Self-Calibration and Metric Reconstruction Inspite of Varying and Unknown Intrinsic Camera Parameters

MARC POLLEFEYS, REINHARD KOCH AND LUC VAN GOOL
ESAT-PSI, K.U. Leuven, Kardinaal Mercierlaan 94, B-3001 Heverlee, Belgium
Marc.Pollefeys@esat.kuleuven.ac.be
Reinhard.Koch@esat.kuleuven.ac.be
Luc.VanGool@esat.kuleuven.ac.be

Abstract. In this paper the theoretical and practical feasibility of self-calibration in the presence of varying intrinsic camera parameters is under investigation. The paper's main contribution is to propose a self-calibration method which efficiently deals with all kinds of constraints on the intrinsic camera parameters. Within this framework a practical method is proposed which can retrieve metric reconstruction from image sequences obtained with uncalibrated zooming/focusing cameras. The feasibility of the approach is illustrated on real and synthetic examples. Besides this a theoretical proof is given which shows that the absence of skew in the image plane is sufficient to allow for self-calibration. A counting argument is developed which—depending on the set of constraints—gives the minimum sequence length for self-calibration and a method to detect critical motion sequences is proposed.

Keywords: self-calibration, metric 3D reconstruction, uncalibrated image sequences, varying camera parameters

1. Introduction

In recent years, researchers have been studying self-calibration methods for cameras. Mostly completely unknown but constant intrinsic camera parameters were assumed. This has the disadvantage that the zoom can not be used and even focusing is prohibited. On the other hand, the proposed perspective model is often too general compared to the range of existing cameras. Mostly the image axes can be assumed orthogonal and often the aspect ratio is known. Therefore a tradeoff can be made and by assuming these parameters to be known, one can allow (some of) the other parameters to vary throughout the image sequence.

Since it became clear that projective reconstructions could be obtained from image sequences alone (Faugeras, 1992; Hartley, 1992), researchers tried to find ways to upgrade these reconstructions to metric (i.e., Euclidean up to unknown scale). Many methods were developed which assumed constant intrinsic camera parameters. Most of these methods are based on

the absolute conic which is the only conic which stays fixed under all Euclidean transformations (Semple and Kneebone, 1995). This conic lies in the plane at infinity and its image is directly related to the intrinsic camera parameters, hence the advantage for self-calibration.

Faugeras et al. (1992), see also (Luong and Faugeras, 1997), proposed to use the Kruppa equations which enforce that the planes through two camera centers, which are tangent to the absolute conic, should also be tangent to both of its images. Later on Zeller and Faugeras (1996) proposed a more robust version of this method.

Heyden and Åström (1996), Triggs (1997) and Pollefeys and Van Gool (1997b) use explicit constraints which relate the absolute conic to its images. These formulations are especially interesting since they can easily be extended to deal with varying intrinsic camera parameters.

Pollefeys and Van Gool (1997a) also proposed a stratified approach which consists of first locating the

plane at infinity using the modulus constraint (i.e., for constant intrinsic camera parameters the infinity homography should be conjugated to a rotation matrix) and then calculating the absolute conic. Hartley (1994) proposed another approach based on the minimization of the difference between the intrinsic camera parameters for the different views.

So far not much work has been done on varying intrinsic camera parameters. Pollefeys et al. (1996) also proposed a stratified approach for the case of a varying focal length, but this method required a pure translation as initialization, along the lines of Armstrong et al. (1994) earlier account for fixed intrinsic camera parameters. Recently Heyden and Åström (1997) proved that self-calibration was possible when the aspect ratio was known and no skew was present. The self-calibration method proposed in their paper is based on bundle adjustment which requires non-linear minimization over all reconstructed points and cameras simultaneously. No method was proposed to obtain a suitable initialization.

In this paper their proof is extended. It will be shown that the absence of skew alone is enough to allow self-calibration. A versatile self-calibration method is proposed which can deal with varying types of constraints. This will then be specialized towards the case where the focal length varies, possibly also the principal point.

Section 2 of this paper introduces notations and some basic principles, Section 3 gives a counting argument for self-calibration and finally shows that imposing the absence of skew is sufficient to restrict the projective ambiguity to the group of similarities (i.e., metric self-calibration). In Section 4 the actual method is developed. A simplified linear version is also given which can be used for initialization. Section 5 summarizes the complete procedure for metric reconstruction of arbitrarily shaped, rigid objects from an uncalibrated image sequence alone. The method is then validated through the experiments of Section 6, in Section 7 some more results illustrate the flexibility of our approach. Section 8 concludes this paper and gives some directions for further research.

2. Notations and Basic Principles

In this section the basic principles and notations used in this paper are introduced. Projective geometry and homogeneous coordinates are used. Metric entities are indicated with a subscript $\mathcal{M}$.

2.1. Cameras

The following equation is used to describe the perspective projection of the scene onto the images

$$m \propto \mathbf{P}M \tag{1}$$

where $\mathbf{P}$ is a 3×4 projection matrix describing the perspective projection process, $M = [X\,Y\,Z\,1]^\top$ and $m = [x\,y\,1]^\top$ are vectors containing the homogeneous coordinates of the world points respectively image points. Notice that $\propto$ will be used throughout this paper to indicate equality up to a non-zero scale factor.

In the metric case the camera projection matrix factorizes as follows:

$$\mathbf{P}_\mathcal{M} = \mathbf{K}[\mathbf{R} \mid -\mathbf{R}t] \tag{2}$$

Here $(\mathbf{R}, t)$ denotes a rigid transformation (i.e., $\mathbf{R}$ is a rotation matrix and t is a translation vector) which indicate the position and orientation of the camera, while the upper triangular calibration matrix $\mathbf{K}$ encodes the intrinsic parameters of the camera:

$$\mathbf{K} = \begin{bmatrix} f_x & s & u_x \\ & f_y & u_y \\ & & 1 \end{bmatrix} \tag{3}$$

where f_x and f_y represent the focal length divided by the pixel width resp. height, (u_x, u_y) represents the principal point and s is a factor which is zero in the absence of skew.

2.2. Conics and Quadrics

In this paper a specific conic and quadric play an important role. Therefore some related notations are introduced here. A conic is represented by a 3×3 symmetric matrix $\mathbf{C}$, a quadric by a 4×4 symmetric matrix $\mathbf{Q}$. A point m on the conic satisfies $m^\top \mathbf{C} m = 0$ and a point M on the quadric satisfies $M^\top \mathbf{Q} M = 0$. A dual (or line) conic is represented by a 3×3 matrix $\mathbf{C}^*$, while a dual (or plane) quadric is represented by a 4×4 matrix $\mathbf{Q}^*$. A line l tangent to the conic $\mathbf{C}$ satisfies $l^\top \mathbf{C}^* l = 0$. A plane L tangent to the quadric $\mathbf{Q}$ satisfies $L^\top \mathbf{Q}^* L = 0$. Provided $\mathbf{C}$ resp. $\mathbf{Q}$ are full rank $\mathbf{C}^* = \mathbf{C}^{-1}$ and $\mathbf{Q}^* = \mathbf{Q}^{-1}$. It can be shown that (Karl et al., 1994),

$$\mathbf{C}^* \propto \mathbf{P}\mathbf{Q}^*\mathbf{P}^\top. \tag{4}$$

This equation describes the projection of the outline of a dual quadric onto a dual image conic.

2.3. Transformations

A projective transformation of 3D space is described by a 4×4 matrix $\mathbf{T}$. In the case of a similarity transformation $\mathbf{T}_\mathcal{M}$ takes on the following form:

$$\mathbf{T}_\mathcal{M} = \begin{bmatrix} \sigma \mathbf{R} & t \\ 0 & 1 \end{bmatrix} \tag{5}$$

with σ a global scale factor ($\sigma = 1$ yields a Euclidean transformation).

Points, planes and projection matrices are transformed as follows:

$$M \to \mathbf{T}M, \quad L \to \mathbf{T}^{-\top}L \quad \text{and} \quad \mathbf{P} \to \mathbf{P}\mathbf{T}^{-1}. \tag{6}$$

On quadrics and dual quadrics the effect of a projective transformation is as follows:

$$\mathbf{Q} \to \mathbf{T}^{-\top}\mathbf{Q}\mathbf{T}^{-1} \quad \text{and} \quad \mathbf{Q}^* \to \mathbf{T}\mathbf{Q}^*\mathbf{T}^\top. \tag{7}$$

3. Some Theory

Before developing a practical self-calibration algorithm some theoretical aspects of the problem are studied in this section. First a counting argument is given which states the minimal sequence length that allows self-calibration from a specific set of constraints. Then a proof is given that self-calibration is possible for the minimal case where the only available constraint is the absence of skew.

