

Constraint Enforcement In Structure And Motion Applied To Closing An Open Sequence

N. Guilbert, F. Kahl, M. Oskarsson and K. Åström

A. Heyden and M. Johansson

Centre for Mathematical Sciences
Lund University
Sweden

Division of Mathematics
Malmö University
Sweden

Abstract

In this paper a system for doing automatic structure and motion estimation given an image sequence taken by a hand-held camera is presented. The main contribution is the ability to stitch together partial reconstructions, specifically in the case of a closed loop motion. One key feature of the system is the possibility to incorporate simple geometrical constraints on the scene in the present case illustrated by two object points being the same, a group of points being the same or several points being coplanar. This is essential when dealing with long image sequences (hundreds of selected images from a sequence of thousands), since the accumulated error necessarily degrades the reconstruction significantly. Experimental validation is presented on both synthetic and real data.

keywords: structure and motion reconstruction, closed image sequences, constraint enforcement.

1. Introduction

Computer vision has matured during the last couple of years and it is now possible to build systems that, given only images taken with unknown cameras undergoing an unknown motion of an unknown scene, calculate the motion of the camera and the structure of the scene, cf. [5, 4].

Such systems depend on several components, like feature extraction and tracking, finding initial solutions to the structure and motion problem and refinement of the solution, see [3] and the references therein.

In building complete systems for solving structure and motion, new questions and research subjects arise naturally. One problem encountered when dealing with long image sequences where features might appear a second time (with a long subsequence not containing the feature in between) is that the re-projection errors might be extremely large. This can for example happen when moving the camera from one room along several other corridors or rooms and then returning to the first room again. A sequence where such

feature correspondences are applied will be referred to as *closed*. One example is illustrated in Figure 5. This phenomenon is due to the fact that small errors and degrees of freedom from partial reconstructions might accumulate to a very large error in both the scene structure and the camera motion. The structure obtained from the feature the first time it is encountered might not fit at all when it is re-projected to the images where the feature appears later due to these accumulated errors.

One way to deal with this problem is to make partial reconstructions from subsequences and then stitch these together. This could be done by minimizing the distances between corresponding points in 3D via homographies of the substructures as it is done in [2]. This approach has several drawbacks, e.g. there is no control over the reprojection errors in the images, a completely new optimization step is needed apart from the bundle adjustment optimization and it is not obvious how to implement this optimization in order to achieve convergence, since there might be very large errors in the starting position. Finally, the substructures might be subject to non-euclidian deformations.

Another way would be to impose soft constraints as described in [6], where a penalty term on the difference between expected identical parameters is included as a Lagrange multiplier in the error function. One could then envisage a gradual increase of the Lagrange multiplier in order not to detract the bundle adjustment algorithm, eventually imposing identity as the multiplier tends towards infinity. However, for long sequences, say containing over 300 cameras and 20000 3D points at the time of writing, each bundle step becomes very expensive, so excessive iterations should imperatively be avoided. Also, the notion of infinity mentioned previously is not clearly defined in this context, consequently there is no straightforward answer to how the different representatives should eventually be merged.

In the present work, a new method is presented for stitching together the structure in large closed image sequences that overcomes the disadvantages of the ones mentioned above. It has the following advantages: It is based on the

reprojection errors, It is an extension of traditional bundle adjustment and thus has a strong theoretical foundation, constraints are enforced exactly and not just approximately, constraints are enforced in a single step.

The proposed method exploits the uncertainties in the structure and motion problem encoded by the covariance matrix. In order to close a sequence, or more generally, to enforce a constraint, the structure and motion parameters are deformed using the covariance information, such that the reprojection error is affected as little as possible. This turns out to be a modified bundle step and it is easily incorporated in the standard bundle structure. Other examples of possible constraints include enforcing coplanarity among 3D points, constancy of intrinsic parameters and more.

A by-product of the modified bundle is that a measure of the likelihood of the features being the same is automatically obtained from the covariance structure. Thus it would be possible to do automatic reasoning under uncertainty, e.g. testing the hypothesis that two sets of parameters are actually duplicates, cf. [1]. This would make it possible to automatically stitch together closed image sequences. We will however not elaborate further on geometric reasoning in the present work.

The paper is organized as follows: In Section 2 we present the general ideas behind constraint enforcement. In Section 3 the theory is specialized to the more specific cases of imposing 3D point identities and/or coplanarity. Experimental validation on synthetic and real data is done in Section 4, and conclusions are given in Section 5.

