# ORIENTATION AND AUTO-CALIBRATION OF IMAGE TRIPLETS AND SEQUENCES

Xiangyang Hao[a], Helmut Mayer[b]

[a]Zhengzhou Institute of Surveying and Mapping, P.R. China – hxy@chxy.com
[b]Institute for Photogrammetry and Cartography, Bundeswehr University Munich, Germany – Helmut.Mayer@UniBw-Muenchen.de

**KEY WORDS:** Auto Calibration, Orientation, Robust, Trifocal Tensor, Visualization.

**ABSTRACT**

The automatic orientation and auto-calibration of image triplets and sequences is the basis for applications in visualization but also for all tasks employing metric information from imagery. Here we propose a hierarchical approach for orientation based on image triplets which is robust in the sense that it works with one set of parameters for a larger number of different sequences, also containing medium- to wide-baseline images, which standard procedures cannot handle. For auto-calibration we have made tests using the absolute dual quadric and the stratified approach based on the modulus constraint, showing that our results correspond with given calibration data, but are not yet totally stable. Finally, we propose a simple and robust method which allows the determination of the two parameters of the principal distance in x- and y-direction for image triplets robustly and reliably.

## 1 INTRODUCTION

As exemplified for instance by (El-Hakim, 2002), visualization of real world scenes becomes more and more sophisticated. As basis for visualization often sequences of images are taken from unknown view points. The tedious manual determination of the orientation can be avoided by automatic matching of points. If more than two images are acquired with fixed parameters of the cameras, it is at least for a general configuration theoretically possible to unambiguously upgrade the projective space defined by the perspective images to a metric space where right angles are defined and where one can measure relative distances. This is called auto-calibration.

In this paper we first summarize notations in Section 2. In Section 3 a robust approach for the automatic (projective) orientation of image triplets is proposed. The orientation builds on the estimation of the trifocal tensor based on the Carlsson-Weinshall duality and hierarchical matching including tracking through the pyramid. Section 4 details the orientation of image sequences by linking triplet by triplet. Old points are tracked and new points are added. Triplet and sequence orientation are robust in that sense that for a larger number of image triplets and sequences reliable results are obtained with all parameters fixed including the initial search-space set to the full image size. In Section 5 three approaches for (linear) auto-calibration are presented. Two are based on the absolute dual quadric and the third employs the stratified approach based on the modulus constraint. After giving results for these three approaches, we propose a simple, but robust approach for the determination of the principal distance in x- and y-direction for image triplets in Section 6. The paper ends with conclusions.

## 2 CALIBRATION MATRIX

We use homogeneous coordinates which are derived from Euclidean, i.e., metric, coordinates by adding an additional coordinate and free scaling. In our notation we distinguish homogeneous 2D and 3D vectors $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{X} = (X_1, X_2, X_3, X_4)$, respectively, as well as matrices $\mathbf{P}$ (bold), from Euclidean vectors $\boldsymbol{x}$ and $\boldsymbol{X}$ as well as matrices $\boldsymbol{R}$ (bold italics).

We model the interior orientation of a camera by principal distances $\alpha_x$ and $\alpha_y$, principal point $(x_0, y_0)$, and skew of the axes $s$. These 5 parameters are collected in the calibration matrix

$$\boldsymbol{K} = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} . \quad (1)$$

Homogeneous 3D points $\mathbf{X}$ are mapped to image points $\mathbf{x}$ via $\mathbf{x} = \mathbf{P}\,\mathbf{X}$. The $3 \times 4$ matrix $\mathbf{P}$ is constructed from a $3 \times 3$ rotation matrix $\boldsymbol{R}$ and the vector $\boldsymbol{t}$ representing the coordinates of the camera in the global coordinate frame by $\mathbf{P} = \boldsymbol{K}\,[\boldsymbol{R} \mid \mathbf{t}]$. Where not stated otherwise, the basic algorithms stem from (Hartley and Zisserman, 2000).

## 3 ROBUST ORIENTATION OF IMAGE TRIPLETS

### 3.1 Estimation of the Trifocal Tensor

While other approaches such as (Pollefeys et al., 2002) use fundamental matrices $\mathbf{F}$ and homographies between image pairs, the later allowing to cope with planar scenes, our basic building block for the orientation of an image sequence is the trifocal tensor $\mathcal{T}$. Its basic advantage is, that it renders it possible to linearly transfer points from two images into a third. This allows to check a match in two images in the third image and therefore helps to rule out blunders. This is not possible for fundamental matrices for which the result of a transfer is one-dimensional, i.e., the epipolar lines.

To estimate the trifocal tensor from a minimum of six point triplets, we employ the Carlsson-Weinshall duality (Carlsson, 1995, Weinshall et al., 1995). Utilizing an algorithm which gives a solution for a minimum number of points is important in two ways: First, for robust estimation based, e.g., on RANSAC (cf. below), this considerably reduces the search space. Second, by taking the minimum number of points we implicitly take the constraints for a tensor to be a trifocal tensor into account.