3.1. A Counting Argument

To restrict the projective ambiguity (15 degrees of freedom) to a metric one (3 degrees of freedom for rotation, 3 for translation and 1 for scale), at least 8 constraints are needed. This, thus determines the minimum length of a sequence from which self-calibration can be obtained, depending on the type of constraints which are available for each view. Knowing an intrinsic camera parameter for n views gives n constraints, fixing one yields only $n - 1$ constraints.

$$n \times (\#known) + (n - 1) \times (\#fixed) \geq 8$$

Table 1. A few examples of minimum sequence length required to allow self-calibration.

Constraints	Known	Fixed	Min no. of images
No skew	s		8
Fixed aspect ratio and absence of skew	s	$\frac{f_y}{f_x}$	5
Known aspect ratio and absence of skew	$s, \frac{f_y}{f_x}$		4
Only focal length is unknown	$s, \frac{f_y}{f_x}, ux, uy$		2
Standard self-calibration problem		f_x, f_y, ux, uy, s	3

Of course this counting argument is only valid for non-critical motion sequences (see Section 4.3).

Therefore, the absence of skew (1 additional constraint per view) should in general be enough to allow self-calibration on a sequence of 8 or more images. In Section 3.2 it will be shown that this simple constraint is not bound to be degenerate. If in addition the aspect ratio is known (e.g., $f_x = f_y$) then four views should be sufficient. When also the principal point is known, a pair of images is enough. A few more examples are given in Table 1.

3.2. Self-Calibration Using Only the Absence of Skew

In this paragraph it is shown that the absence of skew can be sufficient to yield a metric reconstruction. This is an extension of the theorem proposed by Heyden and Åström (1997) which besides orthogonality also requires the aspect ratio to be known.

Theorem 1. *The class of transformations which preserves the absence of skew is the group of similarity transformations.*

The proof is given in the appendix. If a sequence is general enough (in its motion) it follows from this theorem that only a projective representation of the cameras which can be related to the original ones through a similarity transformation (possibly including a mirroring) would satisfy the orthogonality of rows and columns for all views. Using oriented projective geometry (Laveau and Faugeras, 1996) the mirroring ambiguity can easily be eliminated. Therefore self-calibration and metric

reconstruction is possible using this orthogonality constraint only.

Of course adding more constraints will yield more robust results and will diminish the probability of encountering critical motion sequences.

4.   Self-Calibration

It is a well-known result that from image correspondences alone the camera projection matrices and the reconstruction of the scene points can be retrieved up to a projective transformation (Faugeras, 1992; Hartley, 1992). Note that without additional constraints nothing more can be achieved. This can be seen from the following equation.

$$m \propto \mathbf{P}M = (\mathbf{P}\mathbf{T}^{-1})(\mathbf{T}M) = \mathbf{P}'M'$$

with $\mathbf{T}$ an arbitrary projective transformation. Therefore $(\mathbf{P}', M')$ is also a valid reconstruction from the image points m.

In general, however, some additional constraints are available. Some intrinsic parameters are known or can be assumed constant. This yields constraints which should be verified when $\mathbf{P}$ is factorized as in Eq. (2).

It was shown that when no skew is present, the ambiguity of the reconstruction can be restricted to metric (see Section 3.2). Although this is theoretically sufficient, under practical circumstances often more constraints are available and should be used.

In Euclidean space two entities are invariant—setwise, not pointwise—under rigid transformations. The first one is the plane at infinity Π_∞ which allows to compute affine measurements. The second entity is the absolute conic ω which is embedded in the plane at infinity. If besides the plane at infinity Π_∞ the absolute conic ω has also been localized, metric measurements are possible.

When looking at a static scene from different viewpoints the relative position of the camera towards Π_∞ and ω is invariant. If the motion is general enough, only one conic in one specific plane will satisfy this condition. The absolute conic can therefore be used as a virtual calibration pattern which is always present in the scene.

A practical way to encode both the absolute conic ω and the plane at infinity Π_∞ is through the use of the absolute quadric Ω^* (Semple and Kneebone, 1952) (introduced in computer vision by Triggs (1997), see also Heyden and Åström (1996), Pollefeys and Van Gool
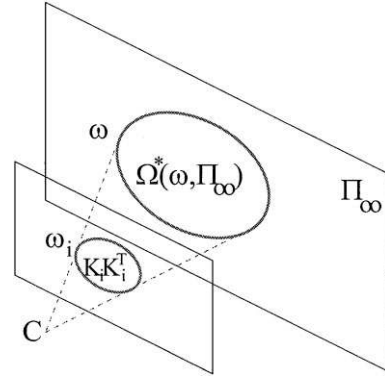

Figure 1.   The absolute quadric Ω^* which encodes both the plane at infinity Π_∞ (affine reference entity) and the absolute conic ω (metric reference entity), projects to the dual image of the absolute conic $\omega_i = \mathbf{K}_i\mathbf{K}_i^\top$. The projection equation allows to translate constraints on the intrinsic parameters to constraints on Ω^*.

(1997b). This dual quadric consists of planes tangent to the absolute conic. Its null-space is the plane at infinity. In a metric frame it is represented by a 4×4 symmetric rank 3 matrix $\Omega^*_\mathcal{M} = [\begin{smallmatrix} \mathbf{I} & 0 \\ 0 & 0 \end{smallmatrix}]$. Using (5) and (7) it can be verified that for a similarity transformation $\mathbf{T}_\mathcal{M}\Omega^*_\mathcal{M}\mathbf{T}_\mathcal{M}^\top \propto \Omega^*_\mathcal{M}$. Similar to (4) the projection of the absolute quadric in the image yields the dual image absolute conic:

$$\omega_i^* = \mathbf{K}_i\mathbf{K}_i^\top \propto \mathbf{P}_i\Omega^*\mathbf{P}_i^\top \qquad (8)$$

independent of the chosen projective basis. Using Eq. (2) this can be verified for a metric basis. Through Eqs. (6) and (7), Eq. (8) can then be verified for any projective basis. Some of these concepts are illustrated in Fig. 1.

Therefore, constraints on the intrinsic camera parameters in $\mathbf{K}_i$ can be translated to constraints on the absolute quadric. If enough constraints are at hand only one quadric will satisfy them all, i.e., the *absolute quadric*. At that point the scene can be transformed to a metric frame (which brings Ω^* to its canonical form).

4.1.   Non-Linear Approach

Equation (8) can be used to obtain the metric calibration from the projective one. The dual image absolute conics ω_i^* should be parameterized in such a way that they enforce the constraints on the calibration parameters. For the absolute quadric Ω^* a minimum parameterization (8 parameters) should be used. This can be

done by putting $\Omega^*_{33} = 1$ and by calculating Ω^*_{44} from the rank 3 constraint. The following parametrization satisfies these requirements:

$$\Omega^* = \begin{bmatrix} \mathbf{KK}^\top & -\mathbf{KK}^\top a \\ -a^\top \mathbf{KK}^\top & a^\top \mathbf{KK}^\top a \end{bmatrix}. \qquad (9)$$

Here a defines the position of the plane at infinity $\Pi_\infty = [\, a^\top \; 1\,]^\top$. In this case the transformation from projective to metric is particularly simple:

$$\mathbf{T}_{\mathcal{P} \to \mathcal{M}} = \begin{bmatrix} \mathbf{K}^{-1} & 0 \\ a^\top & 1 \end{bmatrix} \qquad (10)$$

An approximate solution to these equations can be obtained through non-linear least squares. The following criterion should be minimized:

$$\min \sum_{i=1}^n \left\| \frac{\mathbf{K}_i \mathbf{K}_i^\top}{\left\| \mathbf{K}_i \mathbf{K}_i^\top \right\|_F} - \frac{\mathbf{P}_i \Omega^* \mathbf{P}_i^\top}{\left\| \mathbf{P}_i \Omega^* \mathbf{P}_i^\top \right\|_F} \right\|_F^2. \qquad (11)$$

Remark that to obtain meaningful results $\mathbf{K}_i \mathbf{K}_i^\top$ and $\mathbf{P}_i \Omega^* \mathbf{P}_i^\top$ should both be normalized to have Frobenius norms equal to one.

If one chooses $\mathbf{P}_1 = [\mathbf{I} \mid \mathbf{0}]$, Eq. (8) can be rewritten as follows:

$$\mathbf{K}_i \mathbf{K}_i^\top \propto \mathbf{P}_i \begin{bmatrix} \mathbf{K}_1 \mathbf{K}_1^\top & -\mathbf{K}_1 \mathbf{K}_1^\top a^\top \\ -a \mathbf{K}_1 \mathbf{K}_1^\top & a \mathbf{K}_1 \mathbf{K}_1^\top a^\top \end{bmatrix} \mathbf{P}_i^\top \qquad (12)$$

In this way 5 of the 8 parameters of the absolute conic are eliminated at once, which simplifies convergence issues. On the other hand this formulation implies a bias towards the first view since using this parameterization the equations for the first view are perfectly satisfied, whereas the noise has to be spread over the equations for the other views. In the experiments it will be seen that this is not suitable for longer sequences where in this case the present redundancy can not be used optimally. Therefore it is proposed to first use the simplified version of Eq. (12) and then to refine the results with the unbiased parameterization.

