ACCV2002: The 5th Asian Conference on Computer Vision, 23–25 January 2002, Melbourne, Australia.

A Scheme for Combining Auto-calibration and Scene Constraints*

D. Q. Huynh† A. Heyden‡ S. Khan†
†School of Information Technology, Murdoch University, Perth WA 6150 AUSTRALIA
‡Centre for Mathematical Sciences, Lund University, Box 118, SE-221 00 Lund, SWEDEN
Email: d.huynh@murdoch.edu.au heyden@maths.lth.se s.khan@murdoch.edu.au

Abstract

A scheme for combining auto-calibration and scene constraints in the "structure-from-motion" problem is proposed. This scheme focuses on the recovery of the absolute quadric using auto-calibration while imposing orthogonal scene plane constraints. First, an initial estimate of the absolute quadric is obtained using a linear method. A non-linear constrained optimization step is then applied to refine this quadric and the camera intrinsic parameters to upgrade the estimated projective reconstruction to Euclidean. Finally, a bundle adjustment algorithm optimizes the Euclidean reconstruction to give a statistically optimal result. Constraints from orthogonal scene planes are applied to the initial estimation and refinement steps of the absolute quadric. The performance of the scheme is demonstrated on both simulated and real video data.

1. Introduction

The "structure-from-motion" problem concerns the recovery of both the scene structure and motion of the camera(s) from a set of images. Methods for solving this problem vary largely depending on the camera model chosen, these being, notably, the projective [4, 5] and the affine [20, 16] camera models. The type of image features being used, eg. image lines, points, or conics, on the other hand, determines the number of images required for shape reconstruction [6, 9, 1]. If no special information about the camera or the scene is available then only a *projective* reconstruction of the scene can be obtained, cf. [4, 13, 5]. Since projective structures are not suitable for visualization, it is often desirable to obtain a Euclidean reconstruction up to an unknown similarity transformation.

Traditionally, the Euclidean structure of a scene has been obtained via two different types of methods. The first type is often referred to as *stratification*, since one starts with a projective reconstruction and then finds an affine 'stratum' and finally a Euclidean 'stratum' to give the desired reconstruction. In the upgrade from projective to Euclidean reconstruction, some scene constraints, eg. some distance or angular measurements, cf. [2] may be incorporated. The second type, often referred to as *auto-calibration*, takes into account some a priori information about the intrinsic parameters, eg. known skew and/or aspect ratio (see [12, 21, 8, 15]) or constant intrinsic parameters (see [4]). The main focus of this latter type of method is on finding the intrinsic parameters, ie. auto-calibrating the cameras.

The purpose of this paper is to combine these two methods, making it possible to incorporate both constraints from the scene and constraints on the intrinsic parameters. In particular, we will focus on the use of the natural camera model (ie. a model that has zero skew and unit aspect ratio) and orthogonal scene planes as constraints for the estimation of the absolute quadric and to optimize the Euclidean reconstruction. The proposed method will have a number of advantages compared to using only one type of constraint: (i) using all available constraints will increase the accuracy of reconstruction, (ii) degenerate or near-degenerate cases can be avoided by using different types of constraints, (iii) robustness is increased since the linear initialization step has more constraints.

The use of the natural camera model is justified by the high quality digital and video cameras available today. Even if the skew is non-zero and the aspect ratio of a camera is not unity, these two entities are known to be invariant under the change of focus setting and so they can be pre-calibrated and treated as constant. The use of orthogonal scene planes is justified by the presence of many such planes in man-made objects, eg. indoor scenes, buildings.

2. Background

Auto-calibration which involves the recovery of the absolute conic or the absolute quadric has been attempted by various researchers. The earliest work on recovery of the absolute conic (ie. auto-calibration of the camera(s)) is the algorithm by Faugeras et al [3], where one camera is involved and its intrinsic parameters are assumed fixed. Heyden and Åström [7] later retrieve Euclidean structure via the computation of the absolute quadric, assuming also fixed camera intrinsic parameters. Triggs [21] proposes the use of quasi-linear constraints for the absolute quadric, with the same assumption imposed. Pollefeys et al [14] incor-

*This research was, in part, supported by the Murdoch Special Research Grant MU.AMH.D.413 and the SSF/SRC-JIG project 95-64-222.

porate the so-called modulus constraint to the stratification approach to upgrade projective structure to affine and finally the recovery of the absolute conic to upgrade the structure to Euclidean. In their other work [15], they first assume the principal points in all images vanish to linearize the equations for computing the absolute quadric. A nonlinear optimization process is then followed which allows the intrinsic parameters of the images to vary independently.

