Katholieke
Universiteit
Leuven

Departement Elektrotechniek
Afdeling ESAT/MI2
Kardinaal Mercierlaan 94
B-3001 Heverlee - Belgium

TECHNISCH RAPPORT - TECHNICAL REPORT

Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters

Marc Pollefeys, Reinhard Koch and Luc Van Gool

August 1997

Nr. KUL/ESAT/MI2/9707

Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters

Marc Pollefeys, Reinhard Koch and Luc Van Gool
ESAT-MI2, K.U.Leuven
Kardinaal Mercierlaan 94
B-3001 Heverlee (Leuven)
Belgium
firstname.lastname@esat.kuleuven.ac.be

Abstract

In this report the theoretical and practical feasibility of self-calibration in the presence of varying internal camera parameters is under investigation. A theoretical proof will be given which shows that the absence of skew in the image plane is sufficient to allow for self-calibration. Besides this a self-calibration method is presented which efficiently deals with all kinds of constraints on the internal camera parameters and which can detect critical motion sequences. 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.

Keywords: **Motion and Stereo: (Self-)Calibration, (Uncalibrated) 3D Reconstruction, Multiple-View Geometry**

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 to allow self-calibration 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 (equal to one). 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 [3, 5], researchers tried to find ways to upgrade these reconstructions to metric (i.e. Euclidean but unknown scale). Many methods where developed which assumed constant internal camera parameters. Most of these methods are based on the **absolute conic** which is the only conic which stays fixed under all Euclidean transformations. This conic lays in the plane at infinity and its image is directly related to the internal camera parameters, hence the advantage for self-calibration [16].

Faugeras *et al* [4] (see also [15]) 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 [19] proposed a more robust version of this method.

Heyden and Åström [7], Triggs [18] and Pollefeys and Van Gool [14] 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 internal camera parameters.

Pollefeys *et al* [12, 13] also proposed a stratified approach which consists of first locating the plane at infinity using the modulus constraint (i.e. for constant internal camera parameters the infinity homography should be conjugated to a rotation matrix) and then calculating the absolute conic. Hartley [6] proposed another approach based on the minimization of the difference between the internal camera parameters for the different views.

So far not much work has been done on varying internal camera parameters. Pollefeys *et al* [11] also proposed a stratified approach for the case of a varying focal length, but this method required a pure translation as initialization. Recently Heyden and Åström [8] 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 of limited practical use since it is based on bundle adjustment which requires non-linear minimization over all reconstructed points and cameras simultaneously and they don't deal with the issue of initialization.

In this report 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 were the focal length varies, possibly also the principal point.

Section 2 of this report first introduces some basic principles, then gives a counting ar-

gument 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 3 the actual method is developed. A simplified linear version is also given which can be used for initialization. Section 4 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 5. Section 6 concludes this report and gives some directions for further research.

2 Some theory...

The projection of a scene onto an image can be modeled by the following equation:

$$\lambda m = \mathbf{P} M \tag{1}$$

where $m = [xy1]^\top$ is an image point and $M = [XYZ1]^\top$ is a scene point, $\mathbf{P}$ is the 3×4 camera projection matrix and λ is a scale factor. The camera projection matrix factorizes as follows:

$$\mathbf{P} = \mathbf{K}\,[\mathbf{R}|\text{-}\mathbf{R}\mathbf{t}] \;\text{ with }\; \mathbf{K} = \begin{bmatrix} f_x & s & u \\ & f_y & v \\ & & 1 \end{bmatrix} . \tag{2}$$

Here $(\mathbf{R}, \mathbf{t})$ denotes a rigid transformation (i.e. $\mathbf{R}$ is a rotation matrix and $\mathbf{t}$ is a translation vector), while the upper triangular calibration matrix $\mathbf{K}$ encodes the intrinsic parameters of the camera (i.e. f_x and f_y represent the focal length divided by the pixel width resp. height, (u, v) represents the principal point and s is a factor which is zero in the absence of skew).

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 [3, 5]. Note that without additional constraints nothing more can be achieved. This can be seen from the following equation. $\lambda m = \mathbf{P}M = (\mathbf{P}\mathbf{T}^{-1})(\mathbf{T}M) = \mathbf{P}'M'$ 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 equation 2.

It will be shown that when no skew is present (i.e. $s = 0$ in equation 2), the ambiguity of the reconstruction can be restricted to metric. Although this is theoretically sufficient, under practical circumstances often much more constraints are available and should be used.

2.1 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.

constraints	known	fixed	min #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, $\frac{f_y}{f_x}$, u, v		2
standard self-calibration problem		f_x, f_y, u, v, s	3

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

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 internal camera parameter for n views gives n constraints, fixing one yields only $n-1$ constraints.

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

Of course this counting argument is only valid for non-critical motion sequences (see section 3.3).

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 2.2 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 4 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.

2.2 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 [8] which besides orthogonality also requires the aspect ratio to be known.