2. Constrained Bundle Adjustment

2.1. Problem Description

After running a system of automatic structure and motion estimation for a period of time we have an estimate of the structure and motion as well as an estimate of the covariance structure of these estimates. We are now interested in using this information for inferring geometric constraints. Several types of constraints might be considered:

- Two object points are the same.
- Groups of points are the same.
- Several points are coplanar.
- Two configurations of coplanar points are parallel or orthogonal.

The purpose of this paper is to show how the standard bundle adjustment scheme could be extended to deal with these constraints. Given a solution to the structure and motion problem without these constraints, we show how to use this solution as a starting point for a modified bundle adjustment, where these constraints are taken into account. Furthermore, a measure of the probability of these constraints

being met is obtained, which could be used for hypothesis testing.

2.2. Constraint Enforcement

The method for incorporating constraints in the bundle adjustment is based on the measurement uncertainty. In general we try to find simple methods for assessing how the error function of the optimal estimates is increased as the constraints are enforced on the structure and/or motion.

After a constraint has been incorporated, the structure and motion will be updated with respect to this constraint. It is important that the representation of the uncertainty is also updated, so that the uncertainty is coherent with the geometric properties. For example if some points are coplanar, this geometric property should both be represented in the geometric data base but also in the representation of uncertainty.

We assume that we have a constraint H that we want to incorporate and define \mathcal{M}_H as the manifold of unknown parameters fulfilling H , and thus $\mathcal{M}_H \subset \mathcal{M}$, where \mathcal{M} is the manifold of all the parameters.

Let \mathbf{Y} denote the the residual vector formed by putting all deviations between the measured and the reprojected images coordinates in a column vector. The problem then becomes that of solving

$$\min_{\mathbf{m} \in \mathcal{M}_H} f(\mathbf{m}) = \min_{\mathbf{m} \in \mathcal{M}_H} \mathbf{Y}(\mathbf{m})^T \mathbf{Y}(\mathbf{m}) \quad (1)$$

given that we have the solution to

$$\min_{\mathbf{m} \in \mathcal{M}} f(\mathbf{m}) = \min_{\mathbf{m} \in \mathcal{M}} \mathbf{Y}(\mathbf{m})^T \mathbf{Y}(\mathbf{m}) . \quad (2)$$

The idea here is to use the linearized version of Y similar to the technique in the bundle adjustment. We study the approximate Taylor expansion of f

$$\begin{aligned} f(\Delta \mathbf{x}) &\approx \underbrace{\mathbf{Y}(0)^T \mathbf{Y}(0)}_{f_0} + \underbrace{2\mathbf{Y}(0)^T \frac{\partial \mathbf{Y}}{\partial \Delta \mathbf{x}}(0)}_{0 \text{ at optimum}} \Delta \mathbf{x} + \\ &2\Delta \mathbf{x}^T \underbrace{\frac{\partial \mathbf{Y}}{\partial \Delta \mathbf{x}}(0)^T \frac{\partial \mathbf{Y}}{\partial \Delta \mathbf{x}}(0)}_{C^{-1}} \Delta \mathbf{x} = \\ &f_0 + 2\Delta \mathbf{x}^T C^{-1} \Delta \mathbf{x}. \quad (3) \end{aligned}$$

In the sequel we will need the absolute coordinates of points instead of just local updates, therefore we will use p to denote a vector of the same size as $\Delta \mathbf{x}$ but at those positions of $\Delta \mathbf{x}$ which correspond to changes in a particular point coordinate, the corresponding element of p contains the absolute coordinate of the point. The other elements of p are set to zero. We will use $q = p + \Delta \mathbf{x}$ to represent local changes in parameter. Notice that $q - p = \Delta \mathbf{x}$ and that the parameters of q represent changes from the optimum

for non-point parameters and absolute coordinates for point parameters.

Thus the estimate $m \in \mathcal{M}$ is represented by the vector p , the uncertainty of the estimate $m \in \mathcal{M}$ is given by the covariance matrix C and the vector q represents estimates m_q close to p .

Definition 2.1. The problem of finding an estimate m_q , fulfilling the constraint H , by solving

$$\min_{m_q \in \mathcal{M}_H} (q - p)^T C^{-1} (q - p) \quad (4)$$

is called the **constraint enforcement problem**.