Although Pollefeys et al [15] have indicated that scene constraints can be incorporated into auto-calibration, no further investigations have been reported in the literature on the use of such constraints. Liebowitz and Zisserman's report [10] is perhaps the only one that detects the image projections of parallel and orthogonal scene lines and uses them to estimate the vanishing points and as constraints for estimating the absolute conic. However, their method requires the computation of the fundamental matrix, and that limits their method to working with two images.

In this paper, we show that scene and auto-calibration constraints can be easily combined for the structure-from-motion problem. In particular, we focus on the use of orthogonal scene planes that can be easily found in buildings.

3. The proposed scheme

A number of steps are involved in our scheme. First, using the list of tracked image feature points, a projective reconstruction of the imaged scene is obtained via an iterative factorization and a projective bundle adjustment. An initial estimate of the absolute quadric, which embodies information for Euclidean upgrade, is then computed using linear constraints from a special case of the natural camera and from orthogonal scene planes. This is followed by a nonlinear constrained optimization to refine the estimated absolute quadric and the camera intrinsic parameters. In the optimization process, constraints from the natural camera and from orthogonal scene planes are enforced. By using the refined estimate of the absolute quadric, the projective structure obtained from the first step is upgraded to Euclidean. A final Euclidean bundle adjustment is employed to further optimize the structure and intrinsic camera parameters so as to obtain a statistically optimal result.

3.1. Projective Reconstruction

Given a scene point $\mathbf{X}^j = [X^j\ Y^j\ Z^j\ 1]^\top$, its projection $\mathbf{x}^j = [x^j\ y^j\ 1]^\top$ onto an image plane is governed by:

$$\lambda^j \mathbf{x}^j = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R & -R\mathbf{t} \end{bmatrix} \mathbf{X}^j \qquad (1)$$

where the superscript j denotes the j^{th} scene (or image) point from a list of scene (or image) points and λ^j an unknown scalar. The first matrix is the camera matrix K which embodies the unknown focal length f and principal point (u_0, v_0) of the camera. The second matrix is the motion matrix which contains the unknown rotation R and translation $\mathbf{t}$ of the camera to the scene point. The special form of K arises from the use of the natural camera model.

When none of these parameters are known *a priori*, (1) is often put in the compact form $\lambda^j \mathbf{x}^j = P\mathbf{X}^j$, where P is a 3×4 projection matrix. With the availability of m images and n scene points, the $3m \times 4$ joint projection matrix $\mathbf{P}$, the $3m \times n$ joint image measurement matrix $\mathbf{x}$, and the $4\times n$ joint shape matrix $\mathbf{X}$ are related by

$$\begin{bmatrix} \lambda_1^1\mathbf{x}_1^1 & \cdots & \lambda_1^n\mathbf{x}_1^n \\ \vdots & \ddots & \vdots \\ \lambda_m^1\mathbf{x}_m^1 & \cdots & \lambda_m^n\mathbf{x}_m^n \end{bmatrix} = \begin{bmatrix} P_1 \\ \vdots \\ P_m \end{bmatrix} \begin{bmatrix} \mathbf{X}^1 & \cdots & \mathbf{X}^n \end{bmatrix}$$
$$\Leftrightarrow \mathbf{x} = \mathbf{P}\mathbf{X} \qquad (2)$$

where the subscript i denotes the i^{th} camera.

