Stratified Self-Calibration with the Modulus Constraint

Marc Pollefeys and Luc Van Gool

Abstract— **In computer vision and especially for 3D reconstruction, one of the key issues is the retrieval of the calibration parameters of the camera. These are needed to obtain metric information about the scene from the camera. Often these parameters are obtained through cumbersome calibration procedures. There is a way to avoid explicit calibration of the camera. Self-calibration is based on finding the set of calibration parameters which satisfy some constraints (e.g. constant calibration parameters). Several techniques have been proposed, but it often proved difficult to reach a metric calibration at once. Therefore in this paper a stratified approach is proposed which goes from projective, through affine to metric. The key concept to achieve this is the modulus constraint. It allows retrieval of the affine calibration for constant intrinsic parameters. It is also suited for use in conjunction with scene knowledge. In addition, if the affine calibration is known, it can also be used to cope with a changing focal length.**

I. Introduction

In recent years several methods were proposed to obtain the calibration of a camera from correspondences between several views of the same scene. These methods are based on the rigidity of the scene and on the constancy of the intrinsic camera parameters. Most existing methods start from the projective calibration and then immediately try to solve for the intrinsic parameters. However, they all have to cope with the affine parameters (i.e. the position of the plane at infinity).

Faugeras et al. [7] eliminated these affine parameters yielding two Kruppa equations for each pair of views. A more robust approach was proposed by Zeller and Faugeras [42]. Heyden and Åström [13] and Triggs [38] proposed methods based on the absolute quadric. Hartley [10] does a minimization on all eight parameters to obtain metric projection matrices. Most of these methods encounter problems as they have to solve for many parameters at once from nonlinear equations.

This problem prompted a stratified approach, where starting from a projective reconstruction an affine reconstruction is obtained first and used as the initialization towards metric reconstruction. A similar method has been proposed by Armstrong et al. [1] based on the work of Moons et al. [23]. But this method needs a pure translation which is for example not easy to ensure with a hand-held camera. Successful stratified approaches have also been proposed for the self-calibration of fixed stereo rigs [44], [3], [15].

A first general approach for a single camera based on the modulus constrain needed at least four views [27] to obtain the self-calibration. The method which is developed further in this paper was presented in [28]. This method is more robust and can obtain the metric calibration of a camera setup from only three images. This paper also investigates the possibility of using only two views of the scene. It will be shown that combining constraints on the intrinsic camera parameters with characteristics inferred from the scene can solve the calibration where none of them separately could. Finally the modulus constraint can also be used to obtain the self-calibration in spite of a varying focal length [26].

An alternative approach to self-calibration was recently proposed [30]. In cases where all the intrinsic camera parameters except the focal length are (approximately) known, a solution can be obtained through a linear algorithm (allowing a varying focal length). This solution is then used as an initialization for a non-linear algorithm which allows varying/constant/known camera parameters. This algorithm therefore offers more flexibility, but it also seems to fail in certain cases. Although this needs to be studied more in detail the causes are probably the parametrization and sensitivity of the algorithm to near critical motion sequences. In the cases were both methods are applicable and succeed in their initialization phase the final results will in general be identical since the refinement step is the same for both methods.

This paper is organized as follows. In section II the notation and the geometric concepts used throughout the paper are introduced. In section III the modulus constraint is derived. Section IV explains how the self-calibration problem can be solved using the modulus constraint. In section V two other applications of the modulus constraint are presented. The different methods are validated through the experiments of section VI. Finally, section VII is the conclusion. In addition some more tedious derivations are given in the appendices.

II. Basic principles and notations

In this section the basic principles and notations used in the rest of the paper are introduced. These concepts are organized in projective, affine and metric following the stratification of geometric space. More details on this stratification can be found in [5], [19], [32].

Some research on which this paper draws are also presented in this section.

Marc Pollefeys is with the VISICS-PSI-ESAT group of the Katholieke Universiteit Leuven, Kardinaal Mercierlaan 94, B-3001 Heverlee, Belgium. E-mail: Marc.Pollefeys@esat.kuleuven.ac.be

Luc Van Gool is heading the VISICS-PSI-ESAT group of the Katholieke Universiteit Leuven, Kardinaal Mercierlaan 94, B-3001 Heverlee, Belgium and the Image Science Group (BIWI) of ETH Zürich, Gloriastrasse 35, ETH Zentrum, CH-8092 Zürich, Switzerland, E-mail: Luc.VanGool@esat.kuleuven.ac.be and Luc.VanGool@vision.ee.ethz.ch

A. Projective geometry and reconstruction

Projective geometry is used throughout the paper to describe the perspective projection of the scene onto the images. This projection is described as follow

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

where $\mathbf{P}$ is a 3×4 projection matrix describing the perspective projection process, $M = [XYZ\,1]^\top$ 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.

With $C = [t^\top 1]^\top$ the optical center of the camera and $\mathbf{e}$ the projection of the origin (i.e. $[0\,0\,0\,1]^\top$) in the image, the projection matrices can in general be written as follow:

$$\mathbf{P} = [\mathbf{H}|\ \text{-}\mathbf{H}t] = [\mathbf{H}|\mathbf{e}] \tag{2}$$

with $\mathbf{H}$ the homography describing the projection of points from the reference plane to the image plane. When the geometry is only determined up to a projective (or affine) transformation, the first projection matrix can be chosen as follows $\mathbf{P}_1 = [\mathbf{I}_{3\times3}|\mathbf{0}_{3\times1}]$. This implies that the camera and scene coordinate frames are aligned. In that case the homographies of (2) also describe the transfer from points lying in the reference plane from the first image to the image under consideration.

All these homographies are related through the following equation:

$$\mathbf{H}' = \mathbf{H} + \mathbf{e}.\mathbf{a}^\top \ . \tag{3}$$

With $\mathbf{H}'$ the homography of the plane $[\mathbf{a}^\top\,1]M = 0$ ($\mathbf{a} = [a_1 a_2 a_3]^\top$ is a 3-vector).

It has been shown that it is possible to obtain a projective calibration from an uncalibrated image sequence without needing any constraint on the intrinsic camera parameters [6], [9]. An important effort has been done in the last years to develop robust techniques for automatically obtaining this from image sequences [2], [43], [21], [22], [37], [39], [40]. In our paper the approach of Beardsley et al. [2] was followed for the experiments with real images. For the synthetic experiments an approach similar to [36] was taken.

These algorithms obtain a set of consistent projective camera matrices $\mathbf{P}_1, \mathbf{P}_2, ...\mathbf{P}_n$ for the images of the sequence. This is the starting point of the self-calibration methods proposed in this paper.

B. Affine geometry and the plane at infinity

The affine geometry differs from the projective geometry by the identification of a special plane, i.e. the plane at infinity. Where in 3D projective space all planes are intersecting, in affine space some planes are not intersecting, but are parallel. In fact their intersection line is located in the plane at infinity which is strictly speaking not part of the affine 3D space. This plane is setwise invariant under the group of affine transformations. Therefore any rigid motion of the camera will keep its relative position and orientation towards the plane at infinity unchanged.

The self-calibration method proposed in this paper focuses first on the identification of this plane before retrieving the absolute conic which is embedded in it (see next paragraph).

Since the homographies for all the planes can be expressed as in (3), the homographies of the plane at infinity Π_∞ can be expressed through this equation. This means that for some $\mathbf{a}$ the affine camera projection matrices –with Π_∞ as reference plane– can be obtained starting from any projective representation with an arbitrary reference plane:

$$\mathbf{P}_A = \left[\mathbf{H} + \mathbf{e}\mathbf{a}^\top|\mathbf{e}\right] = [\mathbf{H}_\infty|\mathbf{e}] \ . \tag{4}$$

To obtain an affine calibration it is therefore necessary to determine $\mathbf{a}$. In section III a constraint on $\mathbf{a}$ will be derived.

C. Metric geometry and the absolute conic

When the ambiguity on the geometry is metric (i.e. Euclidean up to an unknown scale factor), the camera projection matrices can be put in the following form:

$$\mathbf{P}_M = \mathbf{K}[\mathbf{R}\,|\text{-}\mathbf{R}t] \tag{5}$$

with t and $\mathbf{R}$ indicating the position and orientation of the camera and $\mathbf{K}$ an upper diagonal 3×3 matrix containing the intrinsic camera parameters:

$$\mathbf{K} = \begin{bmatrix} f_x & s & u_x \\ & f_y & u_y \\ & & 1 \end{bmatrix} \tag{6}$$

where f_x and f_y represent the focal length divided by the horizontal and vertical pixel dimensions, s is a measure of the skew and (u_x, u_y) is the principal point.

The key concept for self-calibration is the absolute conic. Besides the plane at infinity another geometric entity stays constant under all rigid transformations of space. This entity is called the absolute conic and encodes the metric structure of the scene (i.e. Euclidean structure up to scale). It is an imaginary conic situated on the plane at infinity. It corresponds to the equation $X^2 + Y^2 + Z^2 = 0$ and $T = 0$ (for points $M = [X\,Y\,Z\,T]^\top$).

Before going into detail some conic notations are introduced. A conic is represented by a 3×3 symmetric matrix $\mathbf{C}$. A point m on the conic satisfies $m^\top\mathbf{C}m = 0$. A dual (or line) conic is represented by a 3×3 matrix $\mathbf{C}^*$. A line l tangent to the conic $\mathbf{C}$ satisfies $l^\top\mathbf{C}^*l = 0$. Provided $\mathbf{C}$ is full rank $\mathbf{C}^* = \mathbf{C}^{-1}$. Under a homography $m' = \mathbf{H}m$, a conic and a dual conic transform as

$$\mathbf{C}' = \mathbf{H}^{-\top}\mathbf{C}\mathbf{H}^{-1} \text{ and } \mathbf{C}'^* = \mathbf{H}\mathbf{C}^*\mathbf{H}^\top \ . \tag{7}$$

If the camera does not change, then the image of the absolute conic ω and its dual ω^* will also stay the same for all views. These are represented by the following matrices:

$$\omega = \mathbf{K}^{-\top}\mathbf{K}^{-1} \text{ and } \omega^* = \mathbf{K}\mathbf{K}^\top \tag{8}$$

with $\mathbf{K}$ the upper triangular matrix containing the intrinsic parameters of the camera.

The constraint that the dual image of the absolute conic should be the same for all the views can be expressed as follows [10]:

$$\lambda \mathbf{K}\mathbf{K}^\top = \mathbf{H}_{\infty ij}\mathbf{K}\mathbf{K}^\top \mathbf{H}_{\infty ij}^\top \ . \tag{9}$$

Scaling $\det \mathbf{H}_{\infty ij}$ to 1 eliminates the scale factor λ. Therefore once $\mathbf{H}_{\infty ij}$ is known (9) represents a set of linear equations from which the elements of $\mathbf{K}\mathbf{K}^\top$ can be obtained. The intrinsic camera parameters can then be obtained through Cholesky factorization.

