

# Auto-calibration via the Absolute Quadric and Scene Constraints\*

A. Heyden  
School of Technology and Society  
Malmö University  
SE-20506 Malmö SWEDEN  
Email: heyden@ts.mah.se

D. Q. Huynh  
School of Information Technology  
Murdoch University  
Perth WA 6150 AUSTRALIA  
Email: d.huynh@murdoch.edu.au

## Abstract

*A scheme is described for incorporation of scene constraints into the structure from motion problem. Specifically, the absolute quadric is recovered with constraints imposed by orthogonal scene planes. The scheme involves a number of steps. A projective reconstruction is first obtained, followed by a linear technique to form an initial estimate of the absolute quadric. A nonlinear iteration then refines this quadric and the camera intrinsic parameters to upgrade the projective reconstruction to Euclidean. Finally, a bundle adjustment algorithm optimizes the Euclidean reconstruction to give a statistically optimal result. This chain of algorithms is essentially the same as used in auto-calibration and the novelty of this paper is the inclusion of orthogonal scene plane constraints in each step. The algorithms involved are demonstrated on both simulated and real data showing the performance and usability of the proposed scheme.*

## 1. Introduction

### 1.1. Problem description

One of the central problems in computer vision is the so-called "structure from motion problem". If no special information about the camera or the scene is available then only a projective reconstruction of the scene can be obtained, cf. [1, 11, 4]. Since this projective reconstruction might contain severe projective distortions, it is often desirable to obtain a Euclidean reconstruction (up to an unknown similarity transformation) of the scene.

Traditionally, there are two different ways to obtain the Euclidean structure of a scene. The first method, which relies on some a priori information about the scene, e.g. some distance or angular measurements, cf. [2], is often referred to as *stratification*, since one starts with a projective reconstruction and then finds an affine 'stratum' and finally a Euclidean 'stratum', giving the desired reconstruction. The other method, which relies on some a priori information about the intrinsic parameters, e.g. known skew and/or as-

pect ratio (see [10, 18, 6, 13]), is often referred to as *auto-calibration* since the main focus is on finding the intrinsic parameters, i.e. auto-calibrating the cameras, in addition to motion and structure recovery.

The purpose of this paper is to auto-calibrate a camera based on the natural camera model (i.e. unit aspect ratio and no skew) with the incorporation of constraints from orthogonal planes present in the scene. This is achieved via the recovery of the absolute quadric, with the orthogonal scene planes providing the additional equations to constrain this entity. The results are that a more accurate estimate of the absolute quadric is obtained, leading to smaller errors in the estimates of the camera intrinsic parameters and a more accurate Euclidean reconstruction.

The applicability of the proposed algorithm is manifold. First, with high quality digital and video cameras, it is often safe to assume vanishing skew and unit aspect ratio. If one is in doubt, it is always possible to assume constant skew and aspect ratio. The natural camera model is thus accurate for modelling the 3D to 2D projection of all modern cameras. This model reduces the number of intrinsic parameters to be recovered for each image (or camera) to 3: focal length and principal point. Furthermore, for a large number of applications, orthogonal scene planes arise naturally. For example, in images of man-made objects such as buildings, orthogonal walls can be easily detected, making it possible to incorporate such scene constraints into auto-calibration.

### 1.2. Previous work

The earliest work on auto-calibration is the algorithm by Faugeras, Luong and Maybank [3], where the case of constant intrinsic parameters is treated. Later auto-calibration work that involves absolute conic and absolute quadric includes [5, 7, 18, 12, 13]. Although the incorporation of scene constraints is known to be useful (see, e.g. [13]) in auto-calibration, none of these earlier reports have conducted detailed investigation into the use of scene constraints in auto-calibration. While Liebowitz and Zisserman [8] detect the image projections of parallel and orthogonal scene lines and use them to estimate the vanishing points and as constraints in auto-calibration, they apply the constraints to the absolute conic. Also, their work requires

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

the computation of the fundamental matrix, and that limits their method to the use of two images.

## 2. The proposed scheme

The proposed scheme is divided into five steps: projective reconstruction, solving for the absolute quadric, refining the absolute quadric, initial linear Euclidean upgrade, and bundle adjustment with scene constraints. Details of these steps are described below.