Notice that according to (3) the constraint enforcement problem is solved by minimizing the increase in the squared reconstruction error when applying a constraint.

3. Specific Cases

3.1. Merging 3D Points

In this section we address the problem of updating our estimate under the hypothesis that groups of points are in fact the same. For simplicity we will illustrate the method when two points are the same. We will use H as a constraint,

$$H = \{\text{Structure points } i \text{ and } j \text{ are in fact the same}\}. \quad (5)$$

For the case of the constraint being that two points are the same, the constraint enforcement problem can in fact be solved using numerical linear algebra. Let the vector $q_H \in \mathbb{R}^{k+3n-3}$ parametrize k motion and $(3n-3)$ structure parameters with points i and j being identical. In order to be able to compare m_q and m_p we need a mapping $\mathcal{M}_H \mapsto \mathcal{M}$. Given our local parameterization, this mapping is a linear function $q_H \mapsto Eq_H$ where E is the $(k+3n) \times (k+3n-3)$ matrix:

$$E = \begin{bmatrix} \mathbf{I}_{k \times k} & \mathbf{0} \\ \mathbf{0} & E_s \end{bmatrix}, \quad (6)$$

where

$$E_s = \begin{bmatrix} \mathbf{I}_{3 \times 3} & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & & \vdots & & \vdots \\ \mathbf{0} & \dots & \mathbf{I}_{3 \times 3} & \dots & \mathbf{0} \\ \vdots & & \vdots & & \vdots \\ \mathbf{0} & \dots & \mathbf{I}_{3 \times 3} & \dots & \mathbf{0} \\ \vdots & & \vdots & & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{I}_{3 \times 3} \end{bmatrix}. \quad (7)$$

This is almost an identity matrix but the columns for the coordinates for object point j have been removed and instead both point i and j are parameterized by the same parameters. The constraint enforcement problem now becomes

$$\min_{q_H \in \mathbb{R}^{k+3n-3}} (Eq_H - p)^T C^{-1} (Eq_H - p). \quad (8)$$

This is a quadratic problem whose solution is

$$q_{min} = (E^T C^{-1} E)^{-1} E^T C^{-1} p. \quad (9)$$

If the constraints have been applied, the parameters p and the uncertainty C need to be updated. This can be done according to

$$p_{new} = Eq_{min} \text{ and } C_{new} = (E^T C^{-1} E)^{-1}. \quad (10)$$

Enforcing the constraint that a group of points from the beginning of the sequence should match a group of points from the end of the sequence is identical to that of the previous except that the matrix E for m identical points has size $(k+3n) \times (k+3(n-m))$ and that the coordinates of points (i_1, \dots, i_m) and (j_1, \dots, j_m) are parametrized by the same parameters.

Remark. Similar to standard bundle adjustment, one can take advantage of the sparseness of C^{-1} in order to invert $E^T C^{-1} E$ in (9) efficiently.

3.2. Coplanar Points

In order to enforce the constraint of coplanarity of a group of points, two methods are developed. In the first method we assume that the parameters of the plane are fixed, e.g. from the fitting of the plane to the points in the hypothesis generation. This is obviously not optimal. In the second method we assume that both the plane parameters and the adjusted points are found.

For simplicity the derivation here assumes that p and q only contains the points that are coplanar, i.e.

$$p = [x_1 \ y_1 \ z_1 \ \dots \ x_n \ y_n \ z_n]^T.$$

Generalization to the case when these vectors contain parameters of other unconstrained points and camera parameters is straightforward. In the first method we assume also that the points are to lie on a known plane $\pi = [a \ b \ c \ d]^T$. The constraint enforcement amounts to solving

$$\min_{\substack{q, \\ Pq=D}} (q - p)^T C^{-1} (q - p), \quad (11)$$

with

$$P = \begin{bmatrix} a & b & c & \dots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \dots & c \end{bmatrix}$$

and $D^T = [d \ \dots \ d]$. Since the constraints are linear it is straightforward to solve for the constraints and parameterize all valid $q \in \mathcal{M}_H$ as $q = Ax + q_0$. The minimization problem is quadratic

$$\min_x (q_0 - p + Ax)^T C^{-1} (q_0 - p + Ax), \quad (12)$$

with solution

$$x = A(A^T C^{-1} A)^{-1} A^T C^{-1} (q_0 - p) + q_0. \quad (13)$$

Assuming that also the plane is unknown gives a slight variation on the problem, i.e.

$$\min_{\substack{q, \pi, \\ P(\pi)q = D(\pi), \\ |\pi| = 1}} (q - p)^T C^{-1} (q - p). \quad (14)$$

This problem is more difficult. One way to solve it is to start with an approximate solution and then refine iteratively.