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 $\mathbf{P}$ immediately without needing the factorization. A camera projection matrix can be factorized as follows $\mathbf{P} = [\mathbf{H}|\mathbf{e}] = \mathbf{K}[\mathbf{R}|-\mathbf{Rt}]$. In what follows $\mathbf{h}_i$ and $\mathbf{r}_i$ denote the rows of $\mathbf{H}$ and $\mathbf{R}$.

Lemma

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

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

Equipped with this lemma the following theorem can be proven.

Theorem

The class of transformations which preserves the absence of skew is the group of similarity transformations (+mirroring).

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:

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

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:

$$\mathbf{P}' = [\mathbf{H}|\mathbf{e}]\begin{bmatrix}\mathbf{A} & \mathbf{b} \\ \mathbf{c}^\top & d\end{bmatrix} = \left[\mathbf{H}\mathbf{A}+\mathbf{e}\mathbf{c}^\top|\mathbf{H}\mathbf{b}+\mathbf{e}d\right]$$

If $\mathbf{t} = 0$ then $\mathbf{H}' = \mathbf{K}\mathbf{R}\mathbf{A}$ 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^o around the x-, y- and z-axis, imposes the following equations to hold:

$$\begin{aligned}(\mathbf{a}_1 \times \mathbf{a}_2)(\mathbf{a}_3 \times \mathbf{a}_2) &= 0\,, \\ (\mathbf{a}_3 \times \mathbf{a}_1)(\mathbf{a}_2 \times \mathbf{a}_1) &= 0\,, \\ (\mathbf{a}_2 \times \mathbf{a}_3)(\mathbf{a}_1 \times \mathbf{a}_3) &= 0\,.\end{aligned} \tag{3}$$

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^o around the z-axis, the following two equations can be derived:

$$\begin{aligned}((\mathbf{a}_1 + \mathbf{a}_3) \times \mathbf{a}_2)\,((\mathbf{a}_1 - \mathbf{a}_3) \times \mathbf{a}_2) &= 0 \\ ((\mathbf{a}_3 + \mathbf{a}_2) \times \mathbf{a}_1)\,((\mathbf{a}_3 - \mathbf{a}_2) \times \mathbf{a}_1) &= 0 \ .\end{aligned} \tag{4}$$

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 $\left((\mathbf{a}_1 + \mathbf{c}^\top) \times \mathbf{a}_2\right)(\mathbf{a}_3 \times \mathbf{a}_2) = 0$ must hold. Using Equation 3 and $\mathbf{a}_2 \times \mathbf{a}_3 \simeq \mathbf{a}_1$ 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 $\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{c}^\top & d \end{bmatrix}$ is restricted to the following form $\begin{bmatrix} \sigma\mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}$

which concludes the proof. □

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

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 (and possibly a mirroring) would satisfy the orthogonality of rows and columns for all views. Using oriented projective geometry [10] 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.

3 Self-Calibration Methods

A practical way to obtain the calibration parameters from constraints on the internal camera parameters is through the absolute quadric [16] (introduced in computer vision by Triggs [18], see also [7, 14]). In space one quadric (of planes) exists which has the property to be invariant under rigid transformations. This quadric consists of planes tangent to the absolute conic. In a metric frame it is represented by a 4×4 symmetric rank 3 matrix $\Omega = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & 0 \end{bmatrix}$. If $\mathbf{T}$ transforms points $M \to \mathbf{T}M$ (and thus $\mathbf{P} \to \mathbf{P}\mathbf{T}^{-1}$), then it transforms $\Omega \to \mathbf{T}\Omega\mathbf{T}^\top$ (which can be verified to yield Ω when $\mathbf{T}$ is a similarity transformation). The projection of the absolute quadric in the image yields the dual image absolute conic:

$$\omega_k \simeq \mathbf{P}_k \Omega \mathbf{P}_k^\top \simeq \mathbf{K}_k \mathbf{K}_k^\top \tag{5}$$

independent of the chosen projective basis[1]. Therefore constraints on the internal 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 the metric frame (which brings Ω to its canonical form).

[1]Using Equation 2 this can be verified for a metric basis. Transforming $\mathbf{P} \to \mathbf{P}\mathbf{T}^{-1}$ and $\Omega \to \mathbf{T}\Omega\mathbf{T}^\top$ will not change the projection.

3.1 Non-linear Approach

Equation 5 can be used to obtain the metric calibration from the projective one. The dual image absolute conics ω_k should be parameterized in such a way that they enforce the constraints on the calibration parameters. For the absolute quadric $\mathbf{\Omega}$ 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.