The basic idea of the duality is to interchange the roles of points being viewed by several cameras and the projection centers. If one has an algorithm for $n$ views and $m + 4$ points, then there is an algorithm for projective reconstruction from $m$ views of $n + 4$ points. By taking into account $|\mathbf{F}| = 0$, an algorithm can be constructed for the reconstruction of the fundamental matrix from two images for which seven homologous points suffice. This means that $n = 2$ and $m = 3$ and therefore, the duality

results into an algorithm for 3 views and 6 points. To determine the fundamental matrix from seven points, a cubic polynomial for which either one or three real solutions exist has to be solved.

Even though one can reduce mismatches by hierarchical matching (cf. Section 3.2), there are usually far too many for an efficient least squares solution, if the knowledge about the scene and the orientation of the cameras is weak. As our problem is of the type that we only have relatively few parameters and a high redundancy, RANSAC (random sample consensus) (Fischler and Bolles, 1981) is a good choice. RANSAC is based on the idea to select randomly minimum sets of observations. The correctness of a set is evaluated by the number of other observations which confirm it.

As there exist $6^3$ combinations for six triplets, which is considerably more than the $7^2$ for seven pairs, we first calculate the correspondences based on the fundamental matrices of the images one and two as well as one and three and only then match the triplets. For RANSAC we take into account the findings of (Tordoff and Murray, 2002). They state, that the procedure usually used to determine adaptively the number of samples for RANSAC is usually employed in a way neglecting statistical correlations. This leads to a much too low number of samples. Even though the correct solution would be to model the correlations, we have, as proposed in (Tordoff and Murray, 2002), fixed the problem by multiplying the number of samples with a larger factor (we use 500 for the fundamental matrix and 50 for the trifocal tensor), which leads to satisfying results.

### 3.2 Hierarchical Matching

We significantly reduce the search space by means of a hierarchical approach based on image pyramids with a reduction by a factor 2 for each level. With this not only the efficiency, but also the robustness is improved considerably.

Highly precise conjugate points are obtained from a least-squares matching of points obtained from the sub-pixel Förstner operator (Förstner and Gülch, 1987). On the highest level of the pyramids, which consists of about $100 \times 100$ pixels, no reduction of the search space, e.g., by means of epipolar lines, is yet available. To reduce the complexity of the matching, several measures are taken. First, before the actual least-square matching we sort out many points and calculate a first approximation by thresholding and maximizing, respectively, the correlation score among image windows. What is more, we restrict ourselves in the first image to only a few hundred points by regional non-maximum suppression.

Because of the complexity issues detailed in the last section, we compute on the highest pyramid level fundamental matrices and from them the epipolar lines from the first to the second and to the third image. After obtaining a solution for the trifocal tensor on the second or (seldom) third highest level of the pyramid, we have found that it suffices to track the points through the image pyramid. For each level, we scale the point coordinates by a factor of two and then match the point by least-squares matching sub-pixel precisely. This was found to be much faster and equally reliable than extracting points and matching them on each level. The reliability issue was found to be valid even if one takes into account the information from the levels above in the form of the epipolar lines and if one uses the point prediction with the trifocal tensor.

### 3.3 Robust Projective Bundle Adjustment

The linear solutions for the image pair or triplet presented above have the advantage that there is no need for approximate values. Though, the linear solution is algebraic and not geometric and thus the precision is limited. To obtain a highly precise solution, we compute a (projective) bundle adjustment. For this we first calculate (projective) 3D points with a linear algorithm. The bundle adjustment is split into the interleaved optimization of the 3D points and the projection matrices. For the actual optimization the Levenberg-Marquardt algorithm as implemented in the MINPACK public domain package is used.

Even though RANSAC together with other measures more or less guarantees, that the solution is valid, there is still a larger number of blunders in the data which distort the result. To get rid of them, we have implemented a simple scheme which eliminates the observations with the largest residuals as long they are $n$ times larger than the average standard deviation of the observations $\sigma_0 = v^\mathsf{T} v / redundancy$, with $v$ the residuals for the observations and all observations are weighted equally. The redundancy is $3 * \text{number\_points} - 24$ for the image triplet. Every 3D point is determined in three images ($3 * 2 - 3$) and one has to determine $2 * 12$ parameters for two projection matrices. We have found that a factor $n$ of 5 to 8 times $\sigma_0$ leads to reasonable results. This is in accordance with values derived from robust statistics.

The approach was implemented in C++ making use of the commercial image processing package HALCON (www.mvtec.com) and the public domain linear algebra package LAPACK interfaced by the template numerical toolkit (TNT; math.nist.gov/tnt).

Figure 1 shows an example for the orientation of a triplet. The dataset was taken from (Van Gool et al., 2002) as an example for a wide baseline triplet which cannot be oriented by the usual image sequence programs. Our program is not only able to orient this triplet, but as we use the full image as search space it is possible to do this with one and the same set of parameters for a wide range of imagery. This is also true for the image sequence presented below.