III. The modulus constraint

A stratified approach to self-calibration first requires a method to identify the plane at infinity. The property of the homographies for this plane –called infinity homographies in the remainder of this paper– derived in this paragraph will be used for this purpose. The infinity homography from view i to j can be written as a function of the metric entities of (5) or explicitly as functions of projective entities and the position of the plane at infinity in the specific projective reference frame. Both representations are given in the following equation:

$$\begin{aligned} \mathbf{H}_{\infty ij} \ &\propto \ \mathbf{H}_{\infty j}\mathbf{H}_{\infty i}^{-1} \ \propto \ \underbrace{\mathbf{K}\mathbf{R}_j\mathbf{R}_i^{-1}\mathbf{K}^{-1}}_{\sim metric} \\ &\propto \ \underbrace{(\mathbf{H}_j + \mathbf{e}_j\mathbf{a}^\top)(\mathbf{H}_i + \mathbf{e}_i\mathbf{a}^\top)^{-1}}_{\sim projective} \ . \end{aligned} \tag{10}$$

From the Euclidean representation it follows that $\mathbf{H}_{\infty ij}$ is conjugated[1] with a rotation matrix (up to a scale factor) which implies that the 3 eigenvalues of $\mathbf{H}_{\infty ij}$ must have the same moduli, hence the modulus constraint [26], [27]. Note from (10) that this property requires the intrinsic camera parameters to be constant.

This can be made explicit by writing down the characteristic equation for the infinity homography:

$$\det(\mathbf{H}_{\infty ij} - \lambda \mathbf{I}) = l_3\lambda^3 + l_2\lambda^2 + l_1\lambda + l_0 = 0 \tag{11}$$

It can be shown that the following condition is a necessary condition for the roots of (11) to have equal moduli (see Appendix A):

$$l_3 l_1^3 = l_2^3 l_0 \tag{12}$$

This yields a constraint on the 3 affine parameters a_1, a_2, a_3 by expressing l_3, l_2, l_1, l_0 as a function of these parameters (using the projective representation in (10)). The inverse in the projective representation of $\mathbf{H}_{\infty ij}$ can be avoided by using the constraint $\det(\mathbf{H}_{\infty j} - \lambda \mathbf{H}_{\infty i}) = 0$ which is equivalent to (11). Factorizing this expression using the multi-linearity of determinants, l_3, l_2, l_1, l_0 turn out to be *linear* in a_1, a_2, a_3 (see Appendix B). Therefore between every pair of views a modulus constraint is obtained:

$$\mathcal{M}_{ij} : l_{3ij} l_{1ij}^3 - l_{2ij}^3 l_{0ij} = 0 \tag{13}$$

[1] Matrices $\mathbf{A}$ and $\mathbf{B}$ are conjugated if $\mathbf{A} = \mathbf{C}\mathbf{B}\mathbf{C}^{-1}$ for some matrix $\mathbf{C}$. Conjugated matrices have the same eigenvalues.

This results in a polynomial equation of degree four in a_1, a_2 and a_3. In general three of these constraints should only leave a finite number of solutions – not more than 64. With more equations only one possible solution should persist except for some critical motion sequences (e.g. sequences with rotation about parallel axes). See [33] for a more complete discussion of this problem.

IV. Self-calibration from constant intrinsic parameters

In section III a constraint on the location of the plane at infinity was derived for every pair of views. These constraints can thus be used for self-calibration. Once the affine structure has been recovered, it is easy to upgrade the structure to metric using the concepts of section II-C.

A. Finding the plane at infinity

A minimum of three views is needed to allow for self-calibration using the modulus constraint. Three views yield the following constraints $\mathcal{M}_{12}, \mathcal{M}_{13}$ and $\mathcal{M}_{23}$. This third constraint $\mathcal{M}_{23}$ is in general independent of the two first constraints. The fact that $\mathbf{H}_{12}$ and $\mathbf{H}_{13}$ are both conjugated with rotation matrices (i.e. have the same eigenvalues as rotation matrices) does not imply that this is also the case for $\mathbf{H}_{23}$ since the eigenvectors of $\mathbf{H}_{12}$ and $\mathbf{H}_{13}$ could be different.

In the minimal case of three constraints $(\mathcal{M}_{12}, \mathcal{M}_{13}, \mathcal{M}_{23})$ for three unknowns $(\mathbf{a} = [a_1 a_2 a_3]^\top)$ the most adequate method to solve for these unknowns is to use continuation. This is a numerical method which finds all the solutions of a system of polynomial equations (for more details see [24]). Having three polynomial equations of degree 4 a maximum of 64 solutions can be found.

Many of these solutions can be ruled out. Since only real solution are of interest, all complex solutions can be discarded. In addition it should be checked that the eigenvalues of $\mathbf{H}_{ij}$ effectively correspond to these of a rotation matrix (see Appendix A).

For the remaining solutions the intrinsic camera parameters can be computed using the method proposed in section IV-B. Only the solution yielding (quasi)-constant parameters should be taken into account. If more than one solution persists the most plausible one is chosen (no skew, aspect ratio around one, principal point around the middle of the image). This approach worked well for over 90% of the experiments (see Section IV-E). When more views are at hand it is important to take advantage of all the available constraints to obtain a more accurate and robust solution. This can be done by combining the different constraints into a least squares criterion which can be minimized through nonlinear optimization:

$$\mathcal{F} = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathcal{M}_{ij}^2 \ . \tag{14}$$

The recommended method for initialization of the minimization is the continuation method, but trying out a few random starting values often also gives good results.

B. Finding the absolute conic and refining the calibration results

Once the plane at infinity has been obtained, equation (9) can be used to compute the absolute conic as described in section II-C. This is a linear method.

These results can be refined through a non-linear minimization step. The same constraint can be used for this purpose, but $\mathbf{a}$ should appear explicitly in $\mathbf{H}_{\infty ij}$:

$$\lambda \mathbf{K}\mathbf{K}^\top = [\mathbf{H} + \mathbf{e}\mathbf{a}^\top]\mathbf{K}\mathbf{K}^\top[\mathbf{H} + \mathbf{e}\mathbf{a}^\top]^\top \tag{15}$$

In this case however the scale factor can not easily be eliminated. It is possible to consider every scale-factor as an additional unknown [13], but this poses problems and makes the scheme unworkable for longer image sequences. Another way to deal with this scale factor is to use the explicit formula derived in [29]. In practice however it is better to eliminate λ by normalizing both sides of equation (15) to unit Frobenius norm.

These equations can be solved through a nonlinear minimization of the following criterion:

$$\mathcal{C} = \sum_{k=2}^{n} \left\| \frac{\mathbf{H}_{\infty 1k}\mathbf{K}\mathbf{K}^\top\mathbf{H}_{\infty 1k}^\top}{\|\mathbf{H}_{\infty 1k}\mathbf{K}\mathbf{K}^\top\mathbf{H}_{\infty 1k}^\top\|_F} - \frac{\mathbf{K}\mathbf{K}^\top}{\|\mathbf{K}\mathbf{K}^\top\|_F} \right\|_F \tag{16}$$

The implementation presented in this paper uses a Levenberg-Marquard algorithm.

C. Comparison with other methods

In the introduction a brief overview of the existing approaches was given. Here a more detailed discussion of some methods based on the absolute conic are given.

The first method discussed here uses the *Kruppa equations* proposed by Faugeras et al. [7] and refined by Zeller [42]. In this method the affine parameters are eliminated from the equations. Only the fundamental matrices are needed and not a consistent projective frame for all cameras.

The Kruppa equation can be derived starting from (15). The affine parameters a_1, a_2, a_3 can be eliminated from these equations by multiplying with $[\mathbf{e}]_\times$ to the left and $[\mathbf{e}]_\times^\top$ to the right, obtaining

$$[\mathbf{e}]_\times \mathbf{K}\mathbf{K}^\top [\mathbf{e}]_\times^\top \propto \mathbf{F}\mathbf{K}\mathbf{K}^\top\mathbf{F}^\top \tag{17}$$

since the fundamental matrix $\mathbf{F} = [\mathbf{e}]_\times\mathbf{H}_\infty$. From the equations obtained here only 2 are independent. Scale factors can be eliminated by cross-multiplication. The disadvantage of this method is that a consistent supporting plane $\mathbf{\Pi}_\infty$ for ω is only indirectly enforced. The implementation used in Section IV-E uses a criterion similar to the one given in (16):

$$\mathcal{C}_{Kruppa} = \sum_{k=2}^{n} \left\| \frac{\mathbf{F}_{1k}\mathbf{K}\mathbf{K}^\top\mathbf{F}_{1k}^\top}{\|\mathbf{F}_{1k}\mathbf{K}\mathbf{K}^\top\mathbf{F}_{1k}^\top\|_F} - \frac{[e]_\times\mathbf{K}\mathbf{K}^\top[e]_\times^\top}{\|[e]_\times\mathbf{K}\mathbf{K}^\top[e]_\times^\top\|_F} \right\|_F . \tag{18}$$

The next method under consideration was proposed by Heyden and Åström [13]. It can be shown that the constraint proposed in their paper is equivalent to (15) (using $\tilde{\mathbf{a}} = \mathbf{a}\mathbf{K}$):

$$\begin{aligned} \lambda\mathbf{K}\mathbf{K}^\top &= [\mathbf{H}\mathbf{e}] \begin{bmatrix} \mathbf{K}\mathbf{K}^\top & \mathbf{K}\mathbf{K}^\top\mathbf{a}^\top \\ \mathbf{a}\mathbf{K}\mathbf{K}^\top & \mathbf{a}\mathbf{K}\mathbf{K}^\top\mathbf{a}^\top \end{bmatrix} \begin{bmatrix} \mathbf{H}^\top \\ \mathbf{e}^\top \end{bmatrix} \\ &= \mathbf{P} \begin{bmatrix} \mathbf{K}\mathbf{K}^\top & \mathbf{K}\tilde{\mathbf{a}}^\top \\ \tilde{\mathbf{a}}\mathbf{K}^\top & \|\tilde{\mathbf{a}}\|^2 \end{bmatrix} \mathbf{P}^\top \end{aligned} \tag{19}$$

with $\mathbf{P}_1 = [\mathbf{I}\,\mathbf{0}]$. The proposed implementation encountered convergence problems for longer sequences since additional unknowns were introduced to deal with the scale-factors of equation (19).

The constraint proposed by Triggs [38] are very similar to the previous ones. The main contribution being the introduction of the concept of the absolute quadric. This dual quadric (i.e. containing planes and not points) is degenerate and encodes both the plane at infinity and the absolute conic in a 4×4 rank 3 matrix. This formulation does not impose $\mathbf{P}_1 = [\mathbf{I}\,\mathbf{0}]$. This avoids the introduction of a bias in favor of the first view at the cost of some additional unknowns.

The main difference with the method of this paper therefore lies in the initialization of the minimization criterion. In stead of taking a stratified approach, both papers [13], [38] propose to initialize from an initial guess for the intrinsic camera parameters (from which an initialization for Π_∞ can linearly be computed).