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

$$\min \sum_{k=1}^{n} \left\| \mathbf{K}_k \mathbf{K}_k^\top - \mathbf{P}_k \Omega \mathbf{P}_k^\top \right\|_F^2 \quad . \tag{6}$$

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

If one choose $\mathbf{P}_1 = [\mathbf{I}|\mathbf{0}]$, Equation 5 can be rewritten as follows:

$$\mathbf{K}_k \mathbf{K}_k^\top \simeq \mathbf{P}_k \begin{bmatrix} \mathbf{K}_1\mathbf{K}_1^\top & \mathbf{a} \\ \mathbf{a}^\top & \mathbf{a}^\top\mathbf{a} \end{bmatrix} \mathbf{P}_k^\top \tag{7}$$

with $\mathbf{a} = [a_1\, a_2\, a_3]^\top$ a vector which encodes the position of the plane at infinity. 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[2]. Therefore it is proposed to first use the simplified version of equation 7 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[3], it can also be used that the principal point will be close to the center of the image. This leads to the following parameterizations for $\mathbf{K}_k$ (transform the images to have $(0,0)$ in the middle):

$$\mathbf{K}_k = \begin{bmatrix} f & 0 & u \\ & f & v \\ & & 1 \end{bmatrix} \text{ or } \mathbf{K}_k = \begin{bmatrix} f & 0 & 0 \\ & f & 0 \\ & & 1 \end{bmatrix} \quad . \tag{8}$$

These parameterizations can be used in Equation 6. It will be seen in the experiments of Section 5 that this method gives good results on synthetic data as well as on real data.

[2]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 the present redundancy can not be used optimally.

[3]When only a short image sequence is at hand, when the projective calibration is not very accurate or when the motion sequence is close to critical without additional constraints.

3.2 Linear Approach

In the case were 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 $(u, v) \rightarrow (0, 0)$ and the aspect ratio $\frac{f_y}{f_x} \rightarrow 1$. These assumptions simplify Equation 7 as follows:

$$\lambda \begin{bmatrix} f_k^2 & 0 & 0 \\ 0 & f_k^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{P}_k \begin{bmatrix} f_1^2 & 0 & 0 & a_1 \\ 0 & f_1^2 & 0 & a_2 \\ 0 & 0 & 1 & a_3 \\ a_1 & a_2 & a_3 & \|\mathbf{a}\|^2 \end{bmatrix} \mathbf{P}_k^\top \tag{9}$$

This can be regarded as a set of $6(n-1)$ linear equations in $2(n-1)+5$ unknowns: $\lambda_k f_k^2, \lambda_k, f_1^2, a_1, a_2, a_3$ and $\|\mathbf{a}\|^2$. Therefore when only two views are available the solution is only determined up to a one parameter family of solutions. Of course the rank 3 constraint of Ω still has to be enforced. This can be done by imposing that the determinant of the solution should be zero:

$$\det\left(\Omega_a - \gamma\Omega_b\right) = 0 \tag{10}$$

From the left-hand side of Equation 9 it can be seen that the following equations have to be satisfied $\omega_{11} = \omega_{22}, \omega_{12} = \omega_{13} = \omega_{23} = 0$. This can thus be imposed on the right-hand side, yielding $4(n-1)$ linear equations in f_1^2, a_1, a_2, a_3 and $\|\mathbf{a}\|^2$. 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. Imposing the rank 3 constraint in this case results in up to 4 possible solutions.

3.3 Detecting Critical Motion Sequences

It is outside the scope of this report to give a complete analysis of all possible critical motions which can occur for self-calibration. For the case where all internal camera parameters are known to be fixed, such an analysis was carried out by Sturm [17].

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 $\omega = \mathbf{I}$) [4]. In this case it can be verified that $df_x = \frac{1}{2}d\omega_{11}$, $df_y = \frac{1}{2}d\omega_{22}$, $du = d\omega_{13} = d\omega_{31}$, $dv = d\omega_{23} = d\omega_{32}$ and $ds = d\omega_{12} = d\omega_{21}$. 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.

[4] For a real image sequence this implies the transformation of the image points $m \rightarrow \mathbf{K}^{-1}m$.

By linearizing around $\omega = \mathbf{I}$ this yields $d\omega_{12}^{(k)} = 0$, $d\omega_{11}^{(k)} = d\omega_{22}^{(k)}$, $d\omega_{11}^{(k)} = d\omega_{11}^{(l)}$. Which can be rewritten as

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

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

$$\mathbf{C} d\omega = 0 \tag{12}$$

with $d\omega$ a column vector containing the differentials of the coefficients of the dual image absolute conic $\omega^{(k)}$ 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 Equation 12. Using Equation 5 a small change can be modeled as follows:

$$\mathbf{C} d\omega = \mathbf{C} \left[\frac{d\omega}{d\Omega} \right] d\Omega = \mathbf{C}' d\Omega \tag{13}$$

with $\Omega = [\Omega_{11}\Omega_{22}\Omega_{12}\Omega_{31}\Omega_{32}\Omega_{14}\Omega_{24}\Omega_{34}]^\top$ and the Jacobian $\left[\frac{d\omega}{d\Omega}\right]$ evaluated at the solution. To have the expression of Equation 13 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.

4 Overview of the Reconstruction Algorithm

The proposed self-calibration method is embedded in a system to automatically model metric reconstructions of rigid 3D objects from uncalibrated video sequences. The complete procedure for metric 3D reconstruction is summarized here.