## 4 SEQUENCE ORIENTATION

### 4.1 Concatenation of Triplets

For the orientation of a sequence we start by orienting the first triplet as detailed in the previous section. Then we take the next, i.e., fourth image and orient the triplet consisting of the second, the third, and the fourth image. The orientation of a triplet results into three projection matrices $\mathbf{P}^i$, $\mathbf{P}^{i+1}$, $\mathbf{P}^{i+2}$ with the canonical matrix $\mathbf{P}^i = [\mathbf{I} \,|\, 0]$, constructing their own projective coordinate frame. For the orientation of the sequence it is necessary to transform the frames of the triplets into a global reference frame. It is, as in most approaches, also here defined by the first triplet. One way to do this transformation is to determine a $4 \times 4$ projection matrix mapping the coordinate frames.

Here, we take another direction. As we are heading towards a (projective) bundle adjustment based on points, we base the construction of the homogeneous coordinate frame on the points. First, the points in the second and third image of the preceding triplet, which are the same as the first and second image of the new triplet, are transfered via the trifocal tensor for the new triplet into the third image. In this image the coordinates for the point are determined via least-squares matching. To transfer the coordinate frame, we use the direct linear transform (DLT) algorithm. We already have (projective) 3D points and we have new corresponding 2D image points and this enables us to linearly compute the projection matrix for the new image. The same procedure is done for the rest of the triplets and step by step n-fold points are generated.

a b c

Figure 1: Wide baseline triplet from (Van Gool et al., 2002) with matched points and epipolar lines from first to second and third image. $\sigma_0$ of bundle adjustment was $0.044$ pixels.

### 4.2 Addition of Points

The procedure in the last section does not account for the fact that usually only a certain overlap exists between images. Thus, it is necessary to add points which have not been visible before. It is beneficial, to have points well distributed over the scene. Because the geometry of the scene is not known in advance, this can only be simulated by using points well distributed over the image. As shown above, we use non-maximum suppression to control the point distribution in the first image of the triplet.

When adding a new triplet, we want to avoid a lot of close-by points. We use the fact that we already know the orientation of the images and project all 3D points of the sequence via the projection matrices into the first two images of the new triplet. We then take only those image points in the first two images derived for the triplet, which do not lie in the regions defined by dilating the projected 3D points with, e.g., a radius of 3 pixels. For the thus determined points we linearly compute the 3D (projective) coordinates and we are ready for bundle adjustment.

### 4.3 Bundle Adjustment and Tracking

Finally, we use Levenberg Marquardt non-linear optimization as presented above in Subsection 3.3 to obtain a globally optimal (projective) solution. For the sequence we take into account the radial distortion. It is modeled by $\mathcal{R}([\text{x y } 1]^\mathsf{T}) \sim [\text{x y w}]^\mathsf{T}$ with $w^{-1} = (1 + k_1 r^2 + k_2 r^4)$ and $r^2 = x^2 + y^2$.

Also the adjustment for the sequence is done robustly. Here, the redundancy is $2 * \text{number\_points\_in\_all\_images} - (\text{number\_images} - 1) * 12 - \text{number\_3D\_points} * 3 - 2$. The last "2" is for the two parameters of the radial distortion.

To obtain points on the original resolution, we track the points through the image. For the sequence it is possible to find points in images where they have not been found before. We do this by projecting all adjusted 3D points of the sequence via the adjusted projection matrices into the images. The least-squares matching is done pairwise taking the image of the sequence as master image where the point is closest to the center of the image. The rational behind this is that usually the scene is acquired in a way that the image plane is approximately parallel to the scene's major structures for the center of the image.

The results for the six images in Figure 2, comprises 2 3-fold, 12 4-fold, 51 5-fold, and 98 6-fold points after robust estimation, with $\sigma_0 = 0.13$ pixels and $k_1 = 0.000145 \pm 0.0000570$ for ten runs. The latter means, that the radial distortion cannot be reliably determined. Therefore, we did not determine the even more unstable $k_2$. The points are well distributed over the scene and are a good basis for the auto-calibration presented in the next section.

In terms of speed it should be noted, that most of the time is spent in reliably determining the trifocal tensor and also in the outlier elimination in the robust bundle adjustment. The tracking of points is relatively fast even for larger images. For six images from the Rollei D7 metric fix-focus camera of size $2552 \times 1920$ the tracking of more than 200 points starting from $319 \times 240$ images is a matter of a few seconds. The processing of the whole sequence took 147 seconds on average on a 2.5 GHz computer.

## 5 LINEAR SEQUENCE CALIBRATION

### 5.1 Absolute Conic and Absolute Dual Quadric

For the determination of the calibration matrix, which allows to upgrade an affine reconstruction to metric, i.e., Euclidean, the plane at infinity $\pi_\infty$ has to determined. It is a fixed plane under any affine transformation. In affine and metric 3-space it has the canonical position $\pi_\infty = [0, 0, 0, 1]^\mathsf{T}$.