Besides the stratified approach based on the modulus constraint (with and without refinement), the simulations of Section IV-E were also carried out using the the criterion (18) for the Kruppa equations [6] and the criterion (16) for the methods of [13], [38]. In these cases the intrinsic camera parameters were initialized from a close guess.

D. Stratified self-calibration algorithm

step 1: projective calibration
- *step 1.1:* sequentially compute the projective camera matrices for all the views (see [2]).

step 2: affine calibration
- *step 2.1:* formulate the modulus constraint $\mathcal{M}_{ij}$ for all pairs of views
- *step 2.2a:* find a set of initial solutions through continuation
- *step 2.2b:* solve the set of equations $\mathcal{M}_{ij}$ through minimization (for $n > 3$)
- *step 2.3:* compute the affine projection matrices

$$\mathbf{P}_{Ai} = \mathbf{P}_i\mathbf{T}^{-1} \text{ with } \mathbf{T} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{a}^\top & 1 \end{bmatrix}$$

step 3: metric calibration
- *step 3.1:* compute the dual image of the absolute conic from

$$\mathbf{K}\mathbf{K}^\top \propto \mathbf{H}_{\infty 1i}\mathbf{K}\mathbf{K}^\top\mathbf{H}_{\infty 1i}^\top$$

- *step 3.2:* find the intrinsic parameters $\mathbf{K}$ through Cholesky factorization
- *step 3.3:* refine the results through nonlinear minimization of $\mathcal{C}$

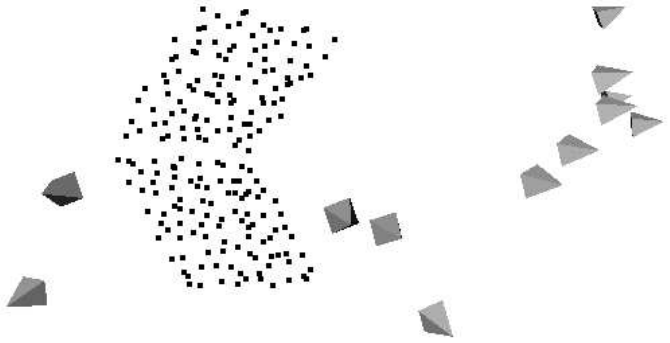


Fig. 1. Example of a sequence used for the simulations. Camera viewpoints are represented by small pyramids.

step 3.4: compute the metric projection matrices

$$\mathbf{P}_{Mi} = \mathbf{P}_i \mathbf{T}^{-1} \text{ with } \mathbf{T} = \begin{bmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{a}^\top & 1 \end{bmatrix}$$

E. Some simulations

The simulations were carried out on sequences ranging from 3 to 36 views. The scene consisted of 200 points which were positioned on 2 orthogonal 10×10 grids and then perturbed. For the calibration matrix the canonical form $\mathbf{K} = \mathbf{I}$ was chosen. The distance to the scene was chosen in such a way that the viewing conditions corresponded to a standard 35mm camera. The views were spaced 10^o apart, then a random perturbation was applied to their position and orientation. An example of such a sequence can be seen in Fig. 1.

The scene points were projected onto the images. Gaussian noise with an equivalent standard deviations of 0, 0.2, 0.5, 1 and 2 pixels for 500×500 images was added to these projections. One hundred sequences were generated for every parameter combination. The projective reconstruction was obtained with a variant of the factorization method proposed in [36]. The self-calibration was carried out using the method proposed in this paper. At first the modulus constraint was used to identify the plane at infinity, then the absolute conic was located. The results were then refined using the method of paragraph IV-B.

For the sake of comparison two alternative methods were included in the tests. The first method is based on the Kruppa equations [7], the second on a method similar to the methods using the absolute quadric [38] or the Kruppa constraints [13]. This second method in fact consists of immediately starting the refinement step of Section IV-B on some prior guess. Both these algorithms were initialized with values close to the exact solution. The focal length was initialized randomly within a factor of 2, the aspect ratio, the principal point and the skew were only slightly perturbed (within ± 0.1) as proposed in [38], [13]. For all these implementations care was taken of normalizing the data according to the recommendations of Hartley [12].

The results of the experiments can be seen in Fig. 2. The 2D backprojection error and the relative 3D error were computed. Since in a number of cases the algorithms fail to obtain satisfying results, the median values of the errors are shown. In addition the failure rate is given.

Notice that for short image sequences the 2D error is small while the 3D error is big. Using longer image sequences (and thus a wider angle between the extreme views) causes the 2D error to increases since an error on the intrinsic camera parameters can be compensated by distorting the structure, but only when the angle between the views is small. This also explains the important 3D error for short image sequences.

The algorithm based on the modulus constraint gives good results, especially when the refinement procedure is carried out. Immediately minimizing the criterion (16) on a prior guess of the intrinsic parameters gives similar results except that this method fails more frequently. The results for the Kruppa equations are poor. This algorithm fails often and the error is much higher than for the other algorithms.

These results tend to indicate that self-calibration results are more dependent on a good estimation of the plane at infinity than on the image of the absolute conic (i.e. the intrinsic camera parameters). In fact even when the modulus constraint itself does not converge to a satisfying calibration, the results are often good enough to succesfully initialize the refinement procedure. On the other hand starting from a prior guess the minimization sometimes even fails in the absence of noise, indicating that it was started in the attraction basin of a local minima. This is an argument in favor of a stratified approach to self-calibration. In addition the modulus constraint does not require any prior guess for the intrinsic camera parameters.

The Kruppa equations are similar to the equations used for refinement except that the position of the plane at infinity is eliminated from it. In the presence of noise this implies that for every image pair a different plane at infinity is implicitly used. All the other algorithms which are used here explicitly impose one consistent plane at infinity. An advantage of the Kruppa equations is that only the fundamental matrices are needed (i.e. pairwise calibration) and not a consistent projective frame over all the views. This also explains the observation of Luong [18] that the calibration results obtained with the Kruppa equations improved once the explicit camera poses were computed, since in this step a consistent frame was introduced.

In Fig. 3 some more results are given for the stratified approach. These results were obtained from 1000 experiments carried out on standard sequences of 12 views (0.5 pixels noise). The different subfigures are histograms of the obtained values for the different calibration parameters f_x, f_y, u_x, u_y, s, the layout was inspired by equation (6). The height of the peaks in the lower-right subfigure indicates how often the algorithm could find a solution. The left part gives the result obtained immediately with the modulus constraint and the right part after refinement. From this figure one can conclude that the calibration results are good and that there is no bias in the estimation. The refinement procedure clearly improves the results.

The succes of the stratified approach of this paper is

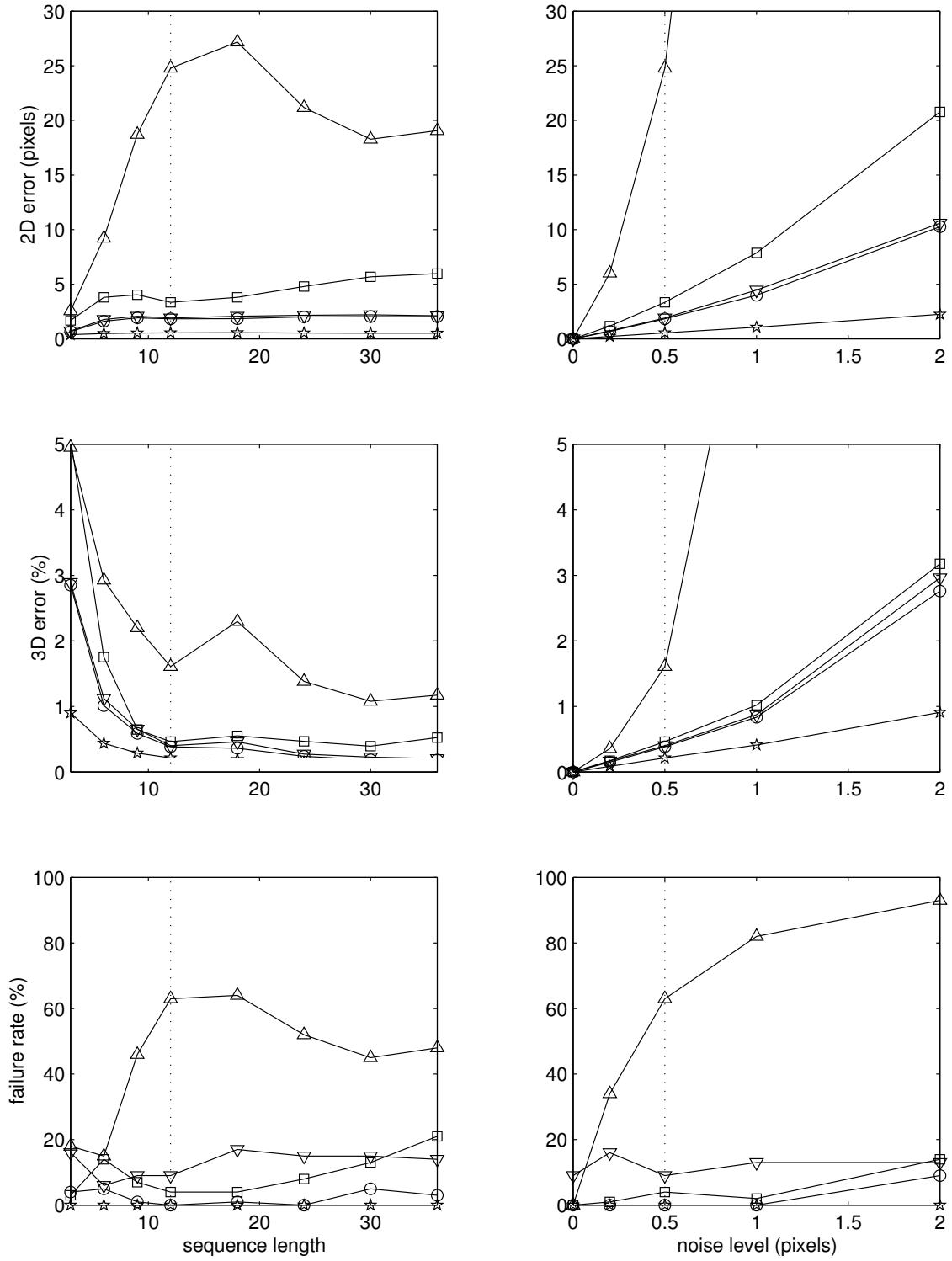


Fig. 2. Results of self-calibration experiment for image sequences varying from 3 to 36 views (left; noise 0.5 pixels) and noise ranging from 0.0 to 2.0 pixels (right; 12 views). The 2D reprojection error (top), the relative 3D error (middle) and the failure rate (bottom) are given. The median of the RMS error for 100 experiments was used as error measure. The algorithms are the modulus constraint $\square$, modulus constraints with refinement $\circ$, absolute conic ∇, Kruppa equations $\triangle$. As a reference the error is also given for the projective reconstruction $\star$.