Setting all the λ_i^j to 1 for the affine camera model, Tomasi and Kanade[20] pioneered the use of the factorization method to retrieve the joint projection matrix $\mathbf{P}$ and joint shape matrix $\mathbf{X}$ from the joint image measurement matrix $\mathbf{x}$. For the projective camera model, Sparr [18] proposes using the subspace method to iteratively refine all the λ_i^j's. In the same year, Sturm and Triggs [19] propose a different scheme that requires estimation of the fundamental matrices and the epipoles. Our Matlab code is based on the method of [18]. Initially, all the λ_i^j's are assumed to be unity. At each iteration, the λ_i^j's are refined with the subspace (4-dimensional subspace of $\Re^n$) constraint on $\mathbf{x}$ being enforced while minimizing the image point reprojection errors. The matrix $\mathbf{x}$ is then updated with the refined values of λ_i^j's and re-factorized to give new $\mathbf{P}$ and $\mathbf{X}$ matrices for the next iteration. The method has been shown to give very good estimates of the λ_i^j's and very rapid convergence.

3.2. Linear estimation of the absolute quadric

The structure contained in the shape matrix $\mathbf{X}$ is projective only, since for any joint projection matrix $\tilde{\mathbf{P}}$ and joint shape matrix $\tilde{\mathbf{X}}$ that satisfy (2), $\tilde{\mathbf{P}}A$ and $A^{-1}\tilde{\mathbf{X}}$ also form a solution, where A is any non-singular 4×4 matrix. This means that the projection matrices can be so arranged that the first projection matrix $P_1 = [\ I \quad \mathbf{0}\ ]$ and the other projection matrices $P_i = [\ Q_i \quad \mathbf{q}_i\ ]$, for some $Q_i \in \Re^{3\times3}$ and $\mathbf{q}_i \in \Re^3$.

To upgrade the projective structure to Euclidean, a matrix A must be recovered for an appropriate projective change of coordinates in the estimated scene points. That is, we estimate matrix A such that

$$P_i A \sim K_i\,[\ R_i \quad -R_i\mathbf{t}_i\ ], \quad \text{for } i = 1, ..., m. \qquad (3)$$

Since $P_1 = [\ I \quad \mathbf{0}\ ]$, matrix A takes the form

$$\begin{bmatrix} K_1 & \mathbf{0} \\ \mathbf{a}^\top & s \end{bmatrix} \tag{4}$$

where $\mathbf{a} = (a_1, a_2, a_3)^\top$ and the unknown, non-zero scalar s is often set to unity. It follows that

$$P_i \tilde{A} \tilde{A}^\top P_i^\top \sim K_i K_i^\top.$$

Here, $\tilde{A}$ represents the matrix that contains the first three columns of A. This leads to

$$\begin{aligned} P_i \begin{bmatrix} K_1 K_1^\top & K_1 \mathbf{a} \\ \mathbf{a}^\top K_1^\top & \|\mathbf{a}\|^2 \end{bmatrix} P_i^\top &\sim K_i K_i^\top \\ \Leftrightarrow P_i \Omega P_i^\top \sim K_i K_i^\top, \quad &\text{for } i = 1, ..., m \end{aligned} \tag{5}$$

The 4×4 matrix, denoted by Ω in (5), is the singular dual quadric that contains the coordinates of the plane at infinity (vector $\mathbf{a}$) for affine reconstruction and the DIAC (dual image of the absolute conic) $K_1 K_1^\top$ for Euclidean reconstruction.

If image feature points from two orthogonal planes in the scene are identified then an additional linear constraint on Ω can be imposed, ie.

$$\mathbf{n}^\top \Omega\, \mathbf{m} = 0 \tag{6}$$

where $\mathbf{n} = (n_1, n_2, n_3, n_4)^\top$ and $\mathbf{m} = (m_1, m_2, m_3, m_4)^\top$ are the coordinates of the two orthogonal planes computed from the projective structure recovered in Section 3.1.