### 2.1. Projective reconstruction

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

$$\begin{aligned} \lambda^j \mathbf{x}^j &= \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} [R \ -Rt] \mathbf{X}^j \\ \Leftrightarrow \lambda^j \mathbf{x}^j &= K [R \ -Rt] \mathbf{X}^j, \end{aligned} \quad (1)$$

where the superscript  $j$  denotes the  $j^{\text{th}}$  scene (or image) point,  $\lambda^j$  is an unknown scalar. The camera matrix, denoted by  $K$ , embodies the unknown camera focal length  $f$  and principal point  $(u_0, v_0)$ . The motion matrix contains the unknown rotation matrix  $R$  and translation vector  $t$  of the camera relative to a coordinate system. The special form of  $K$  here arises from the use of the natural camera model. In the situation where 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  $P \in \mathbb{R}^{3 \times 4}$  is a projection matrix. With the availability of  $m$  images and  $n$  scene points, the joint projection matrix  $\mathbf{P} \in \mathbb{R}^{3m \times 4}$ , the joint image measurement matrix  $\mathbf{x} \in \mathbb{R}^{3m \times n}$ , and the joint shape matrix  $\mathbf{X} \in \mathbb{R}^{4 \times n}$  are related by

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

where the subscript  $i$  indicates the  $i^{\text{th}}$  camera, and  $\lambda_i^j$ 's, known as the *projective depths*, are unknown scalars.

Setting all the  $\lambda_i^j$  to 1 for the affine camera model, Tomasi and Kanade[17] pioneered the factorization method to retrieve the  $\mathbf{P}$  and  $\mathbf{X}$  from  $\mathbf{x}$ . For the projective camera model, the values of  $\lambda_i^j$ 's must be recovered prior to factorization (see [15, 16]). We adopt the method of [15] in our scheme. First, all the  $\lambda_i^j$ 's are assumed to be 1. At each iteration, the  $\lambda_i^j$ 's are refined with the subspace (4D space of  $\mathbb{R}^n$ ) 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  $\lambda_i^j$ 's and re-factorized to give new  $\mathbf{P}$  and  $\mathbf{X}$  matrices for the next iteration. The method has shown to give very good estimates of the  $\lambda_i^j$ 's and very fast convergence.

### 2.2. Solving for the absolute quadric

The structure contained in the shape matrix  $\mathbf{X}$  is projective only, since for any  $\tilde{\mathbf{P}}$  and  $\tilde{\mathbf{X}}$  matrices that satisfy (2),  $\tilde{\mathbf{P}}\tilde{\mathbf{A}}$  and  $\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{X}}$  are also a solution, where  $\tilde{\mathbf{A}} \in \mathbb{R}^{4 \times 4}$  is any non-singular matrix. So, an appropriate  $\tilde{\mathbf{A}}$  matrix satisfying

$$P_i \tilde{\mathbf{A}} \sim K_i [R_i \ -R_i t_i], \quad \text{for } i = 1, \dots, m, \quad (3)$$

must be estimated to upgrade the structure to Euclidean. Let  $P_1 = [I \ 0]$  and  $P_i = [Q_i \ \mathbf{q}_i]$ , for  $i = 2, \dots, m$ . Then  $\tilde{\mathbf{A}}$  takes the form  $\begin{bmatrix} K_1 & \mathbf{0} \\ \mathbf{a}^T & s \end{bmatrix}$ , where  $\mathbf{a} = (a_1, a_2, a_3)^T$  and  $s$  is a non-zero scalar often set to unity. Let  $\tilde{\mathbf{A}}$  be the matrix that contains the first three columns of  $\tilde{\mathbf{A}}$ . It follows that  $P_i \tilde{\mathbf{A}} \tilde{\mathbf{A}}^T P_i^T \sim K_i K_i^T$ , and so

$$\begin{aligned} P_i \begin{bmatrix} K_1 K_1^T & K_1 \mathbf{a} \\ \mathbf{a}^T K_1^T & \|\mathbf{a}\|^2 \end{bmatrix} P_i^T &\sim K_i K_i^T \\ \Leftrightarrow P_i \Omega P_i^T &\sim K_i K_i^T, \quad \text{for } i = 1, \dots, m, \end{aligned} \quad (4)$$

where  $\Omega$  is the *absolute quadric* or the *singular dual quadric* that contains the coordinates,  $\mathbf{a}$ , of the plane at infinity for affine reconstruction and the DIAC (dual image of the absolute conic),  $K_1 K_1^T$ , for Euclidean reconstruction. Furthermore,  $\Omega$  relates the angle  $\theta$  between the projective coordinates of any two scene planes,  $\mathbf{n}$  and  $\mathbf{m}$ , by

$$\cos(\theta) = \mathbf{n}^T \Omega \mathbf{m} / \left( \sqrt{\mathbf{n}^T \Omega \mathbf{n}} \sqrt{\mathbf{m}^T \Omega \mathbf{m}} \right). \quad (5)$$

As the number of unknowns and available equations in (4) are  $6 + 3(m-1)$  and  $5(m-1)$ , the introduction of only one pair of scene planes would reduce the minimum value of  $m$  to 2.