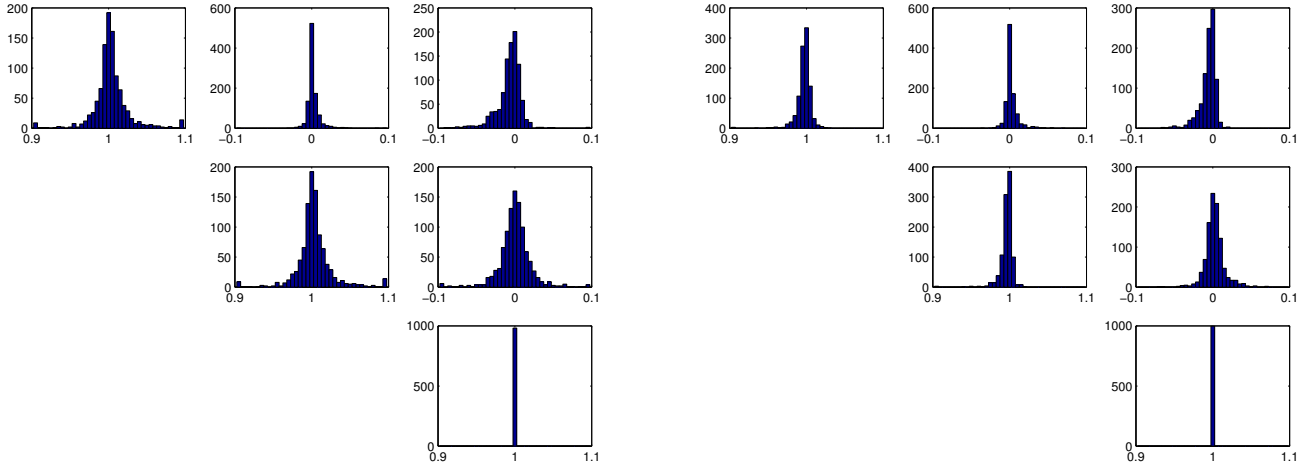


Fig. 3. Computed intrinsic camera parameters for 1000 sequences of 12 views (0.5 pixels noise). The results were obtained with the modulus constraint (left) and after refinement (right).

consistent with the very good calibration results that were obtained when the plane at infinity could be identified a priori (e.g. through pure rotation [11] or pure translation [23], [1]).

V. Special cases

The modulus constraint turns out to be particularly suited for some more specific self-calibration problems. Here two of these cases will be discussed.

A. Two images and two vanishing points

For a pair of images only one modulus constraint exists which is not enough to locate uniquely the plane at infinity. On the other hand, two vanishing points are also insufficient to obtain the plane at infinity. Combining both constraints yields enough information to obtain the affine calibration.

A.1 Vanishing points

Some scenes contain parallel lines. These result in vanishing points in the images. Techniques have been proposed to automatically detect such points [20], [41]. To identify the plane at infinity three vanishing points are necessary. By using the modulus constraint this number can be reduced to two. This reduction can be crucial in practice. For example in the castle sequence (see Fig. 6) two vanishing points could be extracted automatically in all frames using the algorithm described in [41]. The third point could only be found in some images at the end of the sequence. This is typical for a lot of scenes where one vanishing point is extracted for horizontal lines and one for vertical lines. Even when three candidate vanishing points are identified, the modulus constraint can still be very useful by providing a mean to check the hypothesis.

When a vanishing point is identified in the two images it can be used as follows to constraint the homography of the plane at infinity:

$$m_2 \propto \left[\mathbf{H} + \mathbf{e}\mathbf{a}^\top\right] m_1 \tag{20}$$

This results in one linear equation for the coefficients of $\mathbf{a}$ (from the three equations only two are independent due to the epipolar correspondence of m_1 and m_2 and one is needed to eliminate the unknown scale factor).

With two known vanishing points we are thus left with a one parameter family of solutions for $\mathbf{H}_\infty$:

$$\mathbf{H}_\infty = \mathbf{H} + \mathbf{e}(\lambda \mathbf{a}_1^\top + \mathbf{a}_0^\top) \ . \tag{21}$$

A.2 Using the modulus constraint

Applying the modulus constraint is much easier than in the general case. The coefficients l_3, l_2, l_1, l_0 (see (11)) can be evaluated for both $\mathbf{a}_1$ and $\mathbf{a}_0$. The modulus constraint in the two view case then takes on the following form:

$$\begin{aligned}&(\lambda l_3(\mathbf{a}_1) + l_3(\mathbf{a}_0))(\lambda l_1(\mathbf{a}_1) + l_1(\mathbf{a}_0))^3 \\ &= (\lambda l_2(\mathbf{a}_1) + l_2(\mathbf{a}_0))^3(\lambda l_0(\mathbf{a}_1) + l_0(\mathbf{a}_0)) \ .\end{aligned} \tag{22}$$

This results in a polynomial of degree 4 in only one variable λ (not degree 6 as Sturm anticipated [34]). Therefore at most 4 solutions are possible. Because (22) is only a necessary condition for $\mathbf{H}_\infty$ to be conjugated with a scaled rotation matrix, this property should be checked out. This can eliminate several solutions. If different solutions persist at this stage some can still be eliminated in the metric calibration part.

A.3 Metric calibration

Once the affine calibration is known equation (9) can be used. This will however not yield a unique solution. If R_∞ is the intersection point of the plane at infinity with the rotation axis to go from one viewpoint to the other, then not only the absolute conic, but also the degenerate conic $R_\infty R_\infty^\top$ is a fixed conic of the plane at infinity and thus also every linear combination of them. This results in a one parameter family of solutions for $\mathbf{K}\mathbf{K}^\top$.

Additional constraints like some known aspect ratio, orthogonality of the image axes or scene orientations can be used to restrict $\mathbf{K}\mathbf{K}^\top$ to one unique solution. If more than one affine calibration was still under consideration, these

constraints can also help out. Also the fact that $\mathbf{K}\mathbf{K}^\top$ should be positive definite and that the principal point should be more or less in the center of the image can be used to find the true affine, and thus also metric, calibration.

The application of these constraints is not so hard. Here the case of orthogonal orientations in the scene will be discussed. This can for example be applied when it is assumed that the extracted vanishing points correspond to orthogonal orientations.

The points v and v' are two vanishing points in the first image. The corresponding scene points can be obtained from the following equation:

$$v \propto \mathbf{K}[\mathbf{I}|\mathbf{0}] \begin{bmatrix} \tilde{v} \\ 0 \end{bmatrix} \tag{23}$$

Thus $\tilde{v} = \mathbf{K}^{-1}v$ represents the associated direction. Orthogonality now means that $\tilde{v}\tilde{v}' = 0$ or

$$v^\top \mathbf{K}^{-\top}\mathbf{K}^{-1}v' = 0 \ . \tag{24}$$

Therefore it is more appropriate to use the dual equation of (9):

$$\mathbf{K}^{-\top}\mathbf{K}^{-1} \propto \mathbf{H}_\infty^\top \mathbf{K}^{-\top}\mathbf{K}^{-1}\mathbf{H}_\infty \tag{25}$$

which of course also yields a one parameter family of solutions for $\mathbf{K}^{-\top}\mathbf{K}^{-1}$. Imposing (24) resolves this ambiguity.

From $\mathbf{K}^{-\top}\mathbf{K}^{-1}$ first $\mathbf{K}^{-1}$ is extracted by Cholesky factorization and subsequently inverted to obtain $\mathbf{K}$ which contains the camera intrinsic parameters.

A.4 Simulation

Some simulations were performed for the 2 view case with 2 known vanishing points. The same type of simulated data was used as in the general case (see section IV-E). A noise level of 1 pixel was chosen for the correspondences. The localization of the vanishing points was also perturbed by some noise. First the points at infinity (the 3D points corresponding to vanishing points in the images) were chosen as follows:

$$V_X = \begin{bmatrix} 1 & 0 & 0 & \nu_X \end{bmatrix}^\top \text{ and } V_Y = \begin{bmatrix} 0 & 1 & 0 & \nu_Y \end{bmatrix}^\top \tag{26}$$

were V_X and V_Y are the points at infinity corresponding to the X and Y direction, ν_X and ν_Y is a noise term with a mean of zero and some specific standard deviation σ_1. This means that in stead of having a point which is exactly at infinity, mostly a point is obtained of the order of $\frac{1}{\sigma_1}$. The vanishing points used for this simulation are then obtained as follows:

$$v_{Xi} = \mathbf{P}_i V_X \text{ and } v_{Yi} = \mathbf{P}_i V_Y \ . \tag{27}$$

These coordinates were additionally perturbed by some noise with standard deviation σ_2 pixels.

The modulus constraint and the two vanishing points were used to obtain the homography of the plane at infinity. This homography determines the absolute conic up to one parameter. This parameter was retrieved by imposing orthogonality between the directions corresponding to the vanishing points.

A first set of 1000 simulations was carried out with a small amount of noise (i.e. $\sigma_1 = 0.01, \sigma_2 = 10$), a second set of 1000 simulations was carried out with a high amount of noise (i.e. $\sigma_1 = 0.1, \sigma_2 = 100$). The results of these simulations can be seen in Fig. 4. The left part shows the results which were obtained with only a small amount of noise. The layout is similar to Fig. 3.

It can be seen that the results are good for a small amount of noise (left part of Fig. 4). With a lot of noise the results seriously deteriorate (see right part of Fig. 4). Although most solutions are situated around the correct values, a much bigger spread exist. In addition the algorithm only finds a possible solution in 65% of the cases.

B. *Varying focal length*

The modulus constraint can also be used for other purposes than affine calibration (see [26] for more details). The constraint depends on two conditions: the affine calibration *and* constant intrinsic camera parameters. For each view except the first one we get a valid constraint. This means that instead of "spending" the constraint on solving for affine calibration, one can in the traditional scheme – where affine calibration amounts from translation between the first two views [23]– use the constraint to retrieve one changing parameter for each supplementary view. An alternative approach would be to leave the camera parameters unchanged until the affine calibration is retrieved and only then start varying the camera parameters. The most practical application is to allow the focal length to vary. This allows to cope with autofocussing and zooming in the image sequence. A similar approach was proposed for stereo rigs [25].

B.1 Modeling the change in focal length

The first step is to model the effect of changes in focal length. These changes are relatively well described by scaling the image around a fixed focus of expansion c. This can be expressed as follows:

$$m_{f\,ik} = \mathbf{K}_f \, m_{ik} \text{ with } \mathbf{K}_f = \begin{bmatrix} 1 & 0 & (f^{-1}-1)c_x \\ 0 & 1 & (f^{-1}-1)c_y \\ 0 & 0 & f^{-1} \end{bmatrix} \tag{28}$$

where m_{ik} are the points that one would have seen if no change in focal length had occurred, and with $m_{f\,ik}$ the image points for some *relative* focal length f. Note that this equation also makes it possible to cancel the effect of the zoom by *de*zooming a projection matrix using $\mathbf{K}_f^{-1}$.