To apply this self-calibration method to standard zooming/focusing cameras, some assumptions should be made. Often it can be assumed that there is no skew and that the aspect ratio is tuned to one. If necessary (e.g., when only a short image sequence is at hand, when the projective calibration is not accurate enough or when the motion sequence is close to critical without additional constraints), it can also be used that the

principal point is close to the center of the image. This leads to the following parameterizations for $\mathbf{K}_i$ (transform the images to have $(0, 0)$ in the middle):

$$\mathbf{K}_i = \begin{bmatrix} f & 0 & u \\ & f & v \\ & & 1 \end{bmatrix} \quad \text{or} \quad \mathbf{K}_i = \begin{bmatrix} f & 0 & 0 \\ & f & 0 \\ & & 1 \end{bmatrix}. \qquad (13)$$

These parameterizations can be used in (11). It will be seen in the experiments of Section 6 that this method gives good results on synthetic data as well as on real data.

4.2. Linear Approach

In the case where, besides the skew ($s = 0$), both principal point and aspect ratio are (approximately) known, a linear algorithm can be obtained by transforming the principal point $(u, v) \to (0, 0)$ and the aspect ratio $\frac{f_y}{f_x} \to 1$. These assumptions simplify (12) as follows:

$$\lambda \begin{bmatrix} f_i^2 & 0 & 0 \\ 0 & f_i^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{P}_i \begin{bmatrix} b_1 & 0 & 0 & b_2 \\ 0 & b_1 & 0 & b_3 \\ 0 & 0 & 1 & b_4 \\ b_2 & b_3 & b_4 & b_5 \end{bmatrix} \mathbf{P}_i^\top \qquad (14)$$

with $b_1 = f_1^2$, $b_2 = -f_1^2 a_1$, $b_3 = -f_1^2 a_2$, $b_4 = -a_3$ and $b_5 = f_1^4(a_1^2 + a_2^2) + a_3^2$. From the left-hand side of Eq. (14) it can be seen that the following equations have to be satisfied:

$$\omega^*_{11} = \omega^*_{22}, \qquad (15)$$

$$\omega^*_{12} = \omega^*_{13} = \omega^*_{23} = 0 \qquad (16)$$

$$\omega^*_{21} = \omega^*_{31} = \omega^*_{32} = 0. \qquad (17)$$

Note that due to symmetry (16) and (17) result in identical equations. These constraints can thus be imposed on the right-hand side, yielding $4(n - 1)$ independent linear equations in b_1, b_2, b_3, b_4 and b_5:

$$P_i^{(1)} \Omega^* P_i^{(1)\top} = P_i^{(2)} \Omega^* P_i^{(2)\top}$$

$$2 P_i^{(1)} \Omega^* P_i^{(2)\top} = 0$$

$$2 P_i^{(1)} \Omega^* P_i^{(3)\top} = 0$$

$$2 P_i^{(2)} \Omega^* P_i^{(3)\top} = 0$$

with $P_i^{(j)}$ representing row j of $\mathbf{P}_i$ and Ω_* parametrized as in (14). The rank 3 constraint can be imposed by taking the closest rank 3 approximation (using SVD for example).

When only two views are available the solution is only determined up to a one parameter family of solutions $\Omega_a + \gamma \Omega_b$. Imposing the rank 3 constraint in this case should be done through the determinant:

$$\det(\Omega_a + \gamma \Omega_b) = 0. \qquad (18)$$

This results in up to 4 possible solutions. The constraint $b_1 > 0$, see Eq. (14), can be used to eliminate some of these solutions. If more than one solution subsists additional constraints should be used. These can come from knowledge about the camera (e.g., constant focal length) or about the scene (e.g., known angle).

4.3. Detecting Critical Motion Sequences

It is outside the scope of this paper to give a complete analysis of all possible critical motions which can occur for self-calibration. For the case where all intrinsic camera parameters are fixed, such an analysis was carried out by Sturm (1997).

Here a more practical approach is taken. Given an image sequence, a method is given to analyze if that particular sequence is suited for self-calibration. The method can deal with all different combinations of constraints. It is based on a sensitivity analysis towards the constraints. An important advantage of the technique is that it also indicates quasi-critical motion sequences. It can be used on a synthetic motion sequence as well as on a real image sequence from which the rigid motion sequence was obtained through self-calibration.

Without loss of generality the calibration matrix can be chosen to be $\mathbf{K} = \mathbf{I}$ (and thus $\omega_i^* = \mathbf{I}$). In the case of real image sequences this implies that the images should first be transformed with $\mathbf{K}^{-1}$. In this case it can be verified that $df_x = \frac{1}{2}d\omega_{i\,11}^*$, $df_y = \frac{1}{2}d\omega_{i\,22}^*$, $du = d\omega_{i\,13}^* = d\omega_{i\,31}^*$, $dv = d\omega_{i\,23}^* = d\omega_{i\,32}^*$ and $ds = d\omega_{i\,12}^* = d\omega_{i\,21}^*$. Now the typical constraints which are used for self-calibration can all be formulated as linear equations in the coefficients of $\mathbf{K}$. As an example of such a system of equations, consider the case $s = 0$, $\frac{f_y}{f_x} = 1$ and $f_x = $ constant. By linearizing around $\omega^* = \mathbf{I}$ this yields $d\omega_{i\,12}^* = 0$, $d\omega_{i\,11}^* = d\omega_{i\,22}^*$, $d\omega_{i\,11}^* = $

$d\omega_{1\,11}^*$. Which can be rewritten as

$$\begin{bmatrix} 0 & 0 & 1 & \cdots & 0 & \cdots \\ 1 & 0 & 0 & \cdots & -1 & \cdots \\ 1 & -1 & 0 & \cdots & 0 & \cdots \end{bmatrix} \begin{bmatrix} d\omega_{1\,11}^* \\ d\omega_{1\,22}^* \\ d\omega_{1\,12}^* \\ \vdots \\ d\omega_{2\,11}^* \\ \vdots \end{bmatrix} = 0. \qquad (19)$$

More in general the linearized self-calibration equations can be written as follows:

$$\mathbf{C}d\omega^* = 0 \qquad (20)$$

with $d\omega^*$ a column vector containing the differentials of the coefficients of the dual image absolute conic ω_i^* for all views. The matrix $\mathbf{C}$ encodes the imposed set of constraints. Since these equations are satisfied for the exact solution, this solution will be an isolated solution of this system of equations if and only if any arbitrary small change to the solution violates at least one of the conditions of Eq. (20). Using (8) a small change can be modeled as follows:

$$\mathbf{C}d\omega^* = \mathbf{C}\left[\frac{d\omega^*}{d\Omega^*}\right]d\Omega^* = \mathbf{C}'d\Omega^* \qquad (21)$$

with $\Omega^* = [\Omega_{11}^* \Omega_{22}^* \Omega_{12}^* \Omega_{31}^* \Omega_{32}^* \Omega_{14}^* \Omega_{24}^* \Omega_{34}^*]^\top$ and the Jacobian $[\frac{d\omega^*}{d\Omega^*}]$ evaluated at the solution. To have the expression of Eq. (21) different from zero for every possible $d\Omega^*$, means that the matrix $\mathbf{C}'$ should be of rank 8 ($\mathbf{C}'$ should have a right null space of dimension 0). In practice this means that all singular values of $\mathbf{C}'$ should significantly differ from zero, else a small change of the absolute quadric proportional to right singular vectors associated with small singular values will almost not violate the self-calibration constraints.

To use this method on results calculated from a real sequence the camera matrices $\mathbf{P}$ should first be adjusted to have the calculated solution become an exact solution of the self-calibration equations.

5. The Metric Reconstruction Algorithm

The proposed self-calibration method is embedded in a system to automatically model metric reconstructions of rigid 3D objects from uncalibrated image sequences. The complete procedure for metric 3D reconstruction
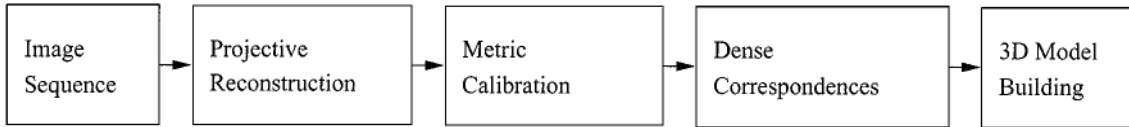
Figure 2. Overview of the different steps in the 3D reconstruction system.

is summarized here. In Fig. 2 the different steps of the 3D reconstruction system are shown.

5.1. Retrieving the Projective Framework

Our approach follows the procedure proposed by Beardsley et al. (1996). The first correspondences are found by extracting points of interest using the Harris corner detector (Harris and Stephens, 1988) in the different images and matching them using a robust tracking algorithm. In conjunction with the matching of the interest points the projective calibration of the setup is calculated in a robust way ( ). This allows to eliminate matches which are inconsistent with the calibration. Using the projective calibration more matches can easily be found and used to refine this calibration. This can be seen in Fig. 3.