Finding the unknowns in (5) is a difficult nonlinear problem. An alternative, as suggested in [15], is to assume further that the camera's principal point is approximately at the centre of the image buffer. So, after setting the origin of the image coordinate system at the centre of the image buffer, (u_{0i}, v_{0i}) can be set to $(0, 0)$ for all i. This leaves f_i as the only unknown intrinsic parameter for each image and each $K_i K_i^\top$ as a diagonal matrix whose first 2 diagonal elements are identical. Except for the first image which provides identity equations to (5), each of the other images provides 4 equations from the special form of $K_i K_i^\top$. The special form of $K_1 K_1^\top$ reduces the number of unknowns in Ω to 5: $\Omega_{11} (= \Omega_{22})$, Ω_{14}, Ω_{24}, Ω_{34}, Ω_{44}. Other elements of Ω are 0, except for Ω_{33} which is 1. The inclusion of a scene constraint (Equation (6)) thus allows us to solve for all the unknowns using 2 images ($m = 2$). In our experiments reported here, more than 2 images from a video sequence were used to recover Ω. More images were chosen since a degenerate configuration or critical motion that might affect any particular image pair is unlikely to affect the entire selected set of images.

3.3. Refining the absolute quadric

The initial estimate of Ω obtained from the linear method above must be refined, taking into account the situations where f_i, u_{0i}, v_{0i}, for all i, are variable in the video sequence. Also, the scene constraints given in (6) have been simplified in order to provide linear constraints for computing Ω. The exact constraints should include a normalization factor that involves Ω, ie.

$$\frac{\mathbf{n}^\top \Omega\, \mathbf{m}}{\sqrt{\mathbf{n}^\top \Omega\, \mathbf{n}}\sqrt{\mathbf{m}^\top \Omega\, \mathbf{m}}} = 0 \tag{7}$$

for two orthogonal planes $\mathbf{n}$ and $\mathbf{m}$ in the scene.

The form of K_i for the natural camera can be imposed in the nonlinear optimization. Writing (5) and (7) in matrix-vector form and dropping the subscript 1 for the first image, we have the following constrained minimization problem:

$$\min_{\substack{a_1, a_2, a_3, \\ f, u_0, v_0, \ldots, \\ f_m, u_{0m}, v_{0m}}} J = \sum_{i=2}^{m} \left\| \frac{M_i \mathbf{q}}{\sqrt{(M_i \mathbf{q})^\top (M_i \mathbf{q})}} - \frac{\mathbf{k}_i}{\sqrt{\mathbf{k}_i^\top \mathbf{k}_i}} \right\|^2 \tag{8}$$

subject to

$$\sum_{j,l} \frac{(N_{jl} \mathbf{q})^\top (N_{jl} \mathbf{q})}{(N_{jj} \mathbf{q})(N_{ll} \mathbf{q})} = 0. \tag{9}$$

Here $M_i \in \Re^{6 \times 10}$, $\mathbf{q} \equiv \left[f^2 + u_0^2, u_0 v_0, u_0, a_1, f_1^2 + v_0^2, v_0, a_2, 1, a_3, a_1^2 + a_2^2 + a_3^2 \right]^\top \in \Re^{10}$ is the vector form containing the elements of Ω, and $\mathbf{k}_i \equiv [\, f_i^2 + u_{0i}^2, u_{0i} v_{0i}, u_{0i}, f_i^2 + v_{0i}^2, v_{0i}, 1 \,]^\top$ contains the elements of $K_i K_i^\top$. Equation (9) is the sum of several orthogonal scene constraints from (7) and $N_{jl} \in \Re^{1 \times 10}$. We note that $\mathbf{q}$ contains only the coordinates of the plane at infinity and the intrinsic parameters of the first image, while $\mathbf{k}_i$ (where $i \neq 1$) contains only the unknown intrinsic parameters of the i^{th} image. The summation in (8) starts at the second image since the camera matrix for the first image satisfies (5) trivially. The total number of unknowns in the above nonlinear constrained optimization is $3m + 3$. This parameterization automatically imposes the condition $\det(\Omega) = 0$.

To get a rapid evaluation of the estimate of the absolute quadric using the above method, we chose the `fmincon` routine of Matlab and submitted to it the analytical gradients of the objective and the constraint functions. `fmincon` minimizes the nonlinear system of equations using the Quasi-Newton method. On the assumption that the initial estimates of the parameters are close to the true solution, the iteration should converge to a minimum close to the starting point. Accordingly, it is appropriate to bound all of the unknowns. For instance, the camera focal length must be within 700 to 3000 pixels; the principal point must be within the range ± 50 pixels of the centre of the image buffer; a_k, for $k = 1, ..., 4$ must be bounded within the interval $a_k^0 \pm 0.5$, where a_k^0 is the initial estimate of a_k. Both the ranges ± 50 and ± 0.5 above are empirically chosen values.