The problem can be formulated as a linear least squares problem with non-linear constraints:

$$\min_{\substack{y, \\ g(y) = 0}} f(y)^T f(y). \quad (15)$$

Here $y = \begin{bmatrix} \pi \\ q \end{bmatrix}$, $f(y) = L(q - p)$, L is the Cholesky factorization of C and $G(y) = \begin{bmatrix} \pi^T \pi - 1 \\ P(\pi)q - D(\pi) \end{bmatrix}$. In each iteration y_k . The estimate is updated according to $z_k = y_k + B\Delta x$, where the columns of the matrix B is a basis for the space orthogonal to ∇g . This means that to the first order the update does not violate the constraints. The step Δx is chosen according to the Gauss-Newton update, where the over-constrained linear problem $\nabla f(y_k) B \Delta x = f(y_k)$ is solved in a least squares sense. The new point z does not necessarily obey the constraints. Therefore a inner loop of optimization is performed with z as a starting point. The result is the new estimate y_{k+1} .

4. Experimental Validation

4.1. Synthetic Data

In this section we present some results from experiments on synthetic data. In the bundle adjustment, camera calibration is assumed to be constant.

Imagine a room with four walls and points evenly distributed on the walls. In this fictive room, 25 cameras were placed on an elliptical trajectory in a plane and 100 points were put on walls around, 25 on each wall. The points were projected onto the cameras and independent Gaussian zero-mean noise was added to the image coordinates. The standard deviation of the noise is given as a percentage of the chosen image size. For example, for an image of size 400×400 pixels, 0.5% noise corresponds to 2 pixels. The image size is set such that around 20 points are visible in each image.

Merging 3D points. A typical example is shown in Figure 1 at a noise level of 0.5%. In (a) the result after a full bundle adjustment without assuming that the sequence is

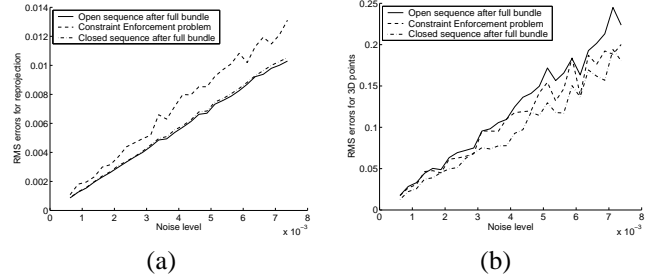


Figure 2: In (a) the RMS reprojection errors for different noise levels are shown. In (b) RMS errors between estimated and ground truth 3D point positions for different noise levels are graphed.

closed is plotted. One can see that there is quite a large discrepancy between the positions of the first and last camera, which should be in the same place.

The result after using the method of Section 3.1 to close the sequence, is drawn in Figure 1(b). One can clearly see that the first and last cameras have moved significantly. The points have also moved closer to their true positions against the walls. The RMS (Root Mean Squares) errors for the 3D points compared to ground truth decreased from 0.84 units to 0.42 units. In (c), the result after a full bundle adjustment assuming a closed sequence is shown. The RMS for the final closed sequence is 0.08 units.

In Figure 2 some results for different noise levels are graphed. In (a) the RMS errors for reprojection in the images are shown. As expected, the image errors increase, but only slightly, when constraints (i.e. merging of points) are enforced. In (b) the RMS errors between the true 3D points and the estimated ones are compared for the three different settings. One can see that the methods in Section 3.1 do indeed decrease the errors in the 3D points compared to ground truth.

Coplanar points. The method of enforcing coplanarity in Section 3.2 has also been tested. Consider again the synthetic room with four walls (cf. Figure 1) with 0.5% noise. Our hypothesis (or constraint) is that the 25 points on the right wall are on a plane¹.

The constraint enforcement problem in (14) is non-linear and an initial solution is required. An initial estimate of the plane is computed by linear plane fitting of the 3D points and an initial solution of the remaining parameters is obtained from the formula in (13). Using Gauss-Newton updates (as explained in Section 3.2), convergence is reached in just a few iterations. The result is illustrated in Fi-

¹The process of generating possible hypotheses can be done by picking three random scene points and check how many of the other points are close to that plane. Experiments on this are omitted due to lack of space.