At first corresponding corners in two images are matched. This defines a projective framework in which the projection matrices of the other views are retrieved one by one. We therefore obtain projection matrices of the following form:

$$\mathbf{P}_1 = [\mathbf{I} \,|\, 0] \quad \text{and} \quad \mathbf{P}_i = [\mathbf{H}_{1i} \,|\, e_{1i}] \qquad (22)$$

with $\mathbf{H}_{1i}$ the homography for some reference plane from view 1 to view i and e_{1i} the corresponding epipole (i.e., the projection of the first camera position in view i).

5.2. Retrieving the Metric Framework

Such a projective calibration is certainly not satisfactory for the purpose of 3D modeling. A reconstruction obtained up to a projective transformation can differ very much from the original scene according to human
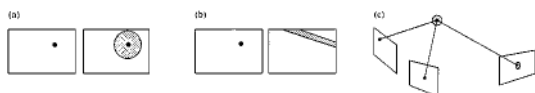

Figure 3. (a) A priori search region, (b) search region based on initial projective geometry, (c) search region after projective reconstruction (used for refinement).

perception; orthogonality and parallelism are in general not preserved, part of the scene can be warped to infinity, etc. Therefore a metric framework should be retrieved. This should be achieved by following the methods described in Section 4. Once the calibration is retrieved it can be used to upgrade the projective reconstruction to a metric one.

5.3. Dense Correspondences

At this point we dispose of a sparse metric reconstruction. Only a restricted number of points are reconstructed. Obtaining a dense reconstruction could be achieved by interpolation, but in practice this does not yield satisfactory results. Often some salient features are missed during the interest point matching and will, therefore, not appear in the reconstruction.

These problems can be avoided by using algorithms which estimate correspondences for almost every point in the images. At this point algorithms can be used which were developed for calibrated 3D systems like stereo rigs. Since we have computed the projective calibration between successive image pairs we can exploit the epipolar constraint that restricts the correspondence search to a 1D search range. In particular it is possible to remap the image pair to standard geometry where the epipolar lines coincide with the image scan lines (Koch, 1996). The correspondence search is then reduced to a matching of the image points along each image scanline. In addition to the epipolar geometry other constraints like preserving the order of neighboring pixels, bidirectional uniqueness of the match, and detection of occlusions can be exploited. These constraints are used to guide the correspondence towards the most probable scanline match using a dynamic programming scheme (Falkenhagen, 1997). The most recent algorithm (Koch et al., 1998) improves the accuracy by using a multi-baseline approach.

5.4. Building the Model

Once a dense correspondence map and the metric camera parameters have been estimated, dense surface

depth maps are computed using depth triangulation. The 3D model surface is constructed as triangular surface patches with the vertices storing the surface geometry and the faces holding the projected image color in texture maps. The texture maps add very much to the visual appearance of the models and augment missing surface detail.

The model building process is at present restricted to partial models computed from single viewpoints and work remains to be done to fuse different viewpoints. Since all the views are registered into one metric framework it is possible to fuse the depth estimate into one consistent model surface (Koch, 1996).

Sometimes it is not possible to obtain a single metric framework for large objects like buildings since one may not be able to record images continuously around it. In that case the different frameworks have to be registered to each other. This will be done using available surface registration schemes (Chen and Medioni, 1991).

6.   Experiments

In this section some experiments are described. First synthetic image sequences were used to assess the quality of the algorithm under simulated circumstances. Both the amount of noise and the length of the sequences were varied. Then results are given for two outdoor video sequences. Both sequences were taken with a standard semi-professional camcorder that was moved freely around the objects. Sequence 1 was filmed with constant camera parameters—like most algorithms require. The new algorithm—which doesn't impose this—could therefore be tested on this. A second sequence was recorded with varying intrinsic parameters. A zoom factor $(2\times)$ was applied while filming.

6.1.   Simulations

The simulations were carried out on sequences of views of a synthetic scene. The scene consisted of 50 points uniformly distributed in a unit sphere with its center at the origin. The intrinsic camera parameters were chosen as follows. The focal length was different for each view, randomly chosen with an expected value of 2.0 and a standard deviation of 0.5. The principal point had an expected value of $(0, 0)$ and a standard deviation of $0.1\sqrt{2}$. In addition the synthetic camera had an aspect

Figure 4.   Example of sequence used for simulations (the views are represented by the optical axis and the image axes of the camera in the different positions).

ratio of one and no skew. The views were taken from all around the sphere and were all more or less pointing towards the origin. An example of such a sequence can be seen in Fig. 4.

The scene points were projected into the images. Gaussian white noise with a known standard deviation was added to these projections. Finally, the self-calibration method proposed in this paper was carried out on the sequence. For the different algorithms the metric error was computed. This is the mean deviation between the scene points and their reconstruction after alignment. The scene and its reconstruction are aligned by applying the metric transformation which minimizes the difference between both. For comparison the same error was also calculated after alignment with a projective transformation. By default the noise had an equivalent standard deviation of 1.0 pixel for a 500×500 image. To obtain significant results every experiment was carried out 10 times and the mean was calculated.

To analyze the influence of noise on the algorithms, noise values of 0, 0.1, 0.2, 0.5, 1, 1.5 and 2 pixels were used on sequences of 6 views. The results can be seen in Fig. 5. It can be seen that for small amounts of noise the more complex models should be preferred. If more noise is added, the simple model gives the best results. This is due to the low redundancy of the system of equations for the models which, beside the focal length, also try to estimate the position of the principal point.

Another experiment was carried out to evaluate the performance of the algorithm for different sequence lengths. Sequences ranging from 4 to 40 views were used. A noise level of one pixel was used. The results are shown in Fig. 6. For short image sequences the results are better when the principal point is assumed in the middle of the image, even though this is not exactly true. For longer image sequences the constraints
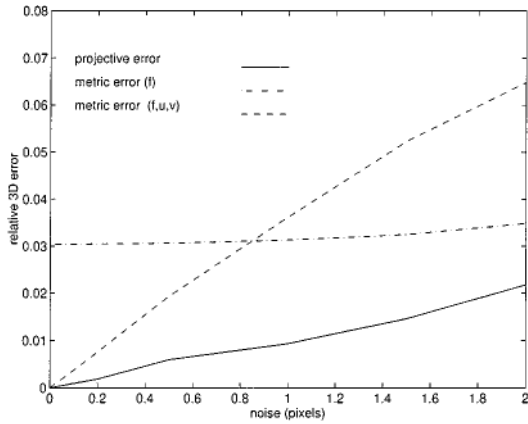
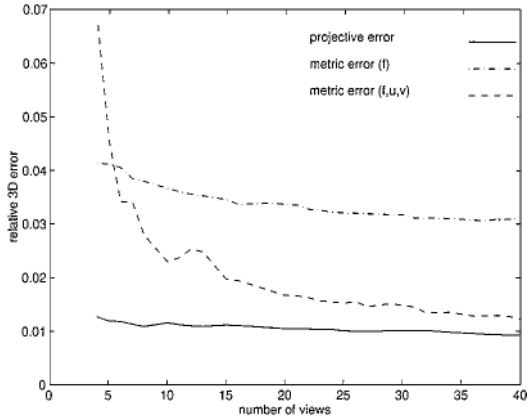*Figure 5.*   Relative 3D error in function of noise.

Figure 6.   Relative 3D error for sequences of different lengths.

on the aspect ratio and the image skew are sufficient to allow an accurate estimation of the metric structure of the scene. In this case fixing the principal point will degrade the results by introducing a bias.

6.2.   Real Sequence 1

The first sequence showing part of an old castle was filmed with a fixed zoom/focus. It is therefore a good test for the algorithms presented in this paper to check if they indeed return constant intrinsic parameters for this sequence. In Fig. 7 some of the images of the sequence are shown. Figure 8 shows the reconstruction together with the estimated viewpoints of the camera. In Fig. 9 another view is shown, both with texture and with shading. The shaded view shows that even small geometrical details (e.g., window indentations) were recovered in the reconstruction. To judge the visual quality of the reconstruction, different perspective views of the model were computed and displayed in Fig. 9. The resulting reconstruction is visually convincing and preserve the metric properties of the original scenes (i.e., parallelism, orthogonality, ...).

A quantitative assessment of these properties can be made by explicitly measuring angles directly on the object surface. For this experiment six lines were placed along prominent surface features, three on each object plane, aligned with the windows. The three lines inside of each object plane should be parallel to each other (angle between them should be $0°$), while the lines of different object planes should be perpendicular to each other (angle between them should be $90°$). The measurement on the object surface shows that this is indeed close to the expected values (see Table 2).

In Fig. 10 the focal length for every view is plotted for the different algorithms and different sets of

Table 2.   Results of metric measurements on the reconstruction angle ($\pm$std. dev.).