When the calibration matrix is known, a projective reconstruction $\{\mathbf{P}^i, \mathbf{X}_j\}$ with $\mathbf{P}^1 = [\mathbf{I} \,|\, 0]$ can be transformed into a metric reconstruction by a matrix H (the 3-vector $\mathbf{p}$ represents $\pi_\infty$):

$$\mathrm{H} = \begin{bmatrix} \boldsymbol{K} & 0 \\ -\mathrm{p}^\mathsf{T}\boldsymbol{K} & 1 \end{bmatrix} \quad . \tag{2}$$

The absolute conic $\Omega_\infty$, which we use for calibration, is a conic on the plane at infinity $\pi_\infty$. In a metric frame the points on the absolute conic $\Omega_\infty$ satisfy $X_1^2 + X_2^2 + X_3^2 = 0$ and $X_4 = 0$. $\Omega_\infty$ is a fixed conic under any similarity transformation and it is a geometric representation of the five degrees of freedom (DOF) of the calibration matrix.

The (degenerate) dual of the absolute conic $\Omega_\infty$ is a degenerate dual quadric in 3-space called the absolute dual quadric $\mathbf{Q}^*_\infty$. Geometrically the latter consists of the planes tangent to $\Omega_\infty$. It is a geometric representation of the eight DOF that are required to specify metric properties in the projective coordinate frame. $\pi_\infty$ is the null-vector of $\mathbf{Q}^*_\infty$, i.e., $\mathbf{Q}^*_\infty \pi_\infty = 0$. The image of the absolute conic (IAC) is the conic $\omega = (\boldsymbol{K}\boldsymbol{K}^\mathsf{T})^{-1}$. Like $\Omega$ it is an imaginary point conic, i.e., it has no real points. Its dual (DIAC) is defined as $\omega^* = \omega^{-1} = \boldsymbol{K}\boldsymbol{K}^\mathsf{T}$ and is the image of $\mathbf{Q}^*$. Once $\omega$

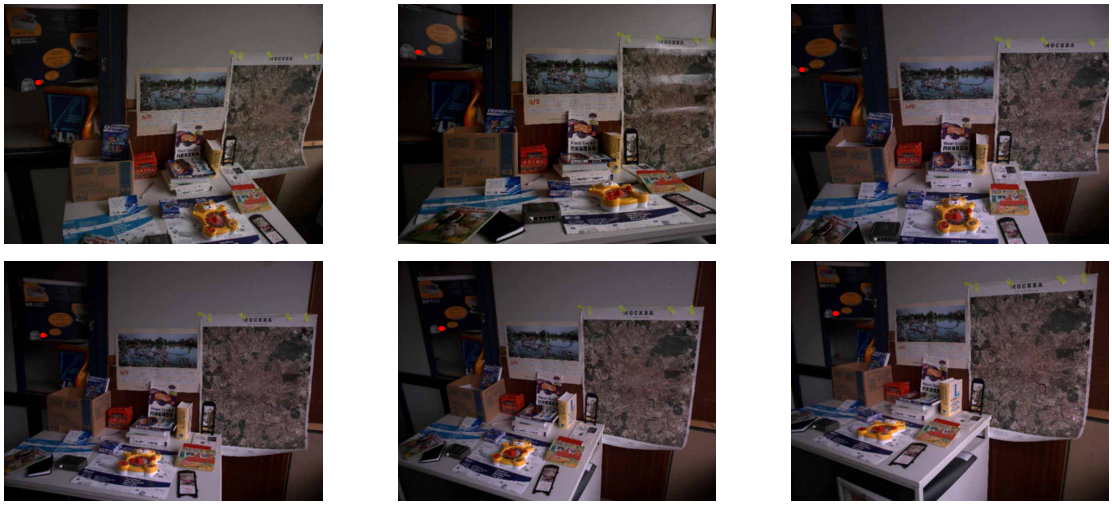


Figure 2: Six convergent images and matched points. 1 3-fold, 6 4-fold, 8 5-fold, and 166 6-fold points before and 2 3-fold, 12 4-fold, 51 5-fold, and 98 6-fold points after robust bundle adjustment ($\sigma_0 = 0.13$ pixels).

and $\omega^*$ have been estimated, also $\boldsymbol{K}$ is determined, as every symmetric matrix can be decomposed into an upper-triangular matrix and its transpose by means of Cholesky factorization.

Auto-calibration based on the DIAC employs

$$\omega^* = \mathbf{P}\mathbf{Q}^*_\infty\mathbf{P}^\mathsf{T} = \boldsymbol{K}\boldsymbol{K}^\mathsf{T} \quad , \tag{3}$$

i.e., the absolute dual quadric $\mathbf{Q}^*_\infty$ projects to the DIAC. Equation (3) is used to transfer a constraint on $\omega^*$ to a constraint on $\mathbf{Q}^*_\infty$ via the known projection matrices $\mathbf{P}^i$. Each element of $\omega^{*i} = \mathbf{P}^i\mathbf{Q}^*_\infty\mathbf{P}^{i\mathsf{T}}$ is linearly related to elements of $\mathbf{Q}^*_\infty$. More specific, linear or quadratic relationships between entries of $\omega^*$ generate linear or quadratic relationships between entries of $\mathbf{Q}^*_\infty$.