Solving for the above unknowns from (4) is a difficult nonlinear problem. An alternative is, as suggested in [13], to use a special case of the natural camera model, namely,  $(u_{0i}, v_{0i}) = (0, 0)$  for all  $i$ . The diagonal form of  $K_i K_i^T$  and the equality of its first two diagonal elements then give 4 linear constraints on  $\Omega$ . Equation (5) can be simplified further to  $\mathbf{n}^T \Omega \mathbf{m} = 0$ , if  $\theta = 90^\circ$ , to linearize the equations from the orthogonal scene planes. In our experiments reported here, we included at least 5 images from each video sequence to recover  $\Omega$ . 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.

### 2.3. Refining the absolute quadric

Writing (4) and (5) in matrix-vector form, we have the following objective functions and constraints:

$$\min_{\substack{a_1, a_2, a_3, a_4, \\ f_1, v_{01}, v_{01}}} J_1 = \left\| \frac{M_i \mathbf{q}}{\sqrt{(M_i \mathbf{q})^T (M_i \mathbf{q})}} - \frac{\mathbf{k}_i}{\sqrt{\mathbf{k}_i^T \mathbf{k}_i}} \right\|^2 \quad (6)$$

$$\min_{\substack{a_1, a_2, a_3, a_4, \\ f_1, v_{01}, v_{01}}} J_2 = \frac{(N_{jt} \mathbf{q})^T (N_{jt} \mathbf{q})}{(N_{jj} \mathbf{q})(N_{tt} \mathbf{q})}, \quad (7)$$

for  $i = 2, \dots, m$ . Here, each  $M_i \in \mathbb{R}^{6 \times 10}$ ,  $N_{ji} \in \mathbb{R}^{1 \times 10}$ , and  $\mathbf{q} \in \mathbb{R}^{10}$  is the vector form containing the elements of  $\Omega$ . Similarly,  $\mathbf{k}_i \in \mathbb{R}^6$  contains the elements of  $K_i K_i^\top$ . The constraint from any two orthogonal planes  $\mathbf{n}_j$  and  $\mathbf{n}_i$  with respect to  $\Omega$  can be written in the form given by (7).

It is useful to apply the Levenberg-Marquardt method to solve the above constrained minimization problem. We introduce the vector  $\mathbf{y}_i$  that embodies all the parameters to be refined at the  $i^{\text{th}}$  iteration. Let  $J = [J_1, \dots, J_2, \dots]$  be the collection of all the minimization functions in (6) and (7), and let  $\mathbf{j}_i \equiv J(\mathbf{y}_i)$  be the residuals at the same iteration. The update  $\Delta \mathbf{y}$  for the vector  $\mathbf{y}_i$  at the  $i^{\text{th}}$  iteration is given by  $\Delta \mathbf{y} = (B^\top B + \epsilon I)^{-1} B^\top \mathbf{j}_i$  where  $B = \frac{\partial J}{\partial \mathbf{y}}(\mathbf{y}_i)$  and  $\epsilon$  is a small positive number. At the  $(i+1)^{\text{th}}$  iteration, we have  $\mathbf{y}_{i+1} = \mathbf{y}_i + \Delta \mathbf{y}$  and so on.

#### 2.4. Initial linear Euclidean upgrade

The objective of this step is to estimate matrix  $A$ . Since  $\tilde{A}\tilde{A}^\top \sim \Omega$ , the easiest way to compute  $A$  is to let  $\tilde{A} \equiv U_3 S_3^{1/2}$  where  $\Omega \equiv USV^\top$  is the SVD of  $\Omega$  and  $U_3, S_3$  are the matrices containing the first 3 columns of  $U$  and  $S$ . By means of the intrinsic parameters refined from (6) and the recovered  $\tilde{A}$  matrix,  $K_i$  can be constructed and  $R_i, \mathbf{t}_i$ , for  $i = 1, \dots, m$  can all be computed. The projective structure  $\mathbf{X}$  estimated from Section 2.1 is then upgraded to  $\mathbf{X}_e \equiv A^{-1}\mathbf{X}$ .