The first thing to do is to retrieve the focus of expansion c. Fortunately, this is easy for a camera with variable focal length, c being the only finite fixed point when varying the focal length without moving the camera. The affine camera calibration can then be retrieved from two views with a different focal length and a pure translation between the two views, using the method described in [26].

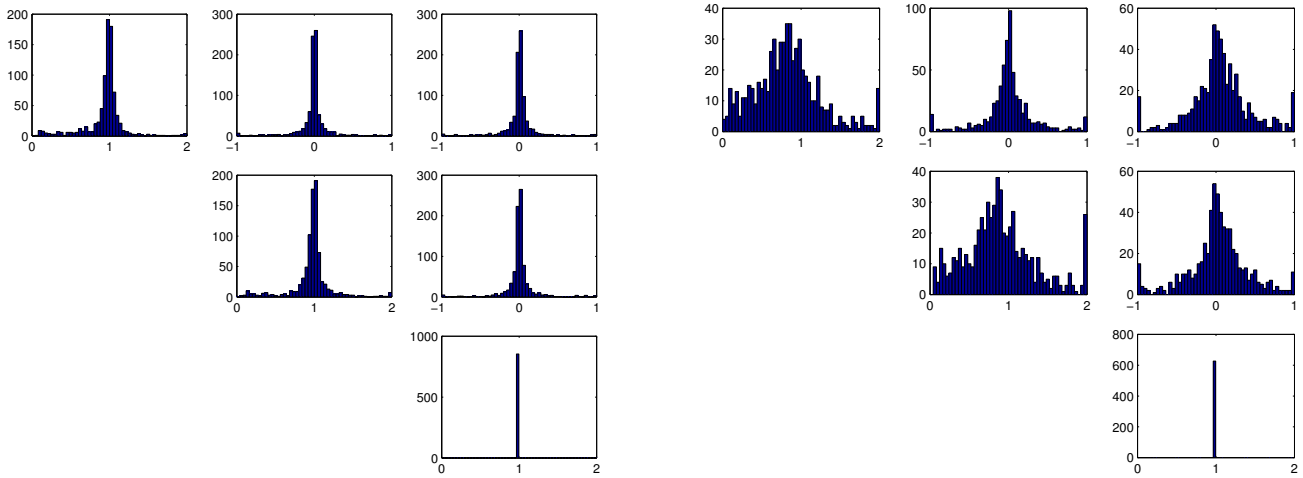


Fig. 4. Computed intrinsic camera parameters for low noise level (left) and high noise level (right).

B.2 Using the modulus constraint

The modulus constraint is only valid for an affinely calibrated camera with constant intrinsic camera parameters or after that the effect of the change in focal length has been taken away. Stated differently, the modulus constraint must be valid for a camera matrix $\mathbf{P}_A^{'} = \mathbf{K}_f^{-1}\mathbf{P}_A$. Writing down the characteristic equation, we get an equation like (11). Substituting the obtained coefficients in (12) we obtain a 4^{th} degree polynomial equation in f:

$$\alpha_4 f^4 + \alpha_3 f^3 + \alpha_2 f^2 + \alpha_1 f + \alpha_0 = 0 \qquad (29)$$

This equation has 4 possible solutions. It can be proven that if f is a real solution, then $-f$ must also be a solution [26]. Imposing this to equation (29) yields the following result:

$$f = \sqrt{\frac{\alpha_1}{\alpha_3}} \ . \qquad (30)$$

Now that f has been retrieved, K_f^{-1} can be used to get normalized images and cameras. These affine camera projection matrices can then be upgraded to metric as described in section IV-B.

B.3 Simulation

Some simulations were also carried out for this case. Here again, the same type of simulated data was used as in the general case (see section IV-E). Different noise levels were used for the simulations. For every experiment four views were generated.

The first two only differ in focal length ($f_1 = 1, f_2 = 2$) which allows us to estimate the focus of expansion. The focal length for the other views is chosen at random between f_1 and f_2. For the third view a pure translation was carried out. From this the affine reconstruction was obtained, in spite of an unknown change in focal length. For the last image a combination of a rotation and a translation was used.

For every noise level one hundred experiments were carried out. The results for 0.2 and 2 pixels of noise are shown in Fig. 5. This corresponds to a low and a high level of noise on the correspondences. The layout is similar to Fig. 3.

It can be seen that the results are very good for a small amount of noise (left part of Fig. 4). With more noise the results are still good. The other computed parameters (i.e. the focus of expansion and the relative focal lengths for the different views) are not shown. In general the focus of expansion and the relative focal length for views 2 and 3 are very accurately obtained. This is due to the high redundancy of the equations in these cases. The quality of the estimate for the relative focal length for view 4 is of the same order as the absolute estimate of the focal length (i.e. f_x).

VI. Experiments

In the previous section some simulation results were already given. In this section experiments carried out on real image sequences are described. The different methods proposed in this paper are successively discussed.

A. Experiments with constant intrinsic parameters

The stratified approach to the classical self-calibration problem that was proposed in this paper has been validated on several real video sequences. Here the results obtained on two sequences of the Arenberg castle will be shown.

A.1 Sequence 1

In this section results obtained from real sequences are presented. The calibration can be evaluated qualitatively by looking at the reconstruction. Different parts of the Arenberg castle in Leuven were filmed. These were recorded with a video camera. The first sequence is shown in Fig. 6. The approach which was followed to generate the 3D model can be briefly summarized as follows. First the corner matching and the projective camera matrices were obtained following the method described in [2]. These camera matrices were upgraded to metric using the self-calibration method described in this text and then a 3D model was generated using these cameras and a dense cor-

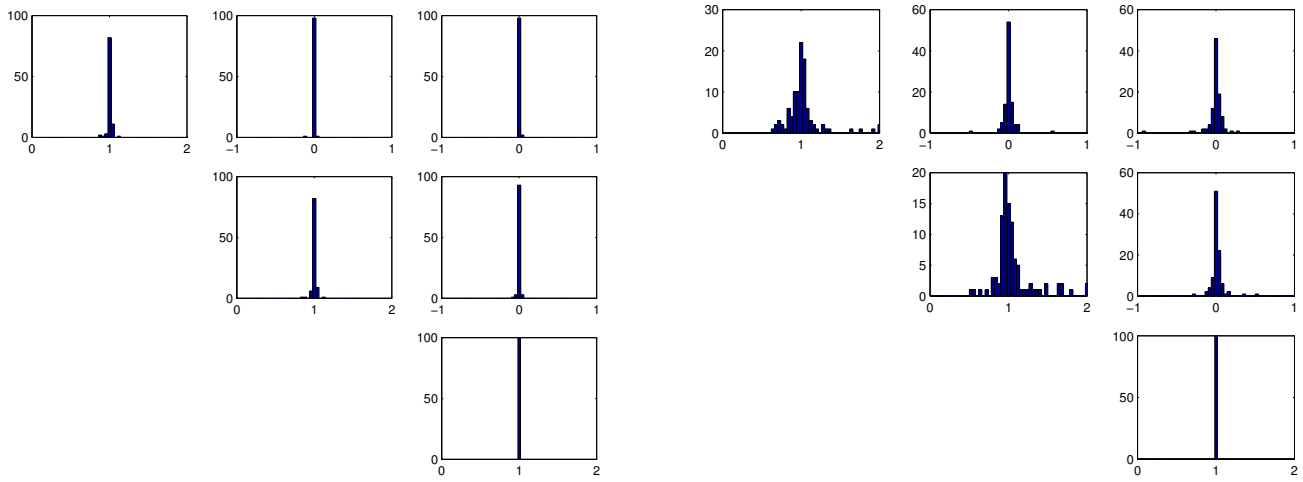


Fig. 5. Computed intrinsic camera parameters for low noise level (left) and high noise level (right).

Fig. 6. Images of the Arenberg castle which were used to generate the 3D model shown in *Fig. 7, 8, 9.*

Fig. 7. Orthographic views of the reconstruction. Notice parallelism and orthogonality.

Fig. 8. Perspective views of the 3D reconstruction obtained from the sequence seen in *Fig. 6.*

Fig. 9. More perspective views of the 3D reconstruction. Notice that once a 3D reconstruction is available it is easy to generate views which where not present in the original sequence.

	angle (±std.dev.)
parallel lines	1.8 ± 1.1 degrees
orthogonal lines	89.7 ± 1.4 degrees

TABLE I
RESULTS OF METRIC MEASUREMENTS ON THE RECONSTRUCTION.

respondence map obtained as in [16]. Remember that this was obtained without any calibration information. The position of the camera for the different views was also unknown and no knowledge about the scene was given to the algorithm. In Fig. 7 one can see 3 orthographic views of the reconstructed scene. Parallelism and orthogonality relations clearly have been retrieved. Look for example at the right angles in the top view or at the rectangular windows. Fig. 8 and 9 contain some perspective views of the reconstruction. Because at this point a dense correspondence map was only calculated for two images there are some inaccuracies left in the reconstruction. This also explains the fact that only a partial model is given here. In the future information from more images will not only be integrated in the calibration stage, but also in the reconstruction stage. A quantitative assessment of these properties can be made by explicitly measuring angles between lines on the object surface. For this experiment lines were manually selected that are aligned with windows and other prominent surface features. Several lines were identified for all of the three main directions. The lines along the same direction should be parallel to each other (angle between them should be 0 degree), while the lines corresponding to different directions 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 I).

A.2 Sequence 2

In Fig. 10 a part of another sequence is shown. This was filmed at the back of the castle. Here also the method proposed in this text was able to extract a metric reconstruction of the scene.

In Fig. 11 some perspective views of the reconstruction

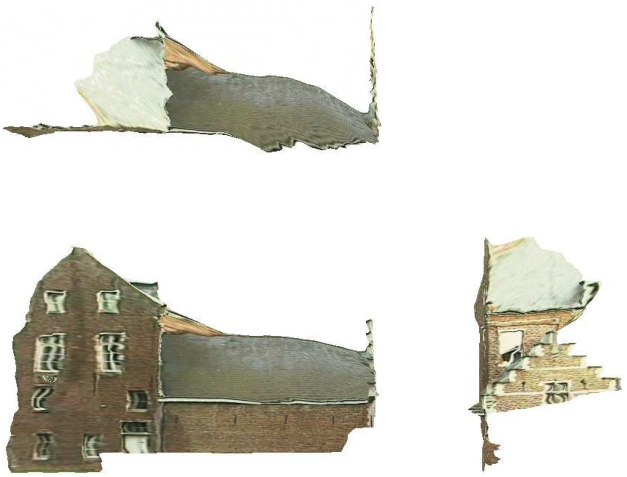


Fig. 12. Orthographic views of the 3D reconstruction.

are shown. To illustrate the metric quality of the reconstruction orthographic views of the reconstruction were computed. These are shown in Fig. 12. Note that this views are extreme views compared to the viewpoints of the sequence. Since dense reconstruction is not the main issue of the paper (i.e. it is only used to illustrated the metric calibration results), only a simple dense correspondence algorithm was used. The dent at the lower part of the roof is due to the alignment of epipolar lines with the gutter which causes an ambiguity for the matcher. Since the left part of the building is only visible in the distance the accuracy of that part of the façade is low. Although some artifacts of the dense reconstruction become clearly visible here the overall metric qualities are clearly recovered by our reconstruction scheme. Parallelism and orthogonality can be observed from the different orthographic views.