3.4. Initial linear Euclidean upgrade

The optimization process above not only refines the estimate of Ω, which allows the projective structure of the scene to be upgraded to Euclidean, it also refines the camera intrinsic parameters of all images. An initial Euclidean structure can be retrieved by estimating A from the refined absolute quadric Ω.

Since $\tilde{A}\tilde{A}^\top \sim \Omega$, the best way to compute A is to let $\tilde{A} \equiv U_3 \hat{S}_3^{1/2}$ where $\Omega \equiv USV^\top$ is the SVD of Ω, and U_3 and S_3 the matrices containing the first 3 columns of U and S. By using the intrinsic parameters refined from (8) and the recovered $\tilde{A}$ matrix, K_i, R_i, $\mathbf{t}_i$, for $i = 1, ..., m$ can all be retrieved. The projective structure $\mathbf{X}$ estimated from Section 3.1 is then upgraded to $\mathbf{X}_e \equiv A^{-1}\mathbf{X}$.

3.5. Bundle Adjustment

The initial Euclidean reconstruction $\mathbf{X}_e$ obtained above can be further improved by minimizing the reprojection errors while bundling all the reconstructed 3D points, camera intrinsic and extrinsic parameters in an iterative refinement process. The number of reprojection errors to be minimized is $2nm$[1] for n image points tracked over m images, while the number of parameters to be refined is $3n + 9m$. We have implemented the bundle adjustment in Matlab using the Levenberg-Marquardt method. The rate of convergence is very fast and requires only 25 iterations. We note that the degree of orthogonality of orthogonal scene planes may reduce slightly (by less than 1°) after bundle adjustment.

4. Experiments

To demonstrate that use of all available constraints increases the accuracy of the reconstruction, we conducted experiments using synthetic data. We used the result obtained from our previous research work on camera calibration to ensure that the synthesized data are realistic. The experiments synthesized the motion of a camera viewing a calibration target that had two orthogonal faces, each of which had 9 calibration points. The lateral motion was always set to positive to simulate a left to right movement of the camera. Small up/down and forward/backward motions were synthesized so that the motion would not be critical. Rotation angles of about $-5°$ to $5°$ were randomly synthesized for the 3 principal axes. The camera's focal length varied up to 10% from the focal length of 1720 pixels computed from our previous camera calibration work. The calibration-target-to-camera distance was approximately 50 cm. In each experiment, 5 camera motions were synthesized as described above, together with 5 different focal lengths and principal points.

[1]There are nm reprojection errors to be minimized for each of the x- and y-directions.

In each experiment, we compared the relative errors of the focal lengths, principal points, reconstruction errors, and orthogonality errors (deviation from the 90° angle) computed from two approaches:

(a) *auto-calibration without scene constraints*, which does not use any scene constraints in the initial estimation and refinement steps of the absolute quadric, and
(b) *auto-calibration with scene constraints*, which incorporates scene constraints in both the initial estimation and refinement steps of the absolute quadric.

In approach (a), the Matlab routine `fminunc` is employed for the refinement of the absolute quadric. Approach (b) is the scheme we propose.

The means and standard deviations of the relative errors mentioned above for 20 experiments on synthesized data are listed in Table 1. The entity $\epsilon_f \equiv |\hat{f} - \bar{f}|/\bar{f}$ is the relative error of the estimated focal length; $\epsilon_{\mathbf{X}} \equiv \|\hat{\mathbf{X}} - \bar{\mathbf{X}}\|$ is Euclidean reconstruction error (in cm); $\epsilon_{(u_0,v_0)} \equiv \|(\hat{u}_0, \hat{v}_0) - (\bar{u}_0, \bar{v}_0)\|$ is the principal point error; $\epsilon_\theta \equiv |\theta - 90|/90$ is relative error of the estimated angle. In the above notations, '$\hat{}$' and '$\bar{}$' denote, respectively, the estimated and true values of the entity, $|.|$ denotes the absolute value, and $\|.\|$ denotes the 2-norm of the vector concerned. The results clearly show that a smaller reconstruction and orthogonality error (2.63% versus 6.87%) can be attained by incorporating scene constraints (if available). Except for the slightly larger mean principal point error, a smaller focal length error (9.17% versus 9.19%) also resulted from the use of scene constraints.