#### 2.5. Bundle adjustment with scene constraints

The initial Euclidean reconstruction obtained above can be improved further by incorporating the reconstructed 3D points and camera intrinsic and extrinsic parameters and minimizing the reprojection errors. To impose orthogonal scene plane constraints into this bundle adjustment, each iteration can be broken into two separate steps:

**Step A:** minimizing the reprojection errors. This is the normal bundle adjustment process.

**Step B:** incorporating constraints of orthogonal scene planes. In this step, all the parameters refined by Step A are fed into a similar operation as described in Section 2.3. The differences are: (1) all the scene plane coordinates must be recomputed, using the Euclidean structure estimated from Step A above, and (2) the absolute quadric  $\Omega$  is replaced with the *update absolute quadric*  $\delta\Omega$  whose initial estimate  $\delta\Omega_0$  is set to  $\begin{bmatrix} I & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{bmatrix}$ . At each iteration, the refined  $\delta\Omega$  is used to update all the intrinsic and extrinsic parameters similar to the procedure described in Section 2.4. To ensure that the camera coordinate systems used in Step A do not undergo major changes due to  $\delta\Omega$ , the update matrix  $\delta\tilde{A}$  (analogous to  $\tilde{A}$  in Section 2.4) is defined as

$$\delta\Omega \equiv \begin{bmatrix} \delta K \delta K^\top & \delta K \delta \mathbf{a} \\ \delta \mathbf{a}^\top \delta K^\top & \|\delta \mathbf{a}\|^2 \end{bmatrix} \Rightarrow \delta\tilde{A} = \begin{bmatrix} \delta K \\ \delta \mathbf{a}^\top \end{bmatrix}.$$

The updated Euclidean structure, intrinsic and extrinsic parameters are then fed back into Step A for the next bundle adjustment iteration.

The rate of convergence for the above modified bundle adjustment is very fast. The extra process in Step B above appears not only to help retain the orthogonality of the scene planes but, on average, give better estimates of all the parameters and smaller 3D reconstruction errors.

### 3. Experiments

To demonstrate that using all the available constraints increases the accuracy of the reconstruction, we first conducted experiments using synthetic data. In each experiment, we compared the relative errors of the focal lengths, principal points, reconstruction errors, and orthogonality errors computed from (i) *auto-calibration without scene constraints*, which does not incorporate any orthogonal scene plane constraints given in (5), (7), and Step B described in Section 2.5, and (ii) *auto-calibration with scene constraints*, which incorporates scene constraints described in Sections 2.2, 2.3, and 2.5. The means and standard deviations of these errors for 20 experiments listed in Table 1 clearly show that a smaller mean reconstruction error (0.4895 versus 1.1095) and a mean orthogonality error (1.05% versus 3.29%) could be achieved by incorporating orthogonal scene plane constraints (if they are available). While a slightly larger mean principal point error was incurred, a smaller focal length error, 0.04%, (5.85% versus 5.81%) also resulted from the use of scene constraints.

|                         | mean                      |                        | std deviation             |                        |
|-------------------------|---------------------------|------------------------|---------------------------|------------------------|
|                         | without scene constraints | with scene constraints | without scene constraints | with scene constraints |
| $\epsilon_f$            | 0.0585                    | 0.0581                 | 0.0472                    | 0.0464                 |
| $\epsilon_{\mathbf{X}}$ | 1.1095                    | 0.4895                 | 0.9284                    | 0.4154                 |
| $\epsilon_{(u_0, v_0)}$ | 19.7890                   | 24.1697                | 10.5831                   | 13.8939                |
| $\epsilon_\theta$       | 0.0329                    | 0.0105                 | 0.0391                    | 0.0123                 |

Table 1: The means and standard deviations of a few error measures with and without the use of scene constraints:  $\epsilon_f = |\hat{f} - \bar{f}|/\bar{f}$  (relative error of estimated focal length in pixels);  $\epsilon_{\mathbf{X}} = \|\hat{\mathbf{X}} - \bar{\mathbf{X}}\|$  (reconstruction error);  $\epsilon_{(u_0, v_0)} = \|(\hat{u}_0, \hat{v}_0) - (\bar{u}_0, \bar{v}_0)\|$  (principal point error in pixels);  $\epsilon_\theta = |\hat{\theta} - 90|/90$  (relative error on orthogonality). The symbols  $\hat{\cdot}$  and  $\bar{\cdot}$  denote the estimated and true values of the entity.