Parallelism	1.0 ± 0.6 degrees
Orthogonality	92.5 ± 0.4 degrees

Figure 7.   Some of the images of the Arenberg castle which were used for the reconstruction.
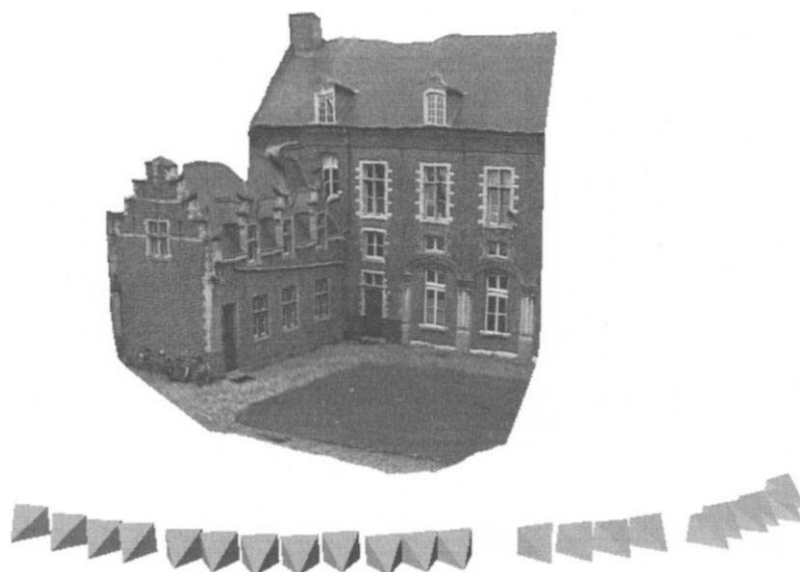
Figure 8.   Perspective view of the reconstruction together with the estimated position of the camera for the different views of the sequence.
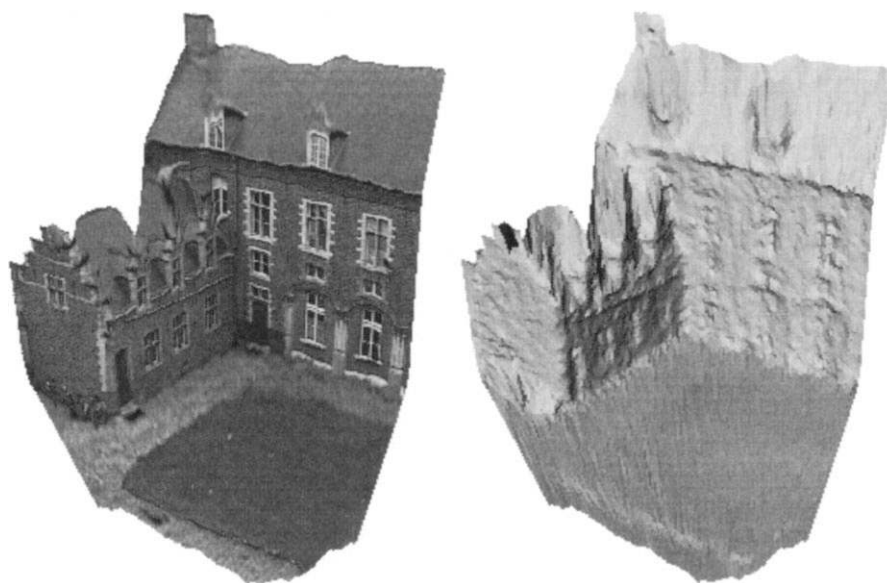

Figure 9.   Two other perspective views of the Arenberg castle reconstruction.

constraints. The calculated focal lengths are almost constant as it should be. In one case also the principal point was estimated (independently for every view), but the results were not so good. Not only did the principal point move a lot (over 100 pixels), but in this case the estimate of the focal length is not as constant anymore. In this case it seems the projective calibration was not accurate enough to allow an accurate

retrieval of the principal point and it could be better to stick with the simplified algorithm. In general it seems that it is hard to accurately determine the absolute value of the focal length, especially when not much perspective distortion is present in the images. This explains why the different algorithms can result in different values for the focal length. On the other hand an inaccurate estimation of the focal length only
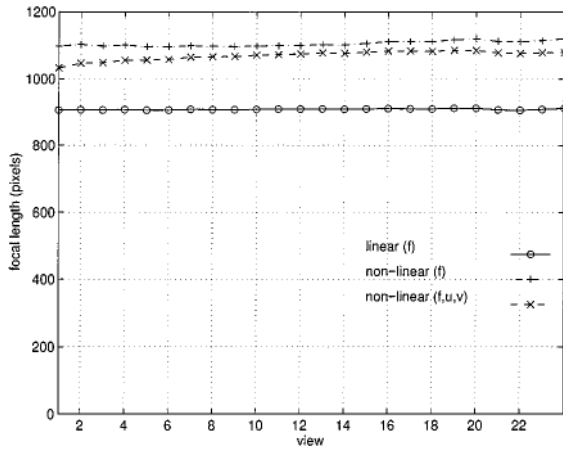
Figure 10. Focal length (in pixels) versus views for the different algorithms.

has a small effect on the reconstruction (Bougnoux, 1998).

6.3. Real Sequence 2

This sequence shows a stone pillar with curved surfaces. While filming and moving away, the zoom was changed to keep the image size of the object constant. The focal length was not changed between the two first images, then it was changed more or less linearly. From the second image to the last image, the focal length has been doubled (if the markings on the camera can be trusted). In Fig. 11, 3 of the 8 images of the sequence can be seen. Notice that the perspective distortion is most visible in the first images (wide angle) and diminishes towards the end of the sequence (longer focal length).

Figure 12 shows a top view of the reconstructed pillar together with the estimated camera viewpoints. These viewpoints are illustrated with small pyramids. Their height is proportional to the focal length. In Fig. 13 perspective views of the reconstruction are given. The view on top is rendered both shaded and with surface texture mapped. The shaded view shows that even most of the small details of the object are retrieved. The bottom part shows a left and a right side view of the reconstructed object. Although there is some distortion at the outer boundary of the object, a highly realistic impression of the object is created. Note the arbitrarily shaped free-form surface that has been reconstructed.

Figure 11. Images 1, 4 and 8 of *Sequence* 2 (note the short focal length/wide angle in the first image and the long focal length in the last image).

Figure 12. Top view of the reconstructed pillar together with the different viewpoints of the camera (note the change in focal length).

Figure 13. Perspective views of the reconstruction (with texture and with shading).

A quantitative assessment of the metric properties for the pillar is not so easy because of the curved surfaces. It is, however, possible to measure some distances on the real object as reference lengths and compare them with the reconstructed model. In this case it is possible to obtain a measure for the absolute scale and verify the consistency of the reconstructed lengths within the model. For this comparison a network of reference lines was placed on the original object and 27 manually measured object distances were compared with the reconstructed distances on the model surface, as seen in Fig. 14. From each comparison the absolute object scale factor was computed. The results are found in Table 3.

Due to the increased reconstruction uncertainty at the outer object silhouette some distances show a larger error than the interior points. This accounts for the
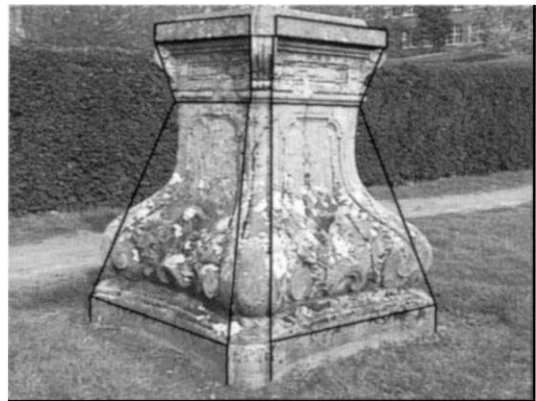

Figure 14. To allow for a quantitative comparison between the real pillar and its reconstruction, some distances, superimposed in black, were measured.

outliers. Averaging all 27 measured distances gave a consistant scale factor of 40.25 with a standard deviation of 5.4% overall. For the interior distances, the reconstruction error dropped to 2.3%. These results demonstrate the metric quality of the reconstruction even for complicated surface shapes and varying focal length. In Fig. 15 the focal length for every view
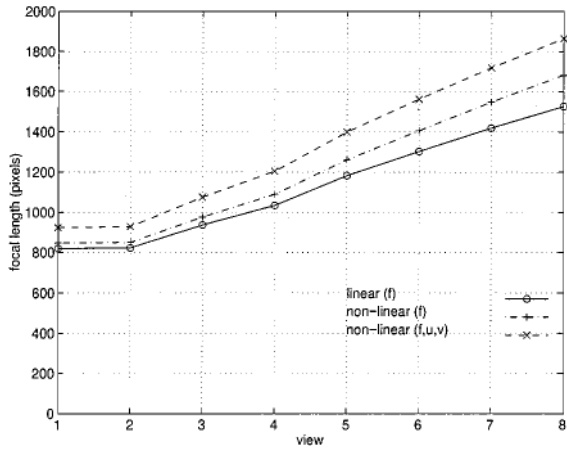
Table 3. Results of metric measurements on the reconstruction ratio (±std. dev.).