1. Corners are matched between the different images of the sequence and the projective camera matrices $\mathbf{P}_k$ are robustly estimated from them. A more detailed explanation of this approach can be found in [1].

2. Self-Calibration

(a) First the focal lengths f_k are estimated using the linear algorithm (assuming that there is no skew and that the principal point is in the middle of the image). These results are refined using the nonlinear algorithm (same assumptions).

(b) (optional) Not only f_k, but also the principal points (u_k, v_k) are estimated for all the views. As initialization the simplified (biased) version with $\mathbf{P}_1 = [\mathbf{I}\,0]$ can be used. This should however be refined with the non-biased version of the algorithm.

(c) Upgrade from projective $\mathbf{P}_k$ to metric using the obtained absolute quadric.

3. Estimation of dense correspondence maps using the known epipolar geometry. The epipolar geometry is enough to yield a pairwise image rectification where all epipolar lines lay along the image scan lines. Dense correspondence matches are then computed along each scan line using hierarchical blockmatching and dynamic programming [2].

4. Reconstruction of a dense 3D model based on metric $\mathbf{P}_k$ and dense correspondence maps. The dense image matches are triangulated in space using the now calibrated cameras and a textured 3D surface model is obtained that can easily be loaded into computer graphic systems for realistic rendering [9].

5 Experiments

In this section a number of experiments are described. First some 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. The results confirmed the validity of the algorithm. A second sequence was recorded with varying internal camera parameters. A zoom factor ($2\times$) was applied while filming. The resulting reconstructions are visually convincing and preserve the metric properties of the original scenes (i.e. parallelism, orthogonality, ...).

5.1 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 internal camera parameters were chosen as follow. 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 ratio of one and no skew. The views were

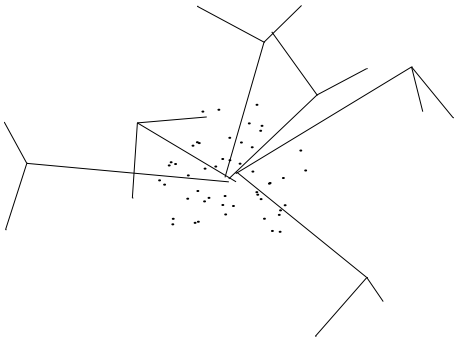

Figure 1: Example of sequence used for simulations (the views are represented by the image axis and optical axes of the camera in the different positions.)

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 figure 1.

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 report 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 values of 0, 0.1, 0.2, 0.5, 1, 1.5 and 2 pixels noise were used on sequences of 6 views. The results can be seen in Figure 2. 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 length. Sequences ranging from 4 to 40 views were used. The results are shown in Figure 3.

5.2 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 report to check if they indeed return constant intrinsic parameters for this sequence. In Figure 4 the images 0, 4, 8, 14, and 20 of the 24 images of the sequence are shown.

In the left plot of Figure 5 the focal length for every view is plotted for the different algorithms. The calculated focal lengths are almost constant as it should be. In the right

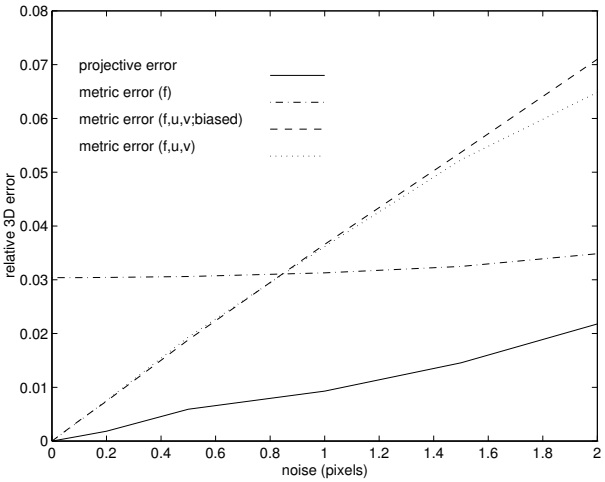

Figure 2: relative 3D error in function of noise for sequences of 6 views.

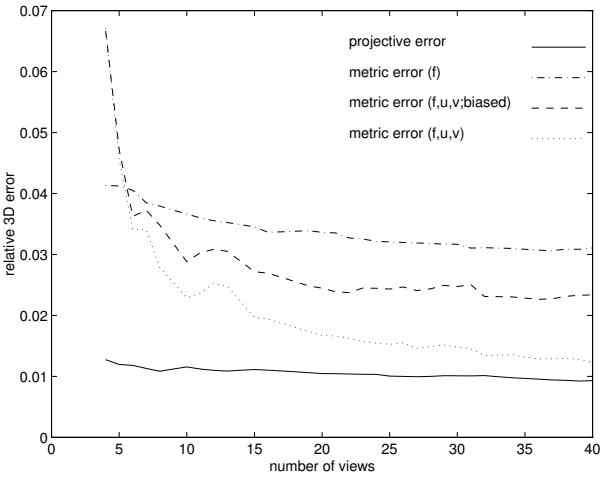

Figure 3: relative 3D error for sequences of different lengths (1.0 pixel of noise)