In the remainder of this paper it is assumed, that the internal parameters of the cameras are the same, i.e., $\omega^{*i} = \omega^{*j}$. From this follows $\mathbf{P}^i\ \mathbf{Q}^*_\infty\ \mathbf{P}^{i\mathsf{T}} = \mathbf{P}^j\ \mathbf{Q}^*_\infty\ \mathbf{P}^{j\mathsf{T}}$. Since the parameters are homogeneous, the equality holds only up to an unknown scale and a set of equations is generated:

$$\begin{aligned}\omega^{*i}_{11}/\omega^{*j}_{11} = \omega^{*i}_{12}/\omega^{*j}_{12} = \omega^{*i}_{13}/\omega^{*j}_{13} \quad = \\ \omega^{*i}_{22}/\omega^{*j}_{22} = \omega^{*i}_{23}/\omega^{*j}_{23} = \omega^{*i}_{33}/\omega^{*j}_{33}\end{aligned} \tag{4}$$

This set of equations corresponds to a set of quadratic equations in the entries of $\mathbf{Q}^*_\infty$. In the minimum case of three views, ten equations result which yield $\mathbf{Q}^*_\infty$.

### 5.2 Auto-Calibration Using the Absolute Dual Quadric

To actually determine $\boldsymbol{K}$, the problem is first transformed in a way making use of the symmetries of the matrices. Instead of $\omega^* = \mathbf{P}\mathbf{Q}^*_\infty\mathbf{P}^\mathsf{T}$ we write $\mathbf{w} = \mathbf{A}\mathbf{x}$, where $\mathbf{w}$ contains the six upper triangular entries of $\omega^*$, $\mathbf{x}$ the 10 elements of the upper triangular parts of $\mathbf{P}\mathbf{Q}^*_\infty$, and $\mathbf{A}$ is an $6 \times 10$ matrix comprising the corresponding elements of $\mathbf{P}\mathbf{P}^\mathsf{T}$. From $\mathbf{P}^1 = [\mathbf{I} \mid 0]$ one can see that $\mathbf{w} = [\mathrm{w}_1, \mathrm{w}_2, \mathrm{w}_3, \mathrm{w}_4, \mathrm{w}_5, \mathrm{w}_6]^\mathsf{T} = [\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \mathrm{x}_5, \mathrm{x}_6, \mathrm{x}_{10}]^\mathsf{T}$

Because $\omega^*$ is a homogeneous matrix, we can set $\mathrm{x}_{10} = \mathrm{w}_6 = \mathbf{Q}^*_{\infty 33} = \omega^*_{33} = 1$. By choosing $\mathbf{P}^1$ as reference for equation (4), five equations $\mathrm{f}_1$ to $\mathrm{f}_5$ of the form

$$\mathrm{f}_1 = \mathrm{x}_1 \sum_{k=1}^{10} \mathrm{A}_{2k}\mathrm{x}_k - \mathrm{x}_2 \sum_{k=1}^{10} \mathrm{A}_{1k}\mathrm{x}_k = 0 \tag{5}$$

arise. As they are non-linear, their linearization yields an equation system $\mathbf{B}\mathbf{x} - \mathbf{f} = \mathbf{v}$ with $\mathbf{B}$ a $5 \times 9$ matrix and $\mathbf{f}$ consisting of $\mathrm{f}_1$ to $\mathrm{f}_5$. Every projection matrix besides the first, canonical matrix adds five equations. If there are $m$, with $m \geq 3$ projection matrices, the number of equations is $5(m-1)$. By assuming approximate values one obtains values for the vectors $\mathbf{x}$ and $\mathbf{w}$ which gives us finally $\omega^*$ and via Cholesky decomposition $\boldsymbol{K}$.

Because the problem is non-linear, the solution strongly depends on the initial values, which should therefore be chosen carefully. As the principal point is for most cameras close to the center of the image and the skew $s$ can mostly be neglected, the following approximate values are usually a good choice: $\mathrm{x}_1 = 1, \mathrm{x}_5 = (w/h)^2, \mathrm{x}_2 = \mathrm{x}_3 = \mathrm{x}_4 = \mathrm{x}_6 = \mathrm{x}_7 = \mathrm{x}_8 = \mathrm{x}_9 = 0$, where $w$ and $h$ are the width and the height of the image, respectively.

For digital cameras one can, at least for the precision obtainable by the above method, safely assume, that the skew is zero. This gives the additional constraint $\omega^*_{12}\omega^*_{33} = \omega^*_{13}\omega^*_{23}$ among the elements of a single matrix $\omega^*$.

### 5.3 Stratified Auto-Calibration Using the Modulus Constraint

An alternative approach to the above one-step solution is to first obtain $\pi_\infty$, or alternatively an affine reconstruction, and only then upgrade it by the matrix $\boldsymbol{K}$ to a metric one. For the latter, a linear solution exists.