All points	40.25 ± 2.2
Interior points	± 0.9

Figure 15. Focal length (in pixels) versus views for the different algorithms.

is plotted for the different algorithms. It can be seen that the calculated values of the focal length correspond to what could be expected. When the principal point was estimated independently for every view, it moved around up to 50 pixels. It is probable that too much noise is present to allow us to estimate the principal point accurately.

7.   Some More Results

In this section some more results are presented which illustrate the flexibility of our reconstruction method. The two first examples were recorded at Sagalassos, an archaeological site in Turkey. The last sequence consists of images of a Jain temple in Ranakpur, India. These images were taken during a tour around India after ICCV'98.

7.1.   The Archeological Site of Sagalassos

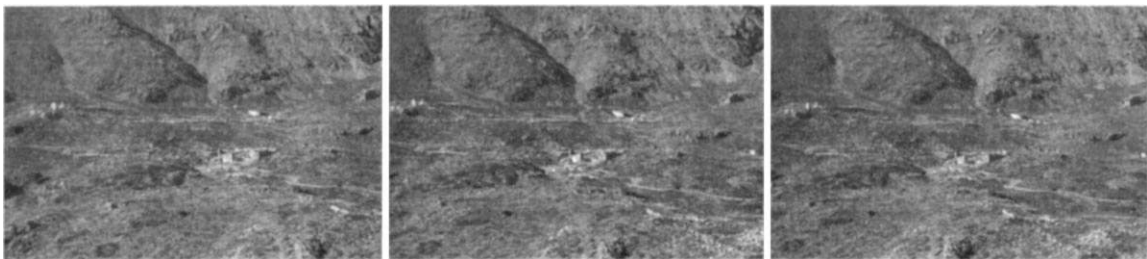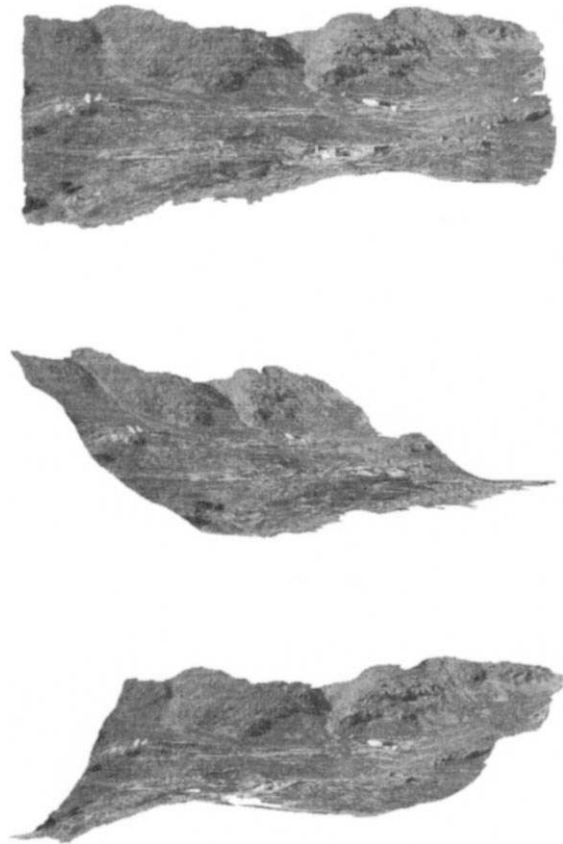In Fig. 16 images of the Sagalassos site sequence (10 images) are shown. They show the landscape

Figure 17. Perspective views of the 3D reconstruction of the Sagalassos site.

surrounding the Sagalassos site. Some views of the reconstruction are shown in Fig. 17. With our technique this model was obtained just as easily as the previous ones. For most active techniques it is impossible to cope with scenes of this size. The use of a stereo rig would also be very hard since a baseline of several tens of meters would be required. Therefore one of the most promising applications of the proposed technique is large scale terrain modeling. In addition one can
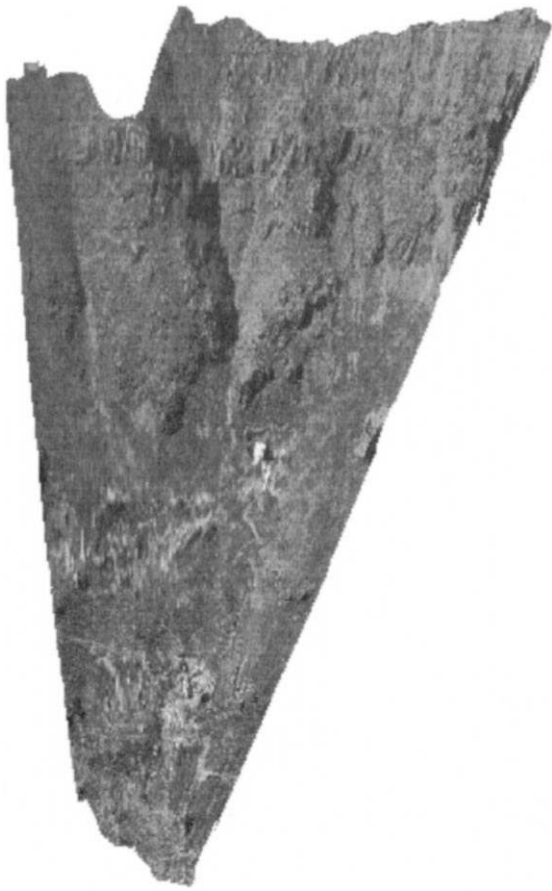

Figure 16. Some of the images of the Sagalassos site sequence.

Figure 18. Top view of the reconstruction of the Sagalassos site.

see from Fig. 18 that this model could also be used to obtain a Digital Terrain Map or an orthomap at low cost. In this case only 3 reference measurements— GPS and altitude—are necessary to localize and orient the model in the world reference frame.

7.2. The Fountain of Sagalassos

Besides the whole site, several monuments were reconstructed separately. As an example, the reconstruction of the remains of an ancient fountain is shown. In Fig. 19 three of the six images used for the reconstruction are shown. All images were taken from the same ground level. They were acquired with a digital camera with a resolution of approximately 1500 × 1000. Half resolution images were used for the computation of the shape. The texture was generated from the full resolution images. The reconstruction can be seen in Fig. 20, the left side shows a view with texture, the right view gives a shaded view of the model without texture. In Fig. 21 two close-up shots of the model are shown.

7.3. The Jain Temple of Ranakpur

These images were taken during a tourist trip after ICCV'98 in India. A sequence of 11 images was taken of some details of one of the smaller Jain temples at Ranakpur, India. These images were taken with a
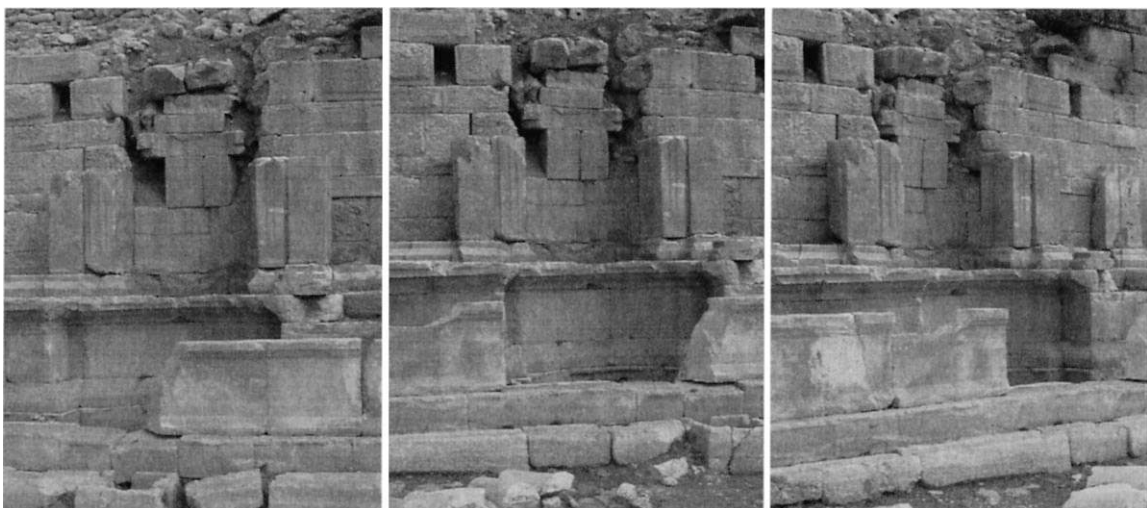

Figure 19. Three of the six images of the fountain sequence.