Figure 4: Some of the Images of the Arenberg castle which were used for the reconstruction

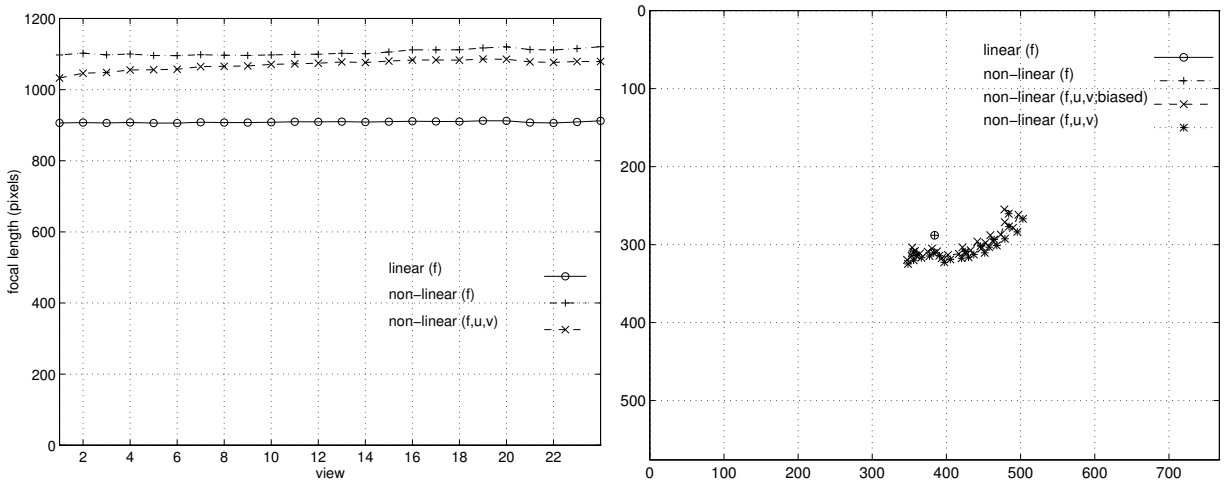

Figure 5: focal length (in pixels) versus views for the different algorithms (left), position of the principal point in the images for the different algorithms (right)

	angle ($\pm$std.dev.)
parallellism	1.0 ± 0.6 degrees
orthogonality	92.5 ± 0.4 degrees

Table 2: Results of metric measurements on the reconstruction

plot the position of the principal point is plotted for the different algorithms (for every view). It can be seen that when the principal point is not fixed by the algorithm, it varies over a few tens of pixels. 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.

To judge the visual quality of the reconstruction, different perspective views of the model were computed and displayed in Figure 6. The result is rather convincing, although at the moment only a partial model from a single image pair in a single view point is reconstructed. The metric quality of the reconstruction can be checked by projecting orthographic views from the main object axis top, left, and right as seen in Figure 7. The model shows a good degree of parallelism and orthogonality.

A quantitative assessment of these properties can be made by explicitely measuring angles directly on the object surface. For this experiment 6 lines were placed along prominent surface features, three on each object plane. They are aligned with the windows as indicated in Figure 8 by the black lines. The three lines inside of each object plane should be parallel to each other (angle between them should be 0 degree), while the lines of different object planes should be perpendicular to each other (angle between them should be 90 degree). The measurement on the object surface shows that this is indeed close to the expected values (see Table 2).

Figure 6: Perspective views of reconstruction

5.3 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). The sequence (8 images) can be seen in Figure 9. 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).

In the left plot of Figure 10 the focal length for every view is plotted for the different algorithms. It can be seen that the calculated values of the focal length correspond to what could be expected. In the right plot the position of the principal point is plotted for the different algorithms (for every view). For the algorithm which is not biased in favor of the first view the motion is smaller, which is more realistic. It is probable that too much noise is present to allow us to estimate the principal point accurately.

In Figure 11 a perspective view of the reconstruction is given, rendered both shaded and with surface texture mapping. The shaded view shows that even most of the small details of the object are retrieved. Figure 12 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 given. Note the arbitrarily shaped free-form surface that has been reconstructed.