For general motions of the camera and constant internal parameters, an effective way to compute $\pi_\infty$ is the so-called modulus constraint (Pollefeys and Van Gool, 1999). It is a polynomial equation in the coordinates of $\pi_\infty$. A projection matrix $\mathbf{P}$ can be rewritten as $[\mathbf{A} \mid \mathbf{a}]$. The homography which describes the projection of points from $\pi_\infty$ to the image plane is $\mathbf{H}_\infty = \mathbf{A} - \mathbf{a}\mathbf{p}^\mathsf{T}$.

If we assume that the internal parameters are constant, we obtain making use of equation (2) for cameras $i$ and $j$ $\mathbf{H}^i_\infty = (\mathbf{A}^i - \mathbf{a}^i\mathbf{p}^\mathsf{T}) = \boldsymbol{K}\boldsymbol{R}^i\boldsymbol{K}^{-1}$ and $\mathbf{H}^j_\infty = (\mathbf{A}^j - \mathbf{a}^j\mathbf{p}^\mathsf{T}) = \boldsymbol{K}\boldsymbol{R}^j\boldsymbol{K}^{-1}$. The infinity homography from $i$ to $j$ can therefore be written as

$$\mathbf{H}^{ij}_\infty = (\mathbf{A}^i - \mathbf{a}^i\mathbf{p}^\mathsf{T})(\mathbf{A}^j - \mathbf{a}^j\mathbf{p}^\mathsf{T}) = \boldsymbol{K}\boldsymbol{R}^j(\boldsymbol{K}\boldsymbol{R}^i)^{-1} \quad . \tag{6}$$

This shows that $\mathbf{H}_\infty^{ij}$ is conjugated with a rotation matrix. Therefore, its three eigenvalues must have the same moduli. The characteristic polynomial of $\mathbf{H}_\infty^{ij}$ is $\det(\mathbf{H}_\infty^{ij} - \lambda\mathbf{I}) = f_3\lambda^3 + f_2\lambda^2 + f_1\lambda + f_0$, where $\lambda_i$ are the three eigenvalues, and $f_i$ are the four coefficients. It was shown that the following condition, called the modulus constraint, is a necessary condition for the roots of the eigenvalues to have equal moduli:

$$f_3 f_1^3 = f_0 f_2^3 \tag{7}$$

This equation yields a constraint on the three elements of the vector $\mathbf{p}$ by expressing $f_0$, $f_1$, $f_2$, and $f_3$ as a function of them. By factorization using the multi-linearity of determinants one finds that $f_0$, $f_1$, $f_2$, and $f_3$ are linear in the elements of $\mathbf{p}$. Putting things together one can see that the modulus constraint is a quartic polynomial in the three elements of $\mathbf{p}$. It is a necessary, but not a sufficient condition. Every pair of views generates a quartic equation. Thus, $\pi_\infty$ can be determined from three views, but only as the intersection of three quartics in three variables. There are overall 64 solutions to these equations. While (Pollefeys and Van Gool, 1999) use continuation, here the problem is again solved by least-squares adjustment of the linearized problem. The solution for this is very sensitive to the given initial values. To compensate for this, a large 3D-space of solutions is searched through.

Once $\mathbf{p}$, i.e., $\pi_\infty$, is known, an affine reconstruction is achieved and the mapping from $\pi_\infty$ is $\mathbf{H}_\infty^i = (\mathbf{A}^i - \mathbf{a}^i\mathbf{p}^\mathsf{T})$. The absolute conic lies on $\pi_\infty$, so its image is mapped between views by $\mathbf{H}_\infty^i$. If the internal parameters are constant over the views, the transformation rules for dual conics lead to $\omega^* = \mathbf{H}_\infty^i\omega^*\mathbf{H}_\infty^{i\mathsf{T}}$. Here, the scale factor can be chosen as unity by normalizing $\det(\mathbf{H}_\infty^i) = 1$. The above formula leads to six equations for the elements of the upper triangular matrix $\omega^*$. By combining equations from more than two views, a linear solution is obtained for $\omega^*$. $\boldsymbol{K}$ is again determined via Cholesky decomposition.

### 5.4 Results

For the sequence of six images presented in Figure 2 we did a calibration for ten projective reconstructions with the three calibration methods, two using the absolute dual quadric and one employing the stratified solution. The second method differs from the first by constraining the skew of the camera to be zero.

The camera used is a Rollei D7 metric camera with highly precisely known camera constant $c = 7.429\,mm$, principal point with $x_{pp} = 0.294\,mm$, $y_{pp} = -0.046\,mm$, and a sensor size of $8.932\,mm \times 6.720\,mm$. This means that $\alpha_x = 0.8317$, $\alpha_y = 1.1055$, $x_0 = 0.0329$ and $y_0 = -0.0095$, and no skew.