Figure 20.   Perspective views of the reconstructed fountain with and without texture.
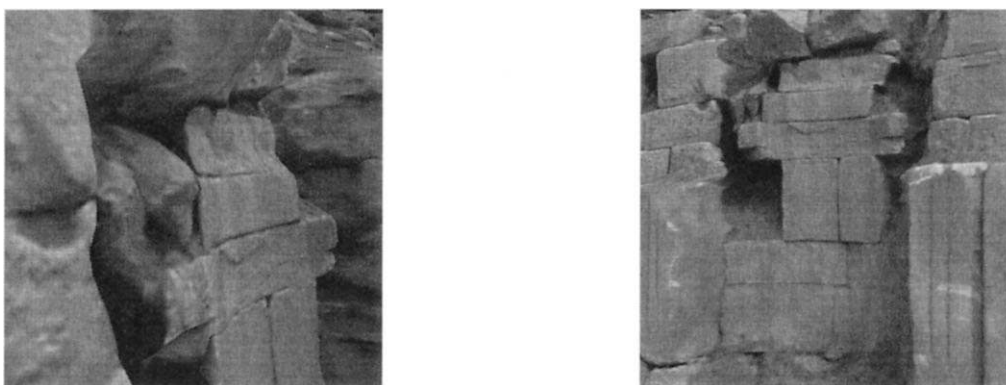

Figure 21.   Close-up views of some details of the reconstructed fountain.

Figure 22.   Three images of a detail of a Jain temple of Ranakpur.

standard Nikon F50 photo camera and then scanned in. Three of them can be seen in Fig. 22. A view of the reconstruction which was obtained from this sequence can be seen in Fig. 23, some details can be seen in Fig. 24. Figure 25 is an orthographic view taken from below the reconstruction. This view allows to verify the orthogonality of the reconstruction.

These reconstructions show that we are able to handle even complex 3D geometries affectively with our reconstruction system.
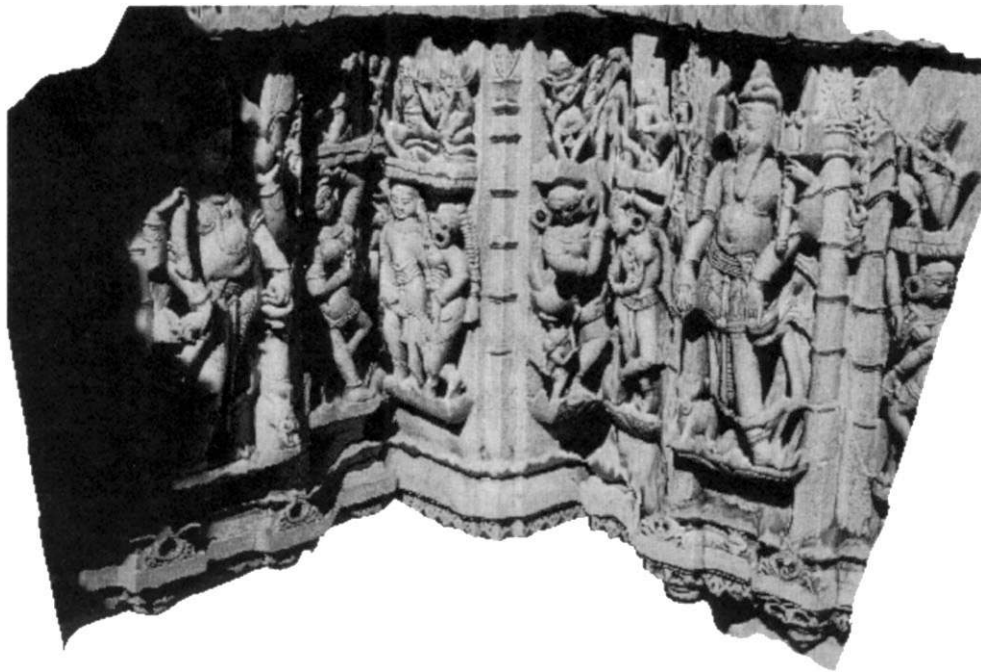
Figure 23.   A perspective view of the reconstruction.

Figure 24.   Some close-ups of the reconstruction.

Figure 25.   Orthographic view from below the reconstruction.

8.   Conclusions

This paper focussed on self-calibration and metric reconstruction in the presence of varying and unknown intrinsic camera parameters. The calibration models used in previous research are on the one hand too restrictive in real imaging situations (constant parameters) and on the other hand too general (all parameters unknown). The more pragmatic approach which is followed in this paper results in more flexibility.

A counting argument was derived which gives the minimum number of views needed for self-calibration depending on which constraints are used. We proved that self-calibration is possible using only the most general constraint (i.e., that image rows and columns are orthogonal). Of course if more constraints are available, this will in general yield better results.

A versatile self-calibration method which can work with different types of constraints (some of the intrinsic camera parameters constant or known) was derived. This method was then specialized towards the practically important case of a zooming/focusing camera (without skew and an aspect ratio $\frac{f_y}{f_x} = 1$). Both known and unknown principal points were considered. It is proposed to always start with the principal point in the center of the image and to first use the linear algorithm. The non-linear minimization is then used to refine the

results, possibly—for longer sequences—allowing the principal point to be different for each image. This can however degrade the results if the projective calibration was not accurate enough, the sequence not long enough, or the motion sequence critical towards the set of constraints. As for all self-calibration algorithms it is important to deal with critical motion sequences. In this paper a general method is proposed which detects critical and quasi-critical motion sequences.

The different methods are validated by experiments which are carried out on synthetic as well as real image sequences. The former are used to analyze noise sensitivity and influence of the length of the sequence. The latter show the practical feasibility of the approach. Some more results were included to demonstrate the flexibility of the approach and the visual quality of the results which can be obtained by incorporating this technique in a 3D reconstruction system.

In the future several problems will be investigated more in depth. Some work is planned on attaching a weight to different constraints. For example, the skew can be very accurately assumed to be zero, whereas the principal point is only known to lay somewhere around the center of the image. Also the critical motion sequence detection should be incorporated in the algorithm and be used to predict the accuracy of the results.

Appendix

In this appendix the proof of Theorem 1 is given. Before starting the actual proof a lemma will be given. This lemma gives a way to check for the absence of skew from the coefficients of $\mathbf{P}$ directly without needing the factorization. A camera projection matrix can be factorized as follows $\mathbf{P} = [\mathbf{H} \mid \mathbf{e}] = \mathbf{K}[\mathbf{R} \mid -\mathbf{Rt}]$. In what follows $\mathbf{h}_i$ and $\mathbf{r}_i$ denote the rows of $\mathbf{H}$ and $\mathbf{R}$.

Lemma 1. *The absence of skew is equivalent with* $(\mathbf{h}_1 \times \mathbf{h}_3)(\mathbf{h}_2 \times \mathbf{h}_3) = 0$.

Proof: It is always possible to factorize $\mathbf{H}$ as $\mathbf{KR}$. Therefore the following can be written:

$$(\mathbf{h}_1 \times \mathbf{h}_3)(\mathbf{h}_2 \times \mathbf{h}_3)$$
$$= ((f_x\mathbf{r}_1 + s\mathbf{r}_2 + u\mathbf{r}_3) \times \mathbf{r}_3)((f_y\mathbf{r}_2 + v\mathbf{r}_3) \times \mathbf{r}_3)$$
$$= ((f_x\mathbf{r}_1 + s\mathbf{r}_2) \times \mathbf{r}_3)(f_y\mathbf{r}_2 \times \mathbf{r}_3)$$
$$= -f_x f_y \mathbf{r}_2\mathbf{r}_1 + s f_y \mathbf{r}_1\mathbf{r}_1 = s f_y.$$

Because $f_y \neq 0$ this concludes the proof.   □

Equipped with this lemma the following theorem can be proven.