Table 1. The means and standard deviations of a few error measures in 20 simulations.

	Without scene constraints		With scene constraints	
	mean	std devn	mean	std devn
ϵ_f	0.0919	0.0615	0.0917	0.0639
$\epsilon_{\mathbf{X}}$	1.1312	0.9273	0.5868	0.4621
$\epsilon_{(u_0,v_0)}$	24.6308	16.4409	27.1651	16.2034
ϵ_θ	0.0687	0.0384	0.0263	0.0192

Two of the real experiments conducted are reported here. In the first experiment, four images (of dimensions 768×576 pixels) of a calibration target (Figure 1) were captured from four different view points. The image feature points were identified and the correspondences were established manually. The centre of the image buffer is at $(384, 288)$ and the principal point of the camera is expected to vary around this point. All of the 72 image points on both faces of the calibration target were used for computing the projective coordinates of two orthogonal scene planes, which were combined to form a scene constraint.

Figure 2 shows the Euclidean reconstruction of the calibration target and the estimated positions, $\{C_i \mid 1 \le i \le 4\}$,

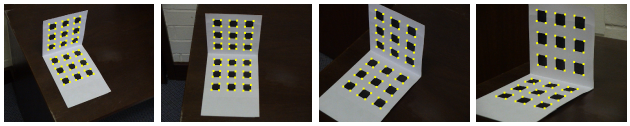

Figure 1. The 4 images in Experiment 1.

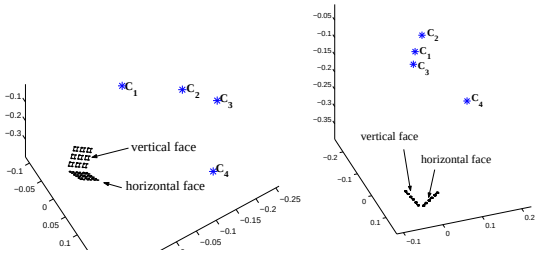

Figure 2. Two perspective views of the Euclidean reconstruction using our proposed scheme.

of the camera. The improvement from the use of scene constraints to this experiment is very small because both approaches (a) and (b) yielded very good initial estimates of the absolute quadric (see Section 3.2). The estimated intrinsic parameters of the camera for the 4 images are given in Table 2. We note that the principal point $(336.25, 70.69)$ of

Table 2. The estimated focal lengths and principal points of the camera in Experiment 1.

image	Without scene constraint		With scene constraint	
	f (pixels)	(u_0, v_0) (pixels)	f (pixels)	(u_0, v_0) (pixels)
1	1271.89	(170.81, 298.48)	1390.39	(326.26, 313.97)
2	1480.16	(230.63, 300.38)	1680.58	(317.27, 245.68)
3	1784.32	(263.76, 232.22)	1866.89	(331.25, 268.77)
4	2294.82	(336.25, 70.69)	1963.78	(366.20, 214.05)

the 4^{th} image from approach (a) is at a significant distance away from the centre of the image buffer.

The angle between the two faces was estimated to be $89.29°$ via approach (a) and $90.84°$ via approach (b). The imposition of a scene constraint gave a slight increase in the mean μ and standard deviation σ of reprojection errors: without the scene constraint, $\mu = 0.70$ pixels and $\sigma = 0.41$ pixels; with the scene constraint, $\mu = 0.84$ pixels and $\sigma = 0.51$ pixels.