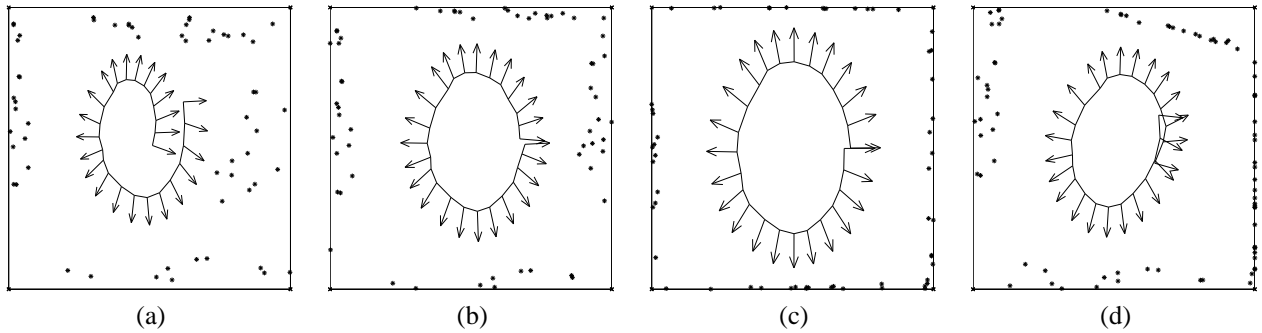


Figure 1: In (a) the result after bundle adjustment without any geometric constraints is shown. (b) is the result after merging 7 points in the first and last image. (c) shows the result for a full bundle adjustment with the merged points. In (d) coplanarity of points on the right wall is enforced.

Figure 1(d) where the starting solution is from the open sequence in Figure 1(a). The 3D points on the right wall are exactly on a plane.

At first, the result seems to be not so good. The camera trajectory is twisted and far from the ground truth. However, there is an important observation to make here. Enforcing coplanarity moves the 3D points in the normal direction of the plane, and hence the cameras are adjusted to be on the correct distance to the plane. There is still a lot of uncertainty in the top-to-bottom direction, which can be seen in the figure. The RMS errors for all the 3D points decreased from 0.84 units to 0.58 units.

4.2. Experiments on Real Data

We give three experiments performed on real data for closing long sequences. The degree of complexity increases as we proceed - the first experiment involves 36 images and the last one deals with several hundreds of frames.

For all three sequences, Euclidean structure and motion was recovered using the standard methods described in [3] assuming constant intrinsic calibration. Since ground truth is not known, the natural error measure is the RMS of the projected features in the images.

Dinosaur. The first sequence consists of 36 images of a dinosaur which was put on a turn-table, cf. Figure 3. The sequence was also tested in [2]. It is characterized by a very regular (circular) motion and precisely tracked features. 375 points were tracked and bundled. Five points were merged in order to close the sequence. Note that the motion being circular is not taken into account - the results are only based on measured image features.

Casino. The third sequence was taken in the Royal Room of the Casino of Madrid (Figure 4). It originally consisted of four slightly overlapping video sequences describing the

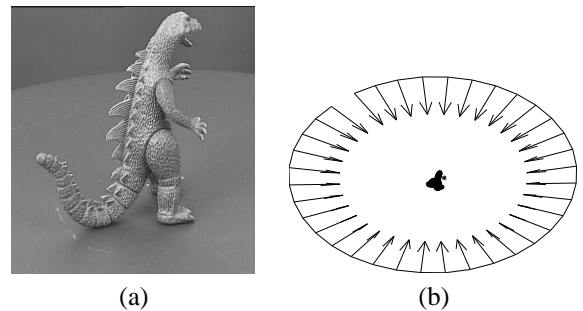


Figure 3: (a) First image in the Dinosaur sequence. (b) Reconstructed Dinosaur sequence (open).

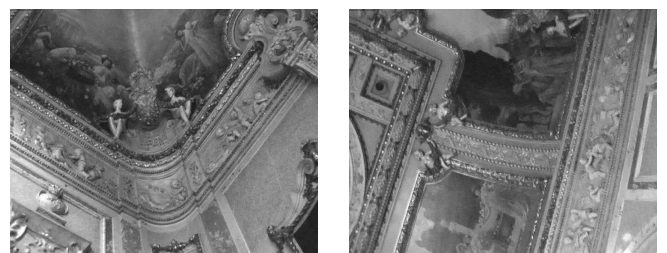


Figure 4: Two images from the Casino sequence.