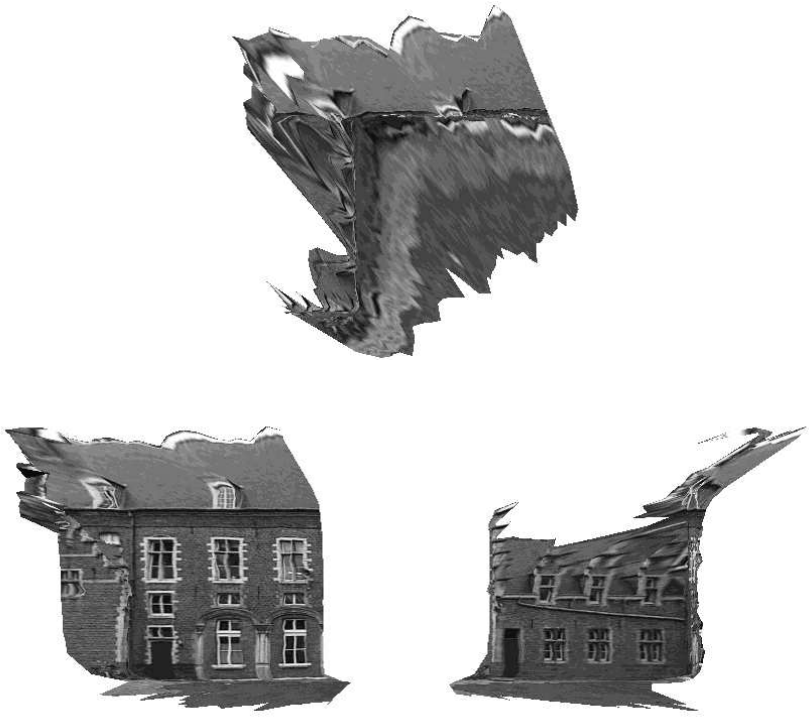

Figure 7: Orthographic views of reconstruction (notice parallelism and orthogonality)

Figure 8: Lines used for quantitative assessment of parallellism and orthogonality (superimposed in black)

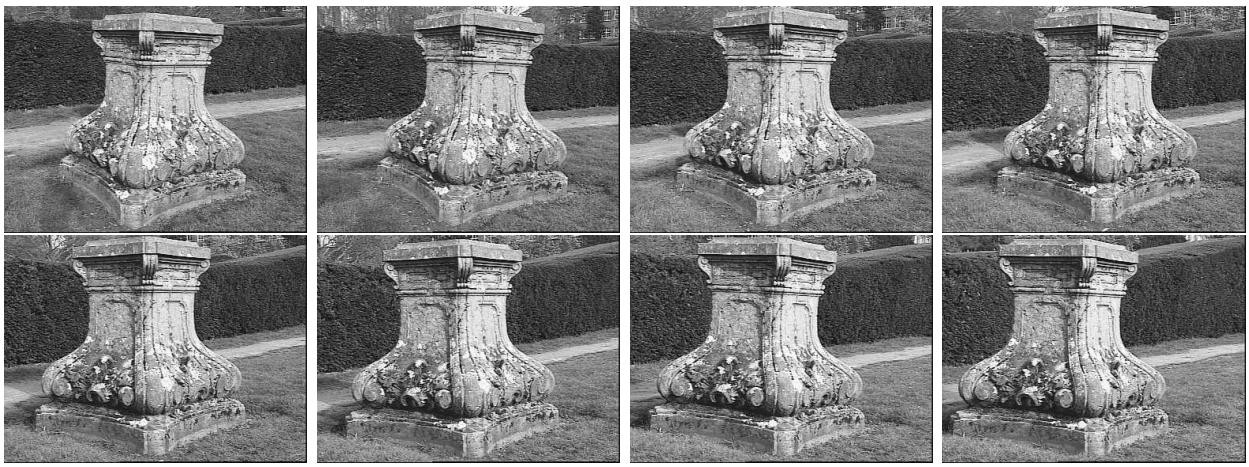

Figure 9: Sequence 2

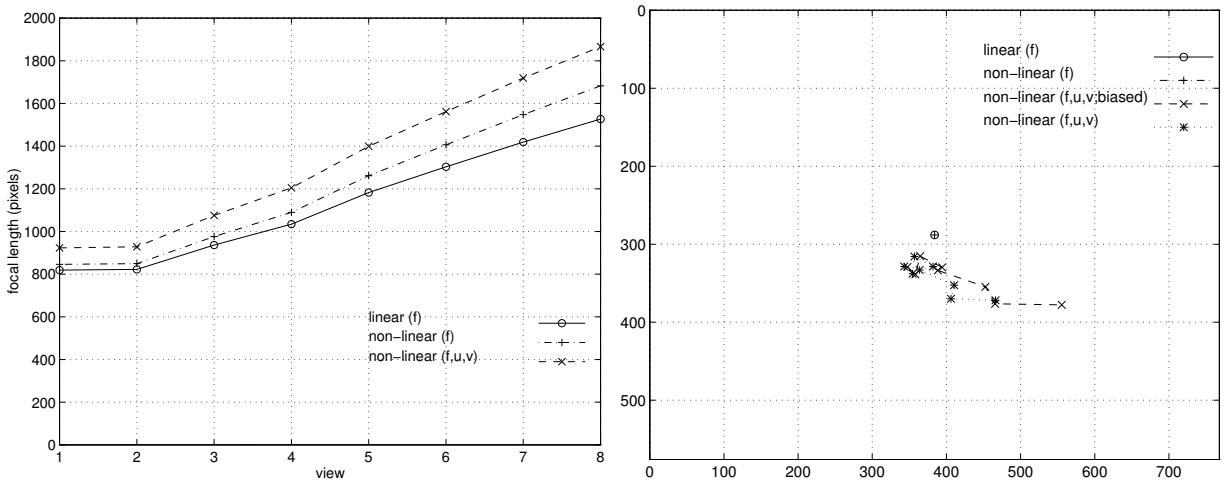

Figure 10: focal length (in pixels) versus views for the different algorithms (left), position of the principal point in the images for the different algorithms (right)