B. Experiments with known vanishing points

It was seen in section V-A that two images could be sufficient for self-calibration when two vanishing points corresponding to orthogonal directions can be identified in these images.

The first step is to obtain the weak calibration for the

Fig. 10. Images of another part of the Arenberg castle. Views from the 3D model generated from this sequence can be seen in *Fig. 11.*

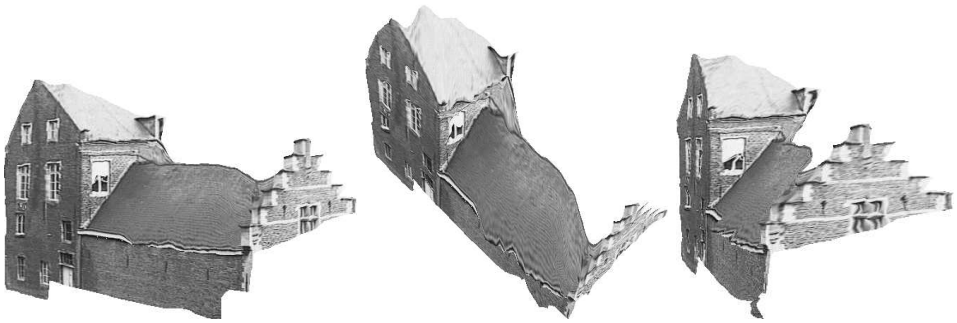


Fig. 11. Perspective views of the 3D reconstruction.

two images. This corresponds to identifying the fundamental matrix. From this a valid instance of the projective camera matrices can be chosen. The second step consists of identifying at least two vanishing points in both images. This was done using a cascaded Hough approach. A first Hough transform is used to identify lines in dual space, a second Hough transform is then used to identify coincidence points. In scenes containing regular structures (e.g. buildings) these points are typically vanishing points. More details of this approach can be found in [41].

Combining these results allows us to follow the steps described in section V-A:

- Affine calibration: (1) writing down the modulus constraint for the two images, (2) filling in the one parameter family of solutions for the infinity homography $\mathbf{H}_\infty$ defined by the two vanishing points and the projective calibration, (3) computing the solutions and extrema of the obtained equation, (4) selecting the best one according to equal moduli of eigenvalues.
- Metric calibration: (1) determining the one parameter family for the absolute conic corresponding with the infinity homography, (2) determining the solution for which the vanishing points correspond to orthogonal directions.

B.1 Sequence

Two images of the castle sequence were used in this experiment. These can be seen in Fig. 13. In addition to the projective calibration two vanishing points were automatically retrieved from these images. The coordinates of the computed vanishing points can be found in Table II. Using the described method the following intrinsic camera parameters were obtained (see Table III). These parameters are relatively close to the parameters obtained with the general method on the whole sequence. Except for the skew which is relatively important (10% of focal length). These are typical results although for some image pairs less accurate results or no results were obtained. This could be expected from the simulations (see Paragraph V-A.4) which showed an important noise sensitivity for this method. For more accurate results a bundle adjustment should be applied on the projective reconstruction and a more accurate localization method should be used for the vanishing points (the main goal of the method that was used [41] was detection and not accurate localization).

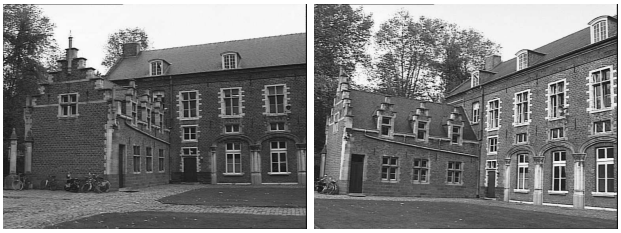


Fig. 13. Images 2 and 24 of the castle sequence.

im 2 (H)	im 2 (V)	im 24 (H)	im 24 (V)
-2490	-1366	-209	1275
37	-41954	437	-10582

TABLE II

COORDINATES OF THE VANISHING POINTS IN IMAGES 2 AND 24.

$$\mathbf{K} = \begin{bmatrix} 875 & 98 & 381 \\ & 880 & 197 \\ & & 1 \end{bmatrix}$$

TABLE III

INTRINSIC CALIBRATION PARAMETERS OBTAINED FROM IMAGE 2 AND 24.

Fig. 14. The 3 images that were used to build a Euclidean reconstruction. The camera was translated between the first two views (the zoom was used to keep the size more or less constant). For the third image the camera was also rotated.

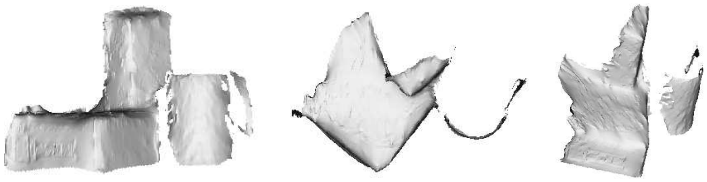


Fig. 15. Different views of the 3D reconstruction.

In some cases this method wasn't able to obtain any solution. These observations can be explained by the absence of redundancy when one tries to extract the metric calibration from a scene from two images only. The self-calibration problem is known to be a hard problem and therefore as much redundancy as possible should be used. Sometimes even this is not enough since many degenerate motion sequences exist (see [35] for an extensive discussion of this problem).

C. Experiments with a varying focal length

In this case an indoor scene was used for our experiment. First the focus of expansion was identified by zooming. Then a pure translation was carried out which allowed to retrieve the affine structure. Finally an additional motion was used to get the metric structure of the scene.

C.1 Sequence

Here some results obtained from a scene consisting of two boxes and a cup are given. The images that were used can be seen in Fig. 14. The scene was chosen to allow a good qualitative evaluation of the metric reconstruction. The boxes have right angles and the cup is cylindrical. These characteristics must be preserved by a *metric* reconstruction, but will in general not be preserved by an *affine* or *projective* reconstruction.

First a corner matcher was used to extract point correspondences between the three images. From these the method allowing varying focal lengths was used to obtain a metric calibration of the cameras. Subsequently, the algorithm of [31] was used to compute dense point correspondences. These were used to build the final 3D reconstruction using the previously recovered calibration.

Fig. 15 shows three views of the reconstructed scene. The left image is a front view, the middle image a top view, while the right image is a side view. Note especially from the top view, that 90^o angles are preserved and that the cup keeps its cylindrical form which is an indication of the quality of the metric reconstruction.

VII. Conclusion

This paper discusses the modulus constraint and its application to self-calibration problems. It is shown that this versatile constraint can be used to obtain the self-calibration of cameras in different circumstances. First the classical self-calibration problem is solved. The modulus constraint allows to follow a stratified approach. The projective calibration is upgraded to affine by identifying the plane at infinity, then the absolute conic is located and a metric calibration is obtained. A nonlinear minimization step allows to refine this calibration.

The experiments on synthetic data show that a stratified approach can be more successful in obtaining a satisfying calibration than an approach based on an a priori guess for the intrinsic camera parameters. These experiments confirm the importance of a good localization of the plane at infinity for a successful calibration. Besides experiments on synthetic data some real image sequences were used to illustrate the feasibility of the method.

Some other applications of the modulus constraint were also proposed. In many circumstances two or more vanishing points are known or can be found in the images. In this case the modulus constraint can even result in self-calibration from two images only. Interesting results have also been obtained for a varying focal length. Once the affine calibration has been obtained (i.e. from a pure translation) the modulus constraint can be used to retrieve the focal length through a closed form equation.

Acknowledgments

Marc Pollefeys acknowledges a specialization grant from the Flemish Institute for Scientific Research in Industry (IWT). Financial support from the EU ESPRIT project LTR 23515 IMPROOFS and the Belgian IUAP4/24 'IMechS' project are also gratefully acknowledged. We thank Andrew Zisserman and his team in Oxford for their comments and for providing us with robust software for projective calibrations. We also thank Reinhard Koch, Marc Proesmans and Tinne Tuytelaars for their contributions to our work.

Appendix A Derivation of the modulus constraint

The roots of equation (11) must obey $|\lambda_1| = |\lambda_2| = |\lambda_3|$. In this appendix a necessary condition is derived. A third order polynomial can be written as follows.

$$l_3\lambda^3 + l_2\lambda^2 + l_1\lambda + l_0 = l_3(\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3) \quad (31)$$

From (31) the following relations follow:

$$\lambda_1 + \lambda_2 + \lambda_3 = -\frac{l_2}{l_3} \quad (32)$$

$$\lambda_1(\lambda_2 + \lambda_3) + \lambda_2\lambda_3 = \frac{l_1}{l_3} \quad (33)$$

$$\lambda_1\lambda_2\lambda_3 = -\frac{l_0}{l_3} \quad (34)$$

At least one of the roots must be real, therefore it can be assumed that λ_1 is real (λ_2 and λ_3 can then be either real or complex). If the roots have the same moduli the following equation must be satisfied.

$$\lambda_1^2 = \lambda_2\lambda_3 \tag{35}$$

Rewriting (33) using (32) and (35) yields

$$\lambda_1(-\frac{l_2}{l_3} - \lambda_1) + \lambda_1^2 = \frac{l_1}{l_3} \tag{36}$$

or

$$\lambda_1 = -\frac{l_1}{l_2} \tag{37}$$

substituting (35) in (34) implies

$$\lambda_1^3 = -\frac{l_0}{l_3} \tag{38}$$

Eliminating λ_1 from the (37) and (38) gives a necessary condition that is only depending on l_3, l_2, l_1, l_0.

$$l_3 l_1^3 = l_2^3 l_0 \tag{39}$$

Note that (37) and thus also (39) are only necessary conditions for the homography to have eigenvalues corresponding to a rotation matrix. These equations can also be satisfied for three real eigenvalues. When multiple solutions persists, solutions for which no two eigenvalues are conjugate can be ruled out.