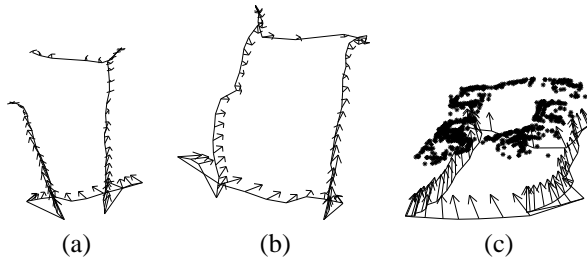


Figure 5: Casino sequence. (a) Trajectory for open sequence, (b) trajectory after closing and (c) final reconstruction of closed sequence with 3D points.

four edges of a near-rectangular camera path. The whole sequence, on which the tracking was performed, lasts several minutes and was in the process reduced to 281 views and 16000 3D points. In order to bundle efficiently, only the 835 longest point tracks were kept in the global bundle adjustment. Five points were merged to close the sequence.

Results. Table 1 presents the evolution of the error. Each sequence has two row entries in the table. The first (overall) indicates the overall RMS (reprojection error) for all image points in the sequence, while the sequence is still open, after imposing the closure constraint (4) and after convergence of the bundle adjustment of the closed sequence, respectively. The second row entry (closure) gives the RMS error when reprojecting the 3D points which are used for merging (and hence visible in both the last and first image) onto the first image.

(pixels)		Open Seq.	Constraint Enforc.	Closed Seq.
Dinosaur	overall	0.3419	0.3543	0.3423
	closure	1.6348	0.2269	0.2245
Casino	overall	0.3340	15.54	0.3667
	closure	777.6	10.55	0.6098

Table 1: Reprojection errors (RMS) in pixels of the reprojected points for the open sequence, after applying the constraint enforcement and for the closed sequence. ‘Overall’ indicates the RMS error for all reprojected points in all images, and ‘closure’ the RMS error of the reprojected 3D points of the closure constraint in the first image.

The precision of the Dinosaur sequence is striking. Although the camera has performed a complete revolution around the scene, the final reprojection error of the 3D points visible in both the first and the last sequence is less than 2 pixels (before closure). Again, the fact that the motion is circular is not enforced, even though it looks perfectly circular in Figure 3(b). So, any method for closing

would probably work on this sequence.

Applying the constraint enforcement reduces the error in the first image to pixel level and the final bundle eventually leads to sub-pixel level errors (cf. Table 1).

For the Casino sequence, the magnitude of the closure error is such that usual bundle adjustment turned out to be insufficient and as can be seen from Figure 5(a), the trajectory mismatch is truly significant. The constraint enforcement was performed using five overlapping points. As can be seen from Figure 5(b) and (c) as well as Table 1, it was successful.

5. Conclusions

In this paper a system for doing structure and motion estimation from closed image sequences taken by hand-held cameras was presented. It is based on an extension of the standard bundle adjustment method that can deal with constraints on both structure and motion, e.g. merging 3D-points and enforcing coplanarity constraints. The extended bundle uses the covariance information obtained before enforcing the constraints in order to find the best modification of the structure and motion obeying the constraints. The experiments show that the proposed method works very well on both simulated and large-scale real experiments. A natural extension to the present formulation would be to perform hypothesis testing, which would allow a fully automatic self-closing system, provided hypotheses can be generated automatically.

References

- [1] C. Baillard and A. Zisserman. Automatic reconstruction of piecewise planar models from multiple views. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 559–565, 1999.
- [2] A. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. In *European Conf. Computer Vision*, volume I, pages 311–326, Freiburg, Germany, 1998.
- [3] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [4] 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*, pages 438–443, 1997.
- [5] P. McLauchlan and D. Murray. A unifying framework for structure and motion recovery from image sequences. In *Proc. Int. Conf. on Computer Vision*, pages 314–320. IEEE Computer Society Press, Los Alamitos, California, 1995.
- [6] B. Triggs, P. McLauchlan, H. R.I., and A. Fitzgibbon. Bundle adjustment - a modern synthesis. In *Vision Algorithms’99*, pages 298–372, in conjunction with ICCV’99, Kerkyra, Greece, 1999.