Theorem 1.   *The class of transformations which preserves the absence of skew is the group of similarity transformations.*

Proof: It is easy to show that the similarity transformations preserve the calibration matrix $\mathbf{K}$ and hence also the orthogonality of the image plane:

$$\mathbf{K}[\mathbf{R} \mid -\mathbf{Rt}]\begin{bmatrix} \mathbf{R}' & \sigma^{-1}\mathbf{t}' \\ 0 & \sigma^{-1} \end{bmatrix} = \mathbf{K}[\mathbf{RR}' \mid \sigma^{-1}(\mathbf{Rt}' - \mathbf{Rt})].$$

Therefore it is now sufficient to prove that the class of transformations which preserve the condition $(\mathbf{h}_1 \times \mathbf{h}_3)(\mathbf{h}_2 \times \mathbf{h}_3) = 0$ is at most the group of similarity transformations. To do this a specific set of positions and orientations of cameras can be chosen, since the absence of skew is supposed to be preserved for *all possible* views. In general $\mathbf{P}$ can be transformed as follows:

$$\mathbf{P}' = [\mathbf{H} \mid \mathbf{e}]\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{c}^\top & d \end{bmatrix} = [\mathbf{HA} + \mathbf{ec}^\top \mid \mathbf{Hb} + \mathbf{e}d]$$

If $\mathbf{t} = 0$ then $\mathbf{H}' = \mathbf{KRA}$ and thus

$$(\mathbf{h}_1' \times \mathbf{h}_3')(\mathbf{h}_2' \times \mathbf{h}_3')$$
$$= ((f_x\mathbf{r}_1 + u\mathbf{r}_3)\mathbf{A} \times \mathbf{r}_3\mathbf{A})((f_y\mathbf{r}_2 + v\mathbf{r}_3)\mathbf{A} \times \mathbf{r}_3\mathbf{A}).$$

Therefore the condition of the lemma is equivalent with

$$(\mathbf{r}_1\mathbf{A} \times \mathbf{r}_3\mathbf{A})(\mathbf{r}_2\mathbf{A} \times \mathbf{r}_3\mathbf{A}) = 0.$$

Choosing the rotation matrices $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{R}_3$ rotations of $90°$ around the x-, y- and z-axis, imposes the following equations to hold:

$$(\mathbf{a}_1 \times \mathbf{a}_2)(\mathbf{a}_3 \times \mathbf{a}_2) = 0 \,,$$
$$(\mathbf{a}_3 \times \mathbf{a}_1)(\mathbf{a}_2 \times \mathbf{a}_1) = 0 \,, \qquad \text{(A1)}$$
$$(\mathbf{a}_2 \times \mathbf{a}_3)(\mathbf{a}_1 \times \mathbf{a}_3) = 0 \,.$$

Hence $(\mathbf{a}_1 \times \mathbf{a}_2)$, $(\mathbf{a}_1 \times \mathbf{a}_3)$ and $(\mathbf{a}_2 \times \mathbf{a}_3)$ define a set of 3 mutually orthogonal planes where $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ form the intersection and are therefore also orthogonal.

Choosing $\mathbf{R}_4$ and $\mathbf{R}_5$ as $\mathbf{R}_1$ and $\mathbf{R}_2$ followed by a rotation of $45°$ around the z-axis, the following two

equations can be derived:

$$((\mathbf{a}_1 + \mathbf{a}_3) \times \mathbf{a}_2)\,((\mathbf{a}_1 - \mathbf{a}_3) \times \mathbf{a}_2) = 0$$
$$((\mathbf{a}_3 + \mathbf{a}_2) \times \mathbf{a}_1)\,((\mathbf{a}_3 - \mathbf{a}_2) \times \mathbf{a}_1) = 0. \tag{A2}$$

Carrying out some algebraic manipulations and using $\mathbf{a}_1 \perp \mathbf{a}_2 \perp \mathbf{a}_3$ this yields the following result:

$$|\mathbf{a}_1|^2 = |\mathbf{a}_2|^2 = |\mathbf{a}_3|^2.$$

These results mean that $\mathbf{A} = \sigma \mathbf{R}$ with σ a scalar and $\mathbf{R}$ an orthonormal matrix. The available constraints are not sufficient to impose $\det \mathbf{R} = 1$, therefore mirroring is possible.

Choose $\mathbf{R}_6 = \mathbf{R}_1$ and $\mathbf{t}_6^\top = [1\,0\,0]$, then $((\mathbf{a}_1 + \mathbf{c}^\top) \times \mathbf{a}_2)(\mathbf{a}_3 \times \mathbf{a}_2) = 0$ must hold. Using (A1) and $\mathbf{a}_2 \times \mathbf{a}_3 \propto \mathbf{a}_1$ this condition is equivalent with $(\mathbf{c} \times \mathbf{a}_2)\mathbf{a}_1 = 0$. Writing $\mathbf{c}$ as $c_1 \mathbf{a_1} + c_2 \mathbf{a_2} + c_3 \mathbf{a_3}$ this boils down to $c_3 = 0$. Taking $\mathbf{R}_7 = \mathbf{R}_2$, $\mathbf{t}_7 = [0\,0\,1]^\top$, $\mathbf{R}_8 = \mathbf{R}_3$ and $\mathbf{t}_8 = [0\,1\,0]^\top$ leads in a similar way to $c_2 = 0$ and $c_1 = 0$ and therefore to $\mathbf{c}^\top = [0\,0\,0]$.

In conclusion the transformation $\left[\begin{smallmatrix}\mathbf{A} & \mathbf{b} \\ \mathbf{c}^\top & d\end{smallmatrix}\right]$ is restricted to the following form $\left[\begin{smallmatrix}\sigma\mathbf{R} & \mathbf{t} \\ 0 & 1\end{smallmatrix}\right]$ which concludes the proof.

$\square$

Remark that 8 views were needed in this proof. This is consistent with the counting argument of the previous paragraph.

Acknowledgments

References

Armstrong, M., Zisserman, A., and Beardsley, P. 1994. Euclidean structure from uncalibrated images. In *Proc. British Machine Vision Conference*, pp. 509–518.

Beardsley, P., Torr, P., and Zisserman, A. 1996. 3D model acquisition from extended image sequences. In *Proc. European Conference on Computer Vision*, Cambridge, UK, vol. 2, pp. 683–695.

Bougnoux, S. 1998. From projective to Euclidean space under any practical situation, a criticism of self-calibration. In *Proc. International Conference on Computer Vision*, Bombay, India, pp. 790–796.

Chen, Y. and Medioni, G. 1991. Object modeling by registration of multiple range images. In *Proc. Int. Conf. on Robotics and Automation*.

Falkenhagen, L. 1997. Hierarchical block-based disparity estimation considering neighbourhood constraints. In *Proc. International Workshop on SNHC and 3D Imaging*, Rhodes, Greece.

Faugeras, O. 1992. What can be seen in three dimensions with an uncalibrated stereo rig. In *Proc. European Conference on Computer Vision*, pp. 563–578.

Faugeras, O., Luong, Q.-T., and Maybank, S. 1992. Camera self-calibration: Theory and experiments. In *Proc. European Conference on Computer Vision*, pp. 321–334.

Harris, C. and Stephens, M. 1988. A combined corner and edge detector. In *Fourth Alvey Vision Conference*, pp. 1447–151.

Hartley, R. 1992. Estimation of relative camera positions for uncalibrated cameras. In *Proc. European Conference on Computer Vision*, pp. 579–587.

Hartley, R. 1994. Euclidean reconstruction from uncalibrated views. *Applications of Invariance in Computer Vision*, LNCS 825, Springer-Verlag, pp. 237–256.

Heyden, A. and Åström, K. 1996. Euclidean reconstruction from constant intrinsic parameters. In *Proc. International Conference on Pattern Recognition*, Vienna, Austria, pp. 339–343.

Heyden, A. and Åström, K. 1997. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Proc. International Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 438–443.

Karl, W., Verghese, G., and Willsky, A. 1994. Reconstructing ellipsoids from projections. *CVGIP; Graphical Models and Image Processing*, 56(2): 124–139.

Koch, R. 1996. Automatische oberflachenmodellierung starrer dreidimensionaler objekte aus stereoskopischen rundum-ansichten. Ph.D. Thesis, University of Hannover, Germany.

Koch, R., Pollefeys, M., and Van Gool, L. 1998. Multi viewpoint stereo from uncalibrated video sequences. In *Proc. European Conference on Computer Vision*, Freiburg, Germany.

Laveau, S. and Faugeras, O. 1996. Oriented projective geometry for computer vision. In *Proc. European Conference on Computer Vision*, Cambridge, UK, vol. 1, pp. 147–156.

Luong, Q.-T. and Faugeras, O. 1997. Self calibration of a moving camera from point correspondences and fundamental matrices. *Internation Journal of Computer Vision*, vol. 22-3.

Pollefeys, M. and Van Gool, L. 1997a. A stratified approach to self-calibration. In *Proc. International Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 407–412.

Pollefeys, M. and Van Gool, L. 1997b. Self-calibration from the absolute conic on the plane at infinity. In *Proc. International Conference on Computer Analysis of Images and Patterns*, Kiel, Germany, pp. 175–182.

Pollefeys, M., Van Gool, L., and Proesmans, M. 1996. Euclidean 3D reconstruction from image sequences with variable focal lengths. In *Proc. European Conference on Computer Vision*, Cambridge, UK, vol. 1, pp. 31–42.

Semple, J.G. and Kneebone, G.T. 1952. *Algebraic Projective Geometry*. Oxford University Press.

Sturm, P. 1997. Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction. In *Proc. International Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 1100–1105.

Torr, P. 1995. Motion segmentation and outlier detection. Ph.D. Thesis, Dept. of Engineering Science, University of Oxford.

Triggs, B. 1997. The absolute quadric. In *Proc. International Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 609–614.

Zeller, C. and Faugeras, O. 1996. Camera self-calibration from video sequences: The Kruppa equations revisited. INRIA, Sophia-Antipolis, France, Research Report 2793.