For the experiments on real video data, we used images captured by a Sony DCR-PC100 digital video camcorder. Due to the limitation of space, we report only one of the experiments conducted. Figure 1 shows 3 images of a video sequence of a house. At the beginning of the sequence, the camera moved from left to right; in the last part of the sequence (about 20 frames), the camera stopped moving but zoomed slowly in to the scene. The KLT feature

tracker [9, 14] was used to track the image feature points in the video sequence. 324 image feature points were detected and 9 images were selected from frames 20 to 100, at every 10<sup>th</sup> frame interval.



Figure 1: Frames 20, 90, 100 of a video sequence of a house with the tracked image features superimposed.

To incorporate scene constraints into the Euclidean reconstruction, 22 and 32 image points of two orthogonal scene planes (the two walls of the house) were manually selected. The coordinates of these two planes from the projective reconstruction (step 1) were computed and were used to constrain the initial estimate and refinement of the absolute quadric. Figure 2 shows the Euclidean reconstruction of the house and the estimated positions of the camera in the scene. Because of the large number of reconstructed scene points and the lack of a texture mapping software for 3D visualization, we only display one third of the reconstructed scene points of and surrounding the house to demonstrate that orthogonality of the walls was retained. From frames 80 to 100 the estimated focal length increased, in accordance to the zooming in to the scene of the camera. The angle between the two walls were estimated to be 75.20° (without scene constraints) and 90.22° (with scene constraints).

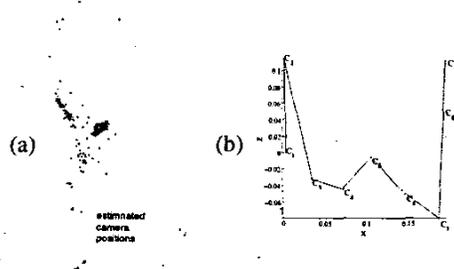


Figure 2: (a) The top view of the Euclidean reconstruction. Some feature points on the two walls of the house are labelled as '\*' and '+'. (b) An enlarged view of the estimated camera positions (projected onto the ground plane).

## 4. Conclusions

We have described a scheme for incorporating orthogonal scene plane constraints into the auto-calibration problem. It involves computing the projective structure of the scene and the estimation of the absolute quadric for Euclidean upgrade

followed by bundle adjustment to statistically optimize the Euclidean reconstruction. Throughout all the steps in the scheme, scene constraints are enforced. Our synthetic and real experiments have shown that known scene constraints can be easily incorporated to improve the estimate of the absolute quadric and subsequently to attain a more accurate Euclidean reconstruction.

## References

- [1] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proc. ECCV*, pp. 563–578, 1992.
- [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] R. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proc. ECCV*, pp. 579–587, 1992.
- [5] A. Heyden and K. Åström. Euclidean reconstruction from constant intrinsic parameters. In *Proc. ICPR*, vol 1, pp. 339–343, 1996.
- [6] 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.
- [7] A. Heyden and K. Åström. Flexible calibration: Minimal cases for auto-calibration. In *Proc. ICCV*, pp. 350–355, 1999.
- [8] D. Liebowitz and A. Zisserman. Combining scene and auto-calibration constraints. In *Proc. ICCV*, 1999.
- [9] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI*, pp. 674–679, 1981.
- [10] Q.-T. Luong and O. Faugeras. Self-calibration of a moving camera from point correspondences and fundamental matrices. *IJCV*, 22(3):261–289, 1997.
- [11] R. Mohr and E. Arbogast. It can be done without camera calibration. *Pattern Recognition Letters*, 12(1):39–43, 1991.
- [12] M. Pollefeys, L. V. Gool, and M. Oosterlinck. The modulus constraint: A new constraint for self-calibration. In *Proc. ICPR*, pp. 349–353, 1996.
- [13] 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.
- [14] J. Shi and C. Tomasi. Good features to track. In *Proc. CVPR*, pp. 593–600, 1994.
- [15] G. Sparr. Simultaneous reconstruction of scene structure and camera locations from uncalibrated image sequences. In *Proc. ICPR*, 1996.
- [16] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *Proc. ECCV*, pp. 709–720, 1996.
- [17] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *IJCV*, 9(2):137–154, 1992.
- [18] B. Triggs. Autocalibration and the absolute quadric. In *Proc. CVPR*, 1997.