Appendix B Expressions for l_3, l_2, l_1 and l_0

In this expressions for l_3, l_2, l_1, l_0 will be derived. They will be expressed in terms of a_1, a_2, a_3 and the projective calibration. Starting from (11) a similar but simpler equation can be derived which avoids the occurence of the matrix inversion.

$$\begin{aligned} \det(\mathbf{H}_{ij} - \lambda\mathbf{I}) &= \det(\mathbf{H}_{1j}\mathbf{H}_{1i}^{-1} - \lambda\mathbf{I}) \qquad (40)\\ &= \det(\mathbf{H}_{1i}^{-1})\det(\mathbf{H}_{1j} - \lambda\mathbf{H}_{1i}) = 0 \\ &\Updownarrow \\ \det(\mathbf{H}_{1j} - \lambda\mathbf{H}_{1i}) &= 0 \qquad (41)\end{aligned}$$

The following notations are used to simplify the expressions: $\mathbf{H}_{1l} = [\mathbf{h}_1\mathbf{h}_2\mathbf{h}_3]$, $\mathbf{H}_{1k} = [\mathbf{h}_1'\mathbf{h}_2'\mathbf{h}_3']$, $\mathbf{e}_{1l} = \mathbf{e}$ and $\mathbf{e}_{1k} = \mathbf{e}'$, $|\mathbf{H}|$ means the determinant of $\mathbf{H}$.

$$\begin{aligned} &\det(\mathbf{H} - \lambda\mathbf{H}') \\ &= |\mathbf{h}_1 - \lambda\mathbf{h}_1'\,\mathbf{h}_2 - \lambda\mathbf{h}_2'\,\mathbf{h}_3 - \lambda\mathbf{h}_3'| \\ &= |\mathbf{h}_1\,\mathbf{h}_2 - \lambda\mathbf{h}_2'\,\mathbf{h}_3 - \lambda\mathbf{h}_3'| \\ &\quad -\lambda|\mathbf{h}_1'\,\mathbf{h}_2 - \lambda\mathbf{h}_2'\,\mathbf{h}_3 - \lambda\mathbf{h}_3'| \\ &= |\mathbf{h}_1\,\mathbf{h}_2\,\mathbf{h}_3 - \lambda\mathbf{h}_3'| - \lambda|\mathbf{h}_1\,\mathbf{h}_2'\,\mathbf{h}_3 - \lambda\mathbf{h}_3'| \\ &\quad -\lambda|\mathbf{h}_1'\,\mathbf{h}_2\,\mathbf{h}_3 - \lambda\mathbf{h}_3'| + \lambda^2|\mathbf{h}_1'\,\mathbf{h}_2'\,\mathbf{h}_3 - \lambda\mathbf{h}_3'| \\ &= |\mathbf{h}_1\mathbf{h}_2\mathbf{h}_3| - \lambda|\mathbf{h}_1'\mathbf{h}_2\mathbf{h}_3| + \lambda^2|\mathbf{h}_1\mathbf{h}_2'\mathbf{h}_3'| - \lambda^3|\mathbf{h}_1'\mathbf{h}_2'\mathbf{h}_3'| \\ &\qquad - \lambda|\mathbf{h}_1\mathbf{h}_2'\mathbf{h}_3| + \lambda^2|\mathbf{h}_1'\mathbf{h}_2\mathbf{h}_3'| \\ &\qquad - \lambda|\mathbf{h}_1\mathbf{h}_2\mathbf{h}_3'| + \lambda^2|\mathbf{h}_1'\mathbf{h}_2'\mathbf{h}_3| \end{aligned} \tag{42}$$

In the above expressions $\mathbf{H} + \mathbf{e}[a_1a_2a_3]$ or a similar expression should be substituted to $[\mathbf{h}_1\mathbf{h}_2\mathbf{h}_3]$, $[\mathbf{h}_1'\mathbf{h}_2\mathbf{h}_3]$, ..., $|\mathbf{h}_1'\mathbf{h}_2'\mathbf{h}_3'|$. Therefore the determinant of $\mathbf{H} + \mathbf{e}[a_1a_2a_3]$ should also be factorized. The other determinants can be factorized in a similar way.

$$\begin{aligned} &\det(\mathbf{H} + \mathbf{e}[a_1a_2a_3]) \\ &= |\mathbf{h}_1 + a_1\mathbf{e}\,\mathbf{h}_2 + a_2\mathbf{e}\,\mathbf{h}_3 + a_3\mathbf{e}| \\ &= |\mathbf{h}_1\,\mathbf{h}_2 + a_2\mathbf{e}\,\mathbf{h}_3 + a_3\mathbf{e}| + a_1|\mathbf{e}\,\mathbf{h}_2 + a_2\mathbf{e}\,\mathbf{h}_3 + a_3\mathbf{e}| \\ &= |\mathbf{h}_1\,\mathbf{h}_2\,\mathbf{h}_3 + a_3\mathbf{e}| + a_1|\mathbf{e}\,\mathbf{h}_2\,\mathbf{h}_3 + a_3\mathbf{e}| \\ &\quad + a_2|\mathbf{h}_1\,\mathbf{e}\,\mathbf{h}_3 + a_3\mathbf{e}| + a_1a_2\underbrace{|\mathbf{e}\,\mathbf{e}\,\mathbf{h}_3 + a_3\mathbf{e}|}_{=0} \\ &= |\mathbf{h}_1\,\mathbf{h}_2\,\mathbf{h}_3| + a_1|\mathbf{e}\,\mathbf{h}_2\,\mathbf{h}_3| + a_2|\mathbf{h}_1\,\mathbf{e}\,\mathbf{h}_3| + a_3|\mathbf{h}_1\,\mathbf{h}_2\,\mathbf{e}| \end{aligned} \tag{43}$$

It follows from this expression that the coefficients of λ^0 and λ^3 of equation (42) are first order polynomials in a_1, a_2, a_3. For λ^1 and λ^2 the derivation is a bit more tedious.

For the third order terms we still have two equal columns (ex. $a_1'a_2a_3|\mathbf{e}'\mathbf{e}\mathbf{e}|$) which means that these determinant vanishes. Some second order terms of the factorization do not vanish at first sight. These are the terms were both $\mathbf{e}$ and $\mathbf{e}'$ appear in the determinants. They can be grouped in pairs (ex. coefficient of λ):

$$\begin{aligned} &a_1a_2(|\mathbf{e}\mathbf{e}'\mathbf{h}_3| + |\mathbf{e}'\mathbf{e}\mathbf{h}_3|) \\ &a_1a_3(|\mathbf{e}\mathbf{h}_2\mathbf{e}'| + |\mathbf{e}'\mathbf{h}_2\mathbf{e}|) \\ &a_2a_3(|\mathbf{h}_1\mathbf{e}\mathbf{e}'| + |\mathbf{h}_1\mathbf{e}'\mathbf{e}|) \end{aligned} \tag{44}$$

All these terms vanish because permutating 2 rows of a determinant changes the sign of that determinant ($|\mathbf{e}'\mathbf{e}\mathbf{h}_3| = -|\mathbf{e}\mathbf{e}'\mathbf{h}_3|$).

This finally yields the following expressions for the 1^{st} order terms:

$$\begin{aligned} &|\mathbf{h}_1'\,\mathbf{h}_2\,\mathbf{h}_3| + |\mathbf{h}_1\,\mathbf{h}_2'\,\mathbf{h}_3| + |\mathbf{h}_1\,\mathbf{h}_2\,\mathbf{h}_3'| \\ &+a_1(|\mathbf{e}'\,\mathbf{h}_2\,\mathbf{h}_3| + |\mathbf{e}\,\mathbf{h}_2'\,\mathbf{h}_3| + |\mathbf{e}\,\mathbf{h}_2\,\mathbf{h}_3'|) \\ &+a_2(|\mathbf{h}_1'\,\mathbf{e}\,\mathbf{h}_3| + |\mathbf{h}_1\,\mathbf{e}'\,\mathbf{h}_3| + |\mathbf{h}_1\,\mathbf{e}\,\mathbf{h}_3'|) \\ &+a_3(|\mathbf{h}_1'\,\mathbf{h}_2\,\mathbf{e}| + |\mathbf{h}_1\,\mathbf{h}_2'\,\mathbf{e}| + |\mathbf{h}_1\,\mathbf{h}_2\,\mathbf{e}'|) \end{aligned} \tag{45}$$

For the second order terms the accents should be inversed.

In conclusion the modulus constraint can be expressed as $l_3l_1^3 = l_2^3l_0$ with

$$\begin{aligned} l_3 =& -|\mathbf{h}_1'\,\mathbf{h}_2'\,\mathbf{h}_3'| - a_1|\mathbf{e}'\,\mathbf{h}_2'\,\mathbf{h}_3'| - a_2|\mathbf{h}_1'\,\mathbf{e}'\,\mathbf{h}_3'| - a_3|\mathbf{h}_1'\,\mathbf{h}_2'\,\mathbf{e}'| \\ l_2 =& \quad (|\mathbf{h}_1\mathbf{h}_2'\mathbf{h}_3'| + |\mathbf{h}_1'\mathbf{h}_2\mathbf{h}_3'| + |\mathbf{h}_1'\mathbf{h}_2'\mathbf{h}_3|) \\ &+a_1\,(|\mathbf{e}\,\mathbf{h}_2'\mathbf{h}_3'| + |\mathbf{e}'\mathbf{h}_2\mathbf{h}_3'| + |\mathbf{e}'\mathbf{h}_2'\mathbf{h}_3|) \\ &+a_2\,(|\mathbf{h}_1\mathbf{e}'\mathbf{h}_3'| + |\mathbf{h}_1'\mathbf{e}\,\mathbf{h}_3'| + |\mathbf{h}_1'\mathbf{e}'\mathbf{h}_3|) \\ &+a_3\,(|\mathbf{h}_1\mathbf{h}_2'\mathbf{e}'| + |\mathbf{h}_1'\mathbf{h}_2\mathbf{e}'| + |\mathbf{h}_1'\mathbf{h}_2'\mathbf{e}|) \\ l_1 =& -(|\mathbf{h}_1'\mathbf{h}_2\mathbf{h}_3| + |\mathbf{h}_1\mathbf{h}_2'\mathbf{h}_3| + |\mathbf{h}_1\mathbf{h}_2\mathbf{h}_3'|) \\ &-a_1\,(|\mathbf{e}'\mathbf{h}_2\mathbf{h}_3| + |\mathbf{e}\,\mathbf{h}_2'\mathbf{h}_3| + |\mathbf{e}\,\mathbf{h}_2\mathbf{h}_3'|) \\ &-a_2\,(|\mathbf{h}_1'\mathbf{e}\,\mathbf{h}_3| + |\mathbf{h}_1\mathbf{e}'\mathbf{h}_3| + |\mathbf{h}_1\mathbf{e}\,\mathbf{h}_3'|) \\ &-a_3\,(|\mathbf{h}_1'\mathbf{h}_2\mathbf{e}| + |\mathbf{h}_1\mathbf{h}_2'\mathbf{e}| + |\mathbf{h}_1\mathbf{h}_2\mathbf{e}'|) \\ l_0 =& \quad |\mathbf{h}_1\,\mathbf{h}_2\,\mathbf{h}_3| + a_1\,|\mathbf{e}\,\mathbf{h}_2\,\mathbf{h}_3| + a_2\,|\mathbf{h}_1\,\mathbf{e}\,\mathbf{h}_3| + a_3\,|\mathbf{h}_1\,\mathbf{h}_2\,\mathbf{e}| \quad . \end{aligned} \tag{46}$$

References

[1] M. Armstrong, A. Zisserman and P. Beardsley, "Euclidean structure from uncalibrated images", *Proc. British Machine Vision Conference*, 1994.