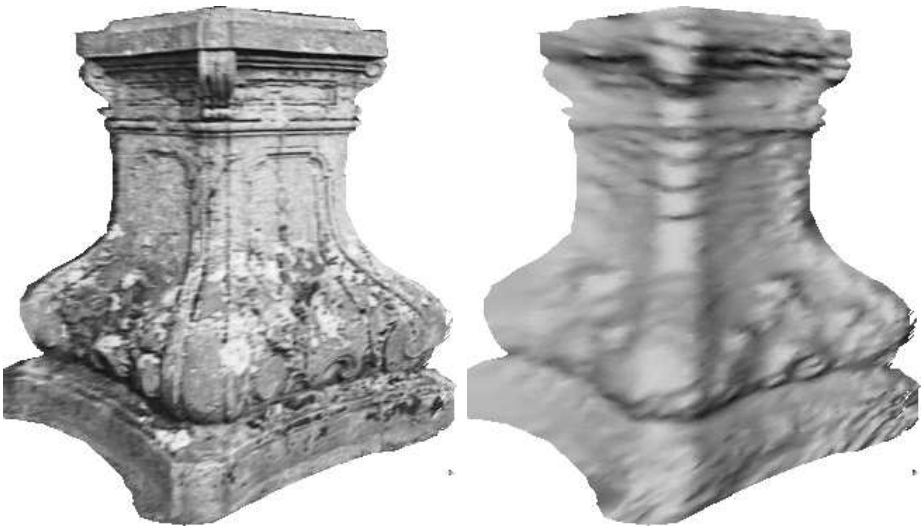

Figure 11: Perspective view of the reconstruction (with texture and with shading).

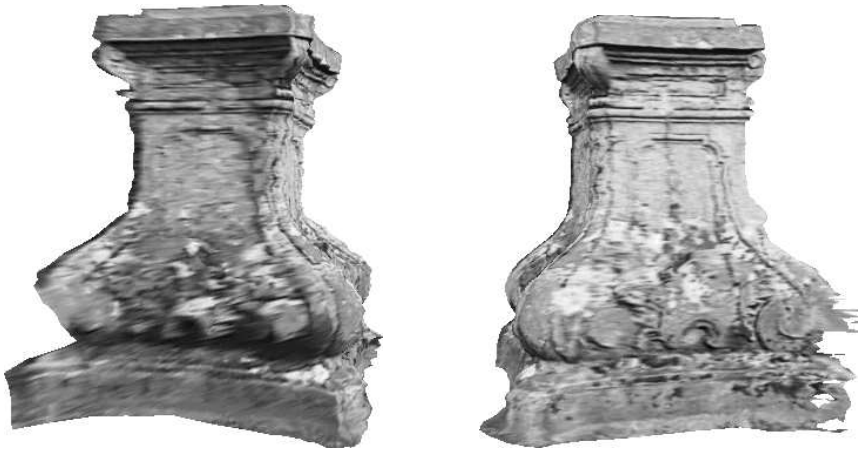

Figure 12: Left and right perspective view of the reconstruction.

