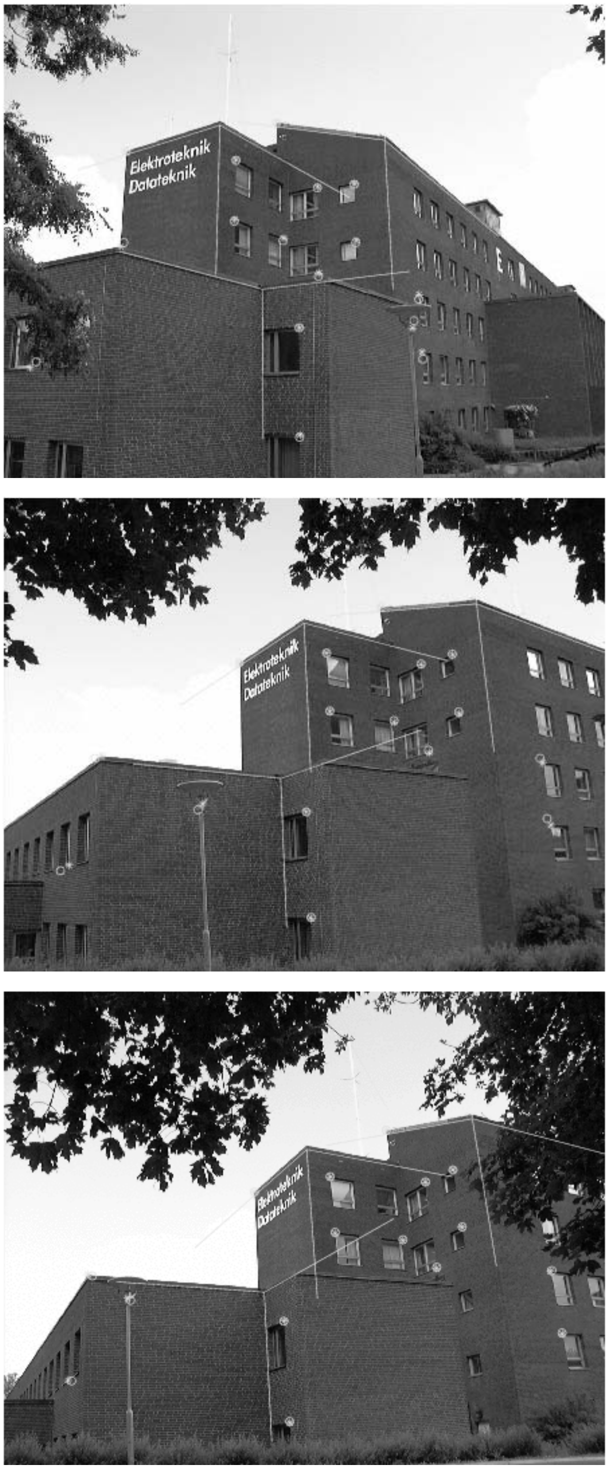

Figure 5: The three images used with three quivers. The reprojected points using the solution obtained from the quivers is also shown with original points as '*' and reprojected points as 'o'.

We have studied three types of quivers, with one, two and three directions respectively. For these types of quivers there are two minimal cases. For 1-quivers, three such features seen in three affine views gives a unique linear solution. For three 3-quivers in three uncalibrated projective views we get up to twelve solutions, but simulations show that in most cases we get a unique solution. Experiments on real data shows that the solutions obtained from the minimal data gives good estimates on the camera geometry.

We have also investigated 2-quivers seen in three views, where simulations show that such features give more stable estimation of the trifocal tensor as opposed to only using line or point information.

References

[1] K. Åström, A. Heyden, F. Kahl, and M. Oskarsson. Structure and motion from lines under affine projections. In *Proc. 7th Int. Conf. on Computer Vision, Kerkyra, Greece*, pages 285–292, 1999.

[2] D. Cox, J. Little, and D. O'Shea. *Using Algebraic Geometry*. Springer Verlag, 1998.

[3] O. Faugeras and T. Papadopoulo. Grassman-cayley algebra for modeling systems of cameras and the algebraic equations of the manifold of trifocal tensors. Technical Report 3225, Institut national de rechereche en informatique et en automatique, july 1997.

[4] O. Faugeras and T. Papadopoulo. A nonlinear method for estimating the projective geometry of three views. In *Proc. 6th Int. Conf. on Computer Vision, Mumbai, India*, 1998.

[5] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications-of-the-ACM*, 24(6):381–95, 1981.

[6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[7] A. Heyden. A common framework for multiple-view tensors. In *Proc. 5th European Conf. on Computer Vision, Freiburg, Germany*, pages 3–19, 1998.

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

[9] J. Verschelde. Phcpack: A general-purpose solver for polynomial systems by homotopy continuation. *ACM Transactions on Mathematical Software*, 25(2):251–276, 1999.

[10] Z. Zhang, R. Deriche, O. D. Faugeras, and Q. T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Technical report, INRIA, 1994.