For the absolute dual quadric one obtains (six of ten runs were successful)

$$\begin{bmatrix} 0.821 \pm 0.0631 & -0.0069 \pm 0.00589 & 0.0329 \pm 0.0389 \\ 0 & 1.1 \pm 0.0718 & -0.0414 \pm 0.0441 \\ 0 & 0 & 1 \end{bmatrix}$$

With the constraint $s = 0$ the following result arises (five of ten runs were successful)

$$\begin{bmatrix} 0.8 \pm 0.056 & 0.00239 \pm 0.00101 & 0.0377 \pm 0.0382 \\ 0 & 1.07 \pm 0.067 & -0.0266 \pm 0.035 \\ 0 & 0 & 1 \end{bmatrix}$$

The modulus constraint gives the following results (as none of ten runs was successful for six images, the result for five images, where three of ten runs succeeded, are given):

$$\begin{bmatrix} 0.711 \pm 0.0264 & 0.0106 \pm 0.00874 & 0.092 \pm 0.0857 \\ 0 & 1 \pm 0.0495 & 0.00773 \pm 0.0733 \\ 0 & 0 & 1 \end{bmatrix}$$

For this sequence the methods based on the absolute dual quadric clearly outperform the stratified auto-calibration. The former two are close to the ground truth, with interestingly the version without the constraint on the skew performing better. Though, as will be shown in the next section, the better performance for the methods based on the absolute quadric is valid only in this case.

## 6 ROBUST CALIBRATION OF IMAGE TRIPLETS

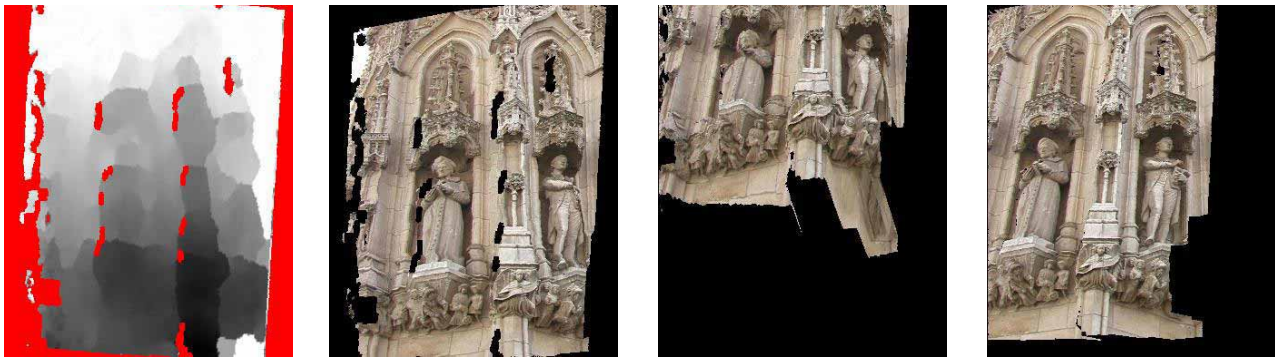
The above auto-calibration is state of the art, but has the disadvantage that we found our preliminary implementation to be unstable for image triplets. Here we present a rather simple, but robust means for the calibration of image triplets.

We again start based on the projective, robustly optimized orientation. Basically, we employ $\mathbf{P} = \boldsymbol{K}\,[\boldsymbol{R} \,|\, \mathbf{t}]$. Also in the calibrated case $\mathbf{P}^1 = [\mathbf{I} \,|\, 0]$ can be chosen. From the trifocal tensor the fundamental matrix $\mathbf{F}_{12}$ from image one to two can be obtained and from it the (calibrated) essential matrix is computed simply as $\boldsymbol{E}_{12} = \boldsymbol{K}^\mathsf{T}\mathbf{F}_{12}\boldsymbol{K}$. The projection matrix $\mathbf{P}^2$ for the second camera defining the metric frame is obtained via singular value decomposition (SVD) of $\boldsymbol{E}_{12} = \mathbf{U}\,diag(1,1,0)\mathbf{V}^\mathsf{T}$, with $diag(1,1,0)$ a diagonal $3 \times 3$ matrix. This leads to four solutions from which the one is chosen where the point is for both cameras in front of them. After defining the metric coordinate frame, 3D Euclidean points can be calculated and $\mathbf{P}^3$ can be determined linearly from the 3D points via DLT.

For the metric bundle adjustment there are in general six parameters to be optimized per projection matrix: three translations in vector $\mathbf{t}$ and three rotations represented implicitly in the matrix $\boldsymbol{R}$. To make the problem well-behaved, rotations are represented via quaternions, e.g., (Förstner, 1999). Because the relative orientation of the first pair is defined by five parameters, e.g., three rotations and two translations, there are eleven parameters to be optimized for the triplet.

The unconstrained optimization of the parameters of the calibration matrix leads to local maxima and thus to unsatisfactory results. Therefore, another way was chosen. We assume that the principal point is approximately in the center of the image and that the ratio principal distance in x- and y- direction is approximately the ratio of the width and the height of the image. We further assume that standard principal distances range from $0.5$ to $2.5$. Then the idea is to just sample the principal distance in x-direction, $\alpha_x$ logarithmically by $2.5 * 0.95^n$ with $0 \leq n \leq 30$ and taking the $\sigma_0$ of the of the least squares adjustment as criterion. For the $\alpha_x$ resulting in the lowest $\sigma_0$, $\alpha_y$ is varied starting from $1.15 * \alpha_x$ with $1.15 * \alpha_x * 0.98^n$ and $0 \leq n \leq 15$.