[2] P. Beardsley, P. Torr and A. Zisserman "3D Model Acquisition from Extended Image Sequences", in : B. Buxton and R. Cipolla (eds.), *Computer Vision - ECCV'96*, Lecture Notes in Computer Science, Vol. **1065**, Springer-Verlag, pp. 683-695, 1996.

[3] F. Devernay and O. Faugeras, "From projective to euclidean reconstruction", *Proc. 1997 Conference on Computer Vision and Pattern Recognition*, IEEE Computer Soc. Press, 1996.

[4] 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.

[5] O. Faugeras, "Stratification of three-dimensional vision: projective, affine, and metric representations", Journal of the Optical Society of America A, pp. 465–483, Vol. 12, No.3, March 1995.

[6] O. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig", *Computer Vision - ECCV'92*, Lecture Notes in Computer Science, Vol. **588**, Springer-Verlag, pp. 563-578, 1992.

[7] O. Faugeras, Q.-T. Luong and S. Maybank. "Camera self-calibration: Theory and experiments", *Computer Vision - ECCV'92*, Lecture Notes in Computer Science, Vol. **588**, Springer-Verlag, pp. 321-334, 1992.

[8] C. Harris and M. Stephens, "A combined corner and edge detector", *Proc. Fourth Alvey Vision Conference*, pp.1447-151, 1988.

[9] R. Hartley, "Estimation of relative camera positions for uncalibrated cameras", *Computer Vision - ECCV'92*, Lecture Notes in Computer Science, Vol. **588**, Springer-Verlag, pp. 579-587, 1992.

[10] R. Hartley, "Euclidean reconstruction from uncalibrated views", in : J.L. Mundy, A. Zisserman, and D. Forsyth (eds.), *Applications of Invariance in Computer Vision*, Lecture Notes in Computer Science, Vol. **825**, Springer, pp. 237-256, 1994.

[11] R. Hartley, "Self-calibration from multiple views with a rotating camera", In : J-O. Eklundh (ed.), *Computer Vision - ECCV'94*, Lecture Notes in Computer Science, Vol. **800-801**, Springer-Verlag, pp. 471-478, 1994.

[12] R. Hartley, "In defense of the eight-point algorithm". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(6):580-593, October 1997.

[13] A. Heyden, K. Åström, "Euclidean Reconstruction from Constant Intrinsic Parameters" *Proc. 13th International Conference on Pattern Recognition*, IEEE Computer Soc. Press, pp. 339-343, 1996.

[14] A. Heyden, K. Åström, "Euclidean Reconstruction from Image Sequences with Varying and Unknown Focal Length and Principal Point", *Proc. 1997 Conference on Computer Vision and Pattern Recognition*, IEEE Computer Soc. Press, pp. 438-443, 1997.

[15] R. Horaud and G. Csurka, "Self-Calibration and Euclidean Reconstruction Using Motions of a Stereo Rig", *Proc. Sixth International Conference on Computer Vision*, Narosa Publishing House, pp. 96-103, 1998.

[16] R. Koch, "3-D Surface Reconstruction from Stereoscopic Image Sequences", *Proc. Fifth International Conference on Computer Vision*, IEEE Computer Soc. Press, pp. 109-114, 1995.

[17] S. Laveau and O. Faugeras, "Oriented Projective Geometry for Computer Vision", *Computer Vision - ECCV'96*, Lecture Notes in Computer Science, Vol. **1064**, Springer-Verlag, pp. 147-156, 1996.

[18] Q.-T. Luong and O. Faugeras, "Self calibration of a moving camera from point correspondences and fundamental matrices", Intl. Journal of Computer Vision, 1996.

[19] Q.-T. Luong and T. Vieville, "Canonic representation for the geometry of multiple projective views", *Computer Vision - ECCV'94*, Lecture Notes in Computer Science, Vol. **800**, Springer-Verlag, pp. 589-600, 1994.

[20] G. McLean and D. Kotturi, "Vanishing point Detection by Line Clustering", *IEEE Trans.PAMI*, Vol.17, No. 11, Nov. 1995, pp.1090-1095.

[21] R. Mohr, F. Veillon and L. Quan, "Relative 3D reconstruction using multiple uncalibrated images", *Proc. Fifth International Conference on Computer Vision*, IEEE Computer Soc. Press, pp.543-548, 1993.

[22] R. Mohr, B. Boufama and P. Brand, "Accurate projective reconstruction", in : J.L. Mundy, A. Zisserman, and D. Forsyth (eds.), *Applications of Invariance in Computer Vision*, Lecture Notes in Computer Science, Vol. **825**, Springer-Verlag, pp. 297-316, 1994.

[23] T. Moons, L. Van Gool, M. Van Diest, and E. Pauwels, "Affine reconstruction from perspective image pairs", in : J.L. Mundy, A. Zisserman, and D. Forsyth (eds.), *Applications of Invariance in Computer Vision*, Lecture Notes in Computer Science, Vol. **825**, Springer-Verlag, pp. 297–316, 1994.

[24] A. Morgan, *Solving polynomial systems using continuation for engineering and scientific problems*, Prentice-Hall Englewood Cliffs (N.J.), 1987.

[25] M. Pollefeys, L. Van Gool, T. Moons. "Euclidean 3D reconstruction from stereo sequences with variable focal lengths", Recent Developments in Computer Vision, Lecture Notes in Computer Science, Vol.1035, Springer-Verlag, pp. 405-414, 1996.

[26] M. Pollefeys, L. Van Gool and M. Proesmans, "Euclidean 3D Reconstruction from Image Sequences with Variable Focal Lengths", in : B. Buxton and R. Cipolla (eds.), *Computer Vision - ECCV'96*, Lecture Notes in Computer Science, Vol. **1064**, Springer-Verlag, pp. 31-42, 1996.

[27] M. Pollefeys, L. Van Gool and A. Oosterlinck, "The Modulus Constraint: A New Constraint for Self-Calibration", *Proc. 13th International Conference on Pattern Recognition*, IEEE Computer Soc. Press, pp. 349-353, 1996.

[28] M. Pollefeys and L. Van Gool, "A stratified approach to self-calibration", *Proc. 1997 Conference on Computer Vision and Pattern Recognition*, IEEE Computer Soc. Press, pp. 407-412, 1997.

[29] M. Pollefeys and L. Van Gool, "Self-calibration from the absolute conic on the plane at infinity", *Proc. CAIP97*, Lecture Notes in Computer Science, Vol. **1296**, Springer-Verlag, pp. 175-182, 1997.

[30] M. Pollefeys, R. Koch and L. Van Gool, "Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters", *Proc. Sixth International Conference on Computer Vision*, Narosa Publishing House, pp. 90-95, 1998.

[31] M. Proesmans, L. Van Gool and A. Oosterlinck, "Determination of optical flow and its discontinuities using non-linear diffusion", *Computer Vision - ECCV'94*, Lecture Notes in Computer Science, Vol. **801**, Springer-Verlag, pp. 295-304, 1994.

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

[33] P. Sturm, "Vision 3D non-calibrée: contributions à la reconstruction projective et études des mouvements critiques pour l'auto-calibrage", Ph.D. Thesis, INP de Grenoble, France , 1997.

[34] P. Sturm and L. Quang, "Affine stereo calibration", In : V. Hlavac and R. Sara (eds.), *Proceedings Computer Analysis of Images and Patterns*, Lecture Notes in Computer Science, Vol. **970**, Springer-Verlag, pp. 838-843, 1995.

[35] P. Sturm, "Critical Motion Sequences for Monocular Self-Calibration and Uncalibrated Euclidean Reconstruction", *Proc. 1997 Conference on Computer Vision and Pattern Recognition*, IEEE Computer Soc. Press, pp. 1100-1105, 1997.

[36] P. Sturm and B. Triggs. "A factorization based algorithm for multi-image projective structure and motion". in : B. Buxton and R. Cipolla (eds.), *Computer Vision - ECCV'96*, Lecture Notes in Computer Science, Vol. **1064**, Springer-Verlag, pp. 709-720, 1996.

[37] R. Szeliski and S. Kang, "Recovering 3D shape and motion from image streams using non-linear least-squares", DEC technical report 93/3, DEC, 1993.

[38] B. Triggs, "The Absolute Quadric", *Proc. 1997 Conference on Computer Vision and Pattern Recognition*, IEEE Computer Soc. Press, Los Alamitos, Calif., 1997, pp. 609-614.

[39] P. Torr, P. Beardsley and D. Murray, "Robust Vision", in *Proc. British Machine Vision Conference*, 1994.

[40] P. Torr, *Motion Segmentation and Outlier Detection*, PhD Thesis, Dept. of Engineering Science, University of Oxford, 1995.

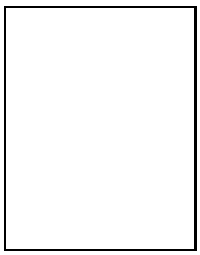
[41] T. Tuytelaars, M. Proesmans, L. Van Gool, "The cascaded Hough transform as support for grouping and finding vanishing points and lines", In *Proceedings international workshop on Algebraic Frames for Perception-Action Cycle*, Lecture Notes in Computer Science, vol. **1315**, Springer-Verlag, pp. 278-289, 1997.

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

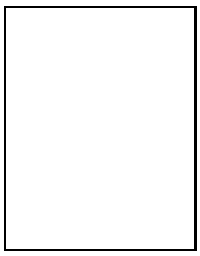
[43] Z. Zhang, R. Deriche, O. Faugeras and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the re-

covery of the unknown epipolar geometry", *Artificial Intelligence Journal*, Vol.78, pp.87-119, October 1995.

[44] A. Zisserman, P. Beardsley and I. Reid, "Metric calibration of a stereo rig", in *Proceedings IEEE Workshop on Representation of Visual Scenes*, Cambridge, 1995, pp. 93-100.



Marc Pollefeys obtained a Master of Electrical Engineering and his PhD from the Katholieke Universiteit Leuven in 1994 and 1999 respectively. His main area of research is computer vision, more precisely self-calibration and metric 3D reconstruction from uncalibrated image sequences. In 1998 he received the David Marr Prize together with Reinhard Koch and Luc Van Gool. He is currently working at the Department of Electrical Engineering at the K.U.Leuven.



Luc Van Gool got a Master of Electricomechanical Engineering and a PhD at the Katholieke Universiteit Leuven in 1981 and 1991 respectively. Currently, he is professor at the Katholieke Universiteit Leuven in Belgium and the ETH in Zurich, Switzerland. He leads computer vision research at both sites, where he also teaches computer vision. He has authored over 100 papers in this field. He has been a programme committee member of several, major vision conferences. His main interests include 3D reconstruction and modeling, object recognition, grouping, and non-linear diffusion for low-level vision. In 1998 he received the David Marr Prize together with Marc Pollefeys and Reinhard Koch, the Flemish Tech-Art Prize and an EITC Prize from the EC. He was the coordinator of several European projects. He is a co-founder of the company Eyetronics.