In the second experiment, a video sequence of 120 images of a wall scene was captured by a hand held digital video camcorder (Sony DCR-PC100). The camera motion was mainly lateral (from left to right), but because of the walking of the cameraman, a small up and down camera motion was involved. In the last 50 frames, while the camera remained stationary, it zoomed slowly into the scene. The KLT feature tracker [11, 17] was applied to the entire video sequence to extract image feature points. Out of the video sequence, 7 images at 20 frames apart were selected. After manually removing all the outliers and adding a few image feature points used for the scene constraints, 140 image feature points in each image resulted. Figure 3 shows 4 of the 7 images (dimensions 768×576 pixels) of the video sequence, superimposed onto each of them are the image feature points.

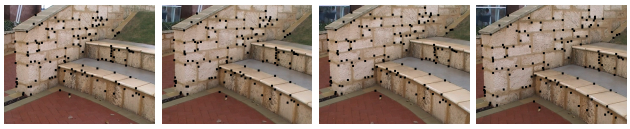

Figure 3. Images 1, 3, 5, 7 of Experiment 2.

Seven image feature points on top of the lower stair and twelve image feature points on the wall were randomly selected to form two sets of points on two orthogonal scene planes. The projective structure recovered for these sets of points was used as a scene constraint for the initial estimation and refinement of the absolute quadric.

Figure 4 shows the Euclidean reconstruction of the scene and the estimated positions of the camera. Those image feature points that were used for composing the scene constraints were labeled as blue $\circ$'s on the wall and as red $*$'s on one of the stairs. The angle between the wall and the

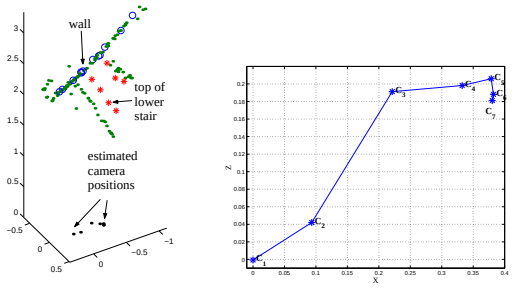

Figure 4. Plan views of the Euclidean reconstruction and the estimated camera positions using our scheme.

lower stair was estimated to be $93.98°$ (without using the scene constraint) and $90.16°$ (with the scene constraint) respectively. The estimated focal length and principal point for each of the 7 images are given in Table 3. Since no ground truth of the camera intrinsic parameters is available, the accuracy of the estimates shown in the table cannot be assessed. The increase in the estimates of the focal length from images 5 to 7, because of the zooming in of the camera, is clearly evident. The imposition of orthogonal scene plane constraints appears to give a slightly larger variation of the camera's principal point in this experiment.

Table 3. The estimated focal lengths and principal points of the camera in Experiment 2.

image	Without scene constraint		With scene constraint	
	f (pixels)	(u_0, v_0) (pixels)	f (pixels)	(u_0, v_0) (pixels)
1	1249.50	(328.29, 310.90)	1264.90	(332.86, 315.69)
2	1236.56	(342.50, 310.86)	1245.45	(357.25, 311.69)
3	1206.03	(379.02, 297.93)	1197.81	(408.41, 300.18)
4	1230.87	(380.63, 278.37)	1225.25	(423.21, 283.74)
5	1224.19	(396.19, 270.75)	1218.43	(442.92, 279.63)
6	1226.59	(390.20, 258.62)	1222.16	(436.54, 264.30)
7	1363.24	(398.74, 265.32)	1358.56	(449.25, 271.49)

5. Discussions

The final Euclidean structure obviously depends on the initial estimate of the absolute quadric obtained by the linear method. If the initial estimate of the absolute quadric is poor then the subsequent steps may not be able to proceed because the initial estimate of the focal length computed from Section 3.2 may be complex. This problem is mainly due to the fact that the initial approximation of the principal point being at the centre of the image buffer is a poor one. However, other factors, such as the distribution of corresponding points in the images and the depth variation of objects in the scene, all make a contribution to the accuracy of the initial estimate of the absolute quadric. Interestingly, in comparison with the approach that does not employ scene constraints, our scheme helps bring the initial estimate of the absolute quadric to the true solution.