First of all, the approach gave a result for all runs of the experiments presented here, but also in all other experiments. For the first triplet of the sequence presented in Figure 2, $\alpha_x = 0.778 \pm 0.020$ and $\alpha_y = 1.079 \pm 0.033$ were obtained. This is in accordance with the given calibration data for the Rollei D7 metric camera as well as the results for the sequence of six images presented above. For the absolute dual quadric including the constraint, that the skew is zero, the result for three images is also reasonable, but without the constraint on the skew and for stratified auto-calibration the result is much worse.



a b c d

Figure 3: a) Disparity map (non-overlapping and occluded areas marked in red/gray) and b) - d) visualization based on calibration (occluded areas in black) for the cathedral example from (Van Gool et al., 2002)

For Figure 1 we do not have ground-truth for the calibration. We have obtained $\alpha_x = 2.37 \pm 0.11$ and $\alpha_y = 1.93 \pm 0.14$ for ten runs. Both methods based on the absolute dual quadric gave no result at all for the ten runs. The stratified auto-calibration gave a totally different result for two runs and for the rest $\alpha_x = 2.18 \pm 0.10$ and $\alpha_y = 1.91 \pm 0.06$ which is for $\alpha_y$ in good accordance, but for $\alpha_x$ there is a larger difference. Similar results were obtained also for a larger number of other image triplets. Figure 3 shows the visualization of the image triplet after auto-calibration. The disparity map in Figure 3 a has been computed based on a improved version of (Zitnick and Kanade, 2000).

## 7 CONCLUSIONS

We have shown methods for the orientation as well as for the auto-calibration of image triplets and sequences. The methods for (projective) orientation are robust in that sense that they generate results, which can be reproduced, with one set of parameters for a larger set of images. The cameras can have a considerably larger baseline than usual video sequences. The results for the auto-calibration of sequences were shown to be correct also in relation to known calibration data, but they are still preliminary in that sense, that they cannot be reliably reproduced. The methods based on the dual quadric perform better for the sequence, while the stratified auto-calibration gives reasonable results for some triplets, where the former methods fail.

For the triplet we have introduced a simple procedure, which yields acceptable results which can be robustly reproduced and are in accordance with given calibration data and the other approaches. Though, this procedure does not give the coordinates of the principal point and will very probably fail, if the principal point is farther away from the image center. On the other hand, this is no problem for most practical applications.

The next thing to be done is the (metric) bundle adjustment of the image sequence using the linearly obtained calibration matrix as a start value. Further ideas go into the direction to use the cheirality inequalities presented in (Hartley and Zisserman, 2000) to reduce the search space for the modulus constraint. And finally, it is probably very useful to implement the approach for constrained auto-calibration proposed by (Pollefeys et al., 2002).

## REFERENCES

Carlsson, S., 1995. Duality of Reconstruction and Positioning from Projective Views. In: IEEE Workshop on Representation of Visual Scenes, Boston, USA.

El-Hakim, S., 2002. Semi-Automatic 3D Reconstruction of Occluded and Unmarked Surfaces from Widely Separated Views. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. (34) 5, pp. 143–148.

Fischler, M. and Bolles, R., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the ACM 24(6), pp. 381–395.

Förstner, W., 1999. On Estimating Rotations. In: Festschrift für Prof. Dr.-Ing. Heinrich Ebner zum 60. Geburtstag, Lehrstuhl für Photogrammetrie und Fernerkundung der Technischen Universität München, pp. 85–96.

Förstner, W. and Gülch, E., 1987. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. In: ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, pp. 281–305.

Hartley, R. and Zisserman, A., 2000. Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, UK.

Pollefeys, M. and Van Gool, L., 1999. Stratified Self-Calibration with the Modulus Constraint. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(8), pp. 707–724.

Pollefeys, M., Verbiest, F. and Van Gool, L., 2002. Surviving Dominant Planes in Uncalibrated Structure and Motion Recovery. In: Seventh European Conference on Computer Vision, Vol. II, pp. 837–851.

Tordoff, B. and Murray, D., 2002. Guided Sampling and Consensus for Motion Estimation. In: Seventh European Conference on Computer Vision, Vol. I, pp. 82–96.

Van Gool, L., Tuytelaars, T., Ferrari, V., Strecha, C., Vanden Wyngaerd, J. and Vergauwen, M., 2002. 3D Modeling and Registration under Wide Baseline Conditions. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. (34) 3A, pp. 3–14.

Weinshall, D., Werman, M. and Shashua, A., 1995. Shape Descriptors: Bilinear, Trilinear, and Quadrilinear Relations for Multi-Point Geometry and Linear Projective Reconstruction. In: IEEE Workshop on Representation of Visual Scenes, Boston, USA, pp. 55–65.

Zitnick, C. and Kanade, T., 2000. A Cooperative Algorithm for Stereo Matching and Occlusion Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(7), pp. 675–684.