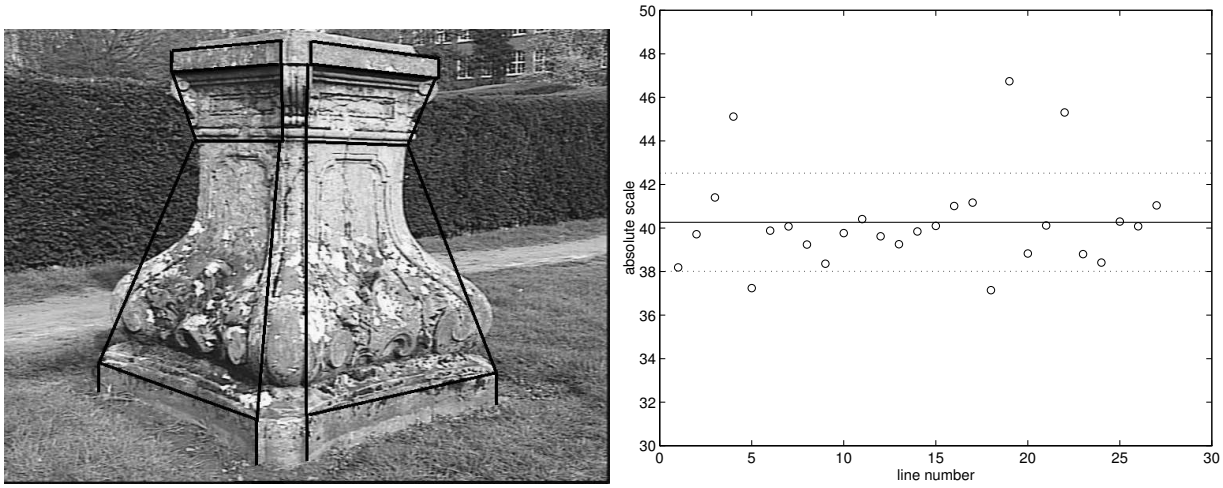

Figure 13: Comparison between real and reconstructed object distances: reference lines superimposed in black (left), measurements of absolute scale for all lines (right).

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 length 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 Figure 13. From each comparison the absolute object scale factor was computed and plotted as function of the evaluated line number. 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 outliers. Averaging all 27 measured distances gave a consistant scale factor of 40.25 (solid line) with a standard deviation of 5.4% overall (dotted line). 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.

6 Conclusions

This report focussed on self-calibration and metric reconstruction in the presence of varying and unknown internal camera parameters. The calibration models used in previous research are on 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 report 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 internal 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 point were considered. It is proposed to always start with the principal point in the center of the image and 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 report a general method is proposed which detects critical and quasi-critical motion sequences.

The different methods are validated with experiments which are carried out on real as well as synthetic image sequences. The latter ones are used to analyze noise sensitivity and influence of the length of the sequence. The former ones show the practical feasibility of the approach.

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.

Acknowledgements

We would like to thank Andrew Zisserman and his group from Oxford for supplying us with robust projective reconstruction software. A specialization grant from the Flemish Institute for Scientific Research in Industry (IWT) and the financial support from the EU ACTS project AC074 'VANGUARD' are also gratefully acknowledged.

References

[1] P. Beardsley, P. Torr and A. Zisserman 3D Model Acquisition from Extended Image Sequences *Proc. ECCV'96*, vol.2, pp.683-695

[2] L. Falkenhagen, Depth Estimation from Stereoscopic Image Pairs assuming Piecewise Continuous Surfaces, *European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Productions*, Hamburg, Germany, Nov. 1994.

[3] O. Faugeras, What can be seen in three dimensions with an uncalibrated stereo rig, *Proc. ECCV'92*, pp.563-578.

[4] O. Faugeras, Q.-T. Luong and S. Maybank. Camera self-calibration: Theory and experiments, *Proc. ECCV'92*, pp.321-334.

[5] R. Hartley, Estimation of relative camera positions for uncalibrated cameras, *Proc. ECCV'92*, pp.579-587.

[6] R. Hartley, Euclidean reconstruction from uncalibrated views, Applications of invariance in Computer Vision, LNCS 825, Springer-Verlag, 1994.

[7] A. Heyden, K. Åström, Euclidean Reconstruction from Constant Intrinsic Parameters *Proc. ICPR'96*.

[8] A. Heyden, K. Åström, Euclidean Reconstruction from Image Sequences with Varying and Unknown Focal Length and Principal Point *Proc. CVPR'97*.

[9] R. Koch, 3-D Surface Reconstruction from Stereoscopic Image Sequences, *Proc. ICCV'95*, Cambridge, USA, June 1995.

[10] S. Laveau, O. Faugeras, Oriented Projective Geometry for Computer Vision, *Proc. ECCV'96*, vol.1, pp. 147-156.

[11] M. Pollefeys, L. Van Gool and M. Proesmans, Euclidean 3D Reconstruction from Image Sequences with Variable Focal Lengths, *Proc. ECCV'96*, vol.1, pp. 31-42.

[12] M. Pollefeys, L. Van Gool and A. Oosterlinck, The Modulus Constraint: A New Constraint for Self-Calibration, *Proc. ICPR'96*, pp. 349-353.

[13] M. Pollefeys and L. Van Gool, A stratified approach to self-calibration, *Proc. CVPR'97*.

[14] M. Pollefeys and L. Van Gool, Self-calibration from the absolute conic on the plane at infinity, *Proc. CAIP'97*.

[15] Q.-T. Luong and O. Faugeras, Self Calibration of a moving camera from point correspondences and fundamental matrices, IJCV, vol.22-3, 1997.

[16] J.G. Semple and G.T. Kneebone, Algebraic Projective Geometry, Oxford University Press, 1952.

[17] P. Sturm, Critical Motion Sequences for Monocular Self-Calibration and Uncalibrated Euclidean Reconstruction, *Proc. CVPR'97*.

[18] B. Triggs, The Absolute Quadric, *Proc. CVPR'97*.

[19] C. Zeller and O. Faugeras, Camera self-calibration from video sequences: the Kruppa equations revisited. Research Report 2793, INRIA, 1996.