We have only been able to show Euclidean structure in sparse 3D point form. If more dense correspondences of image feature points are identified then, after some post-processing on the 3D data, the Euclidean structure could be represented as 3D facets and would be more suitable for 3D visualization. We note that since the camera intrinsic parameters and its motion have been recovered, more accurate image point correspondences can be established.

6. Conclusions

We have described a scheme for combining auto-calibration and scene constraints for the structure-from-motion problem. It involves computation of the projective structure of the scene and estimation of the absolute quadric for Euclidean upgrade followed by bundle adjustment to statistically optimize the Euclidean reconstruction. Throughout all the steps, auto-calibration constraints are imposed. In addition, constraints from orthogonal scene planes are enforced in the initial estimation and refinement steps of the absolute quadric. Our experiments have shown that known scene constraints can be easily incorporated to improve the estimate of the absolute quadric and subsequently to obtain a more accurate Euclidean reconstruction.

References

[1] K. Åström, A. Heyden, F. Kahl, and M. Oskarsson. Structure and Motion from Lines under Affine Projections. In *Proc. ICCV*, vol. 1, pp. 285–292, 1999.

[2] O. Faugeras. Stratification of three-dimensional projective, affine and metric representations. *J. Opt. Soc. America*, 12(3):465–484, 1995.

[3] O. Faugeras, Q.-T. Luong, and S. Maybank. Camera self-calibration: Theory and experiments. In *Proc. ECCV*, pp. 321–334, 1992.

[4] O. D. Faugeras. What Can be Seen in Three Dimensions with an Uncalibrated Stereo Rig? In *Proc. ECCV*, pp. 565–578, 1992.

[5] R. I. Hartley. Estimation of Relative Camera Positions for Uncalibrated Cameras. In *Proc. ECCV*, pp. 579–587, 1992.

[6] R. I. Hartley. A Linear Method for Reconstruction from Lines and Points. In *Proc. ICCV*, pp. 882–887, 1995.

[7] A. Heyden and K. Åström. Euclidean reconstruction from constant intrinsic parameters. In *Proc. ICPR*, vol. 1, pp. 339–343. 1996.

[8] A. Heyden and K. Åström. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Proc. CVPR*, pp. 438–443, 1997.

[9] F. Kahl and A. Heyden. Using Conic Correspondences in Two Images to Estimate the Epipolar Geometry. In *Proc. ICCV*, pp. 761–766, 1998.

[10] D. Liebowitz and A. Zisserman. Combining scene and auto-calibration constraints. In *Proc. ICCV*, 1999.

[11] B. D. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *IJCAI*, pp. 674–679, 1981.

[12] Q.-T. Luong and O. Faugeras. Self-calibration of a moving camera from point correspondances and fundamental matrices. *IJCV*, 22(3):261–289, 1997.

[13] R. Mohr and E. Arbogast. It can be done without camera calibration. *Pattern Recognition Letters*, 12(1):39–43, 1991.

[14] M. Pollefeys, L. V. Gool, and M. Oosterlinck. The modulus constraint: A new constraint for self-calibration. In *Proc. ICPR*, pp. 349–353, 1996.

[15] M. Pollefeys, R. Koch, and L. V. Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proc. ICCV*, 1998.

[16] L. Quan. Conic Reconstruction and Correspondence from Two Views. *IEEE Trans. on PAMI*, 18(2):151–160, 1996.

[17] J. Shi and C. Tomasi. Good Features to Track. In *Proc. CVPR*, pp. 593–600, 1994.

[18] G. Sparr. Simultaneous Reconstruction of Scene Structure and Camera Locations from Uncalibrated Image Sequences. In *Proc. ICPR*, vol. 1, pp. 328–333, 1996.

[19] P. Sturm and B. Triggs. A Factorization Based Algorithm for Multi-Image Projective Structure and Motion. In *Proc. ECCV*, pp. 709–720, 1996.

[20] C. Tomasi and T. Kanade. Shape and Motion from Image Streams under Orthography: a Factorization Method. *IJCV*, 9(2):137–154, 1992.

[21] B. Triggs. Autocalibration and the Absolute Quadric. In *Proc. CVPR*, pp. 609–614, 1997.