EXPERIMENTAL RESULTS ON THE DETERMINATION OF THE TRIFOCAL TENSOR USING NEARLY COPLANAR POINT CORRESPONDENCES

Camillo RESSL
Institute of Photogrammetry and Remote Sensing
University of Technology, Vienna, Austria
car@ipf.tuwien.ac.at

Working Group III/1

KEY WORDS: Orientation, Algorithms, Performance Test, Orientation, Calibration, Bundle block adjustment .

ABSTRACT

In this article we examine the computation of the trifocal tensor from different view points: the minimization of algebraic or reprojection error, the consideration of the internal constraints, and the effect of nearly coplanar object points. It is shown using synthetic data, that a correct solution for the trifocal tensor can be obtained as long as the object points deviate from a common plane by at least 1 % of the viewing distance. Using real data it is shown that the orientation parameters derived from the tensor can be successfully used to initialize a subsequent bundle block adjustment.

1 INTRODUCTION

Projective geometry is widely used in Computer Vision because it enables linear and simple representations for several orientation methods; e.g. the trifocal tensor for the relative orientation of three uncalibrated images [Hartley 1997].

These alternative representations based on projective geometry, however, are still not very popular in Photogrammetry, because there are several drawbacks associated with them:

- The linearity of these representations is achieved by **over-parameterization**, i.e. more parameters than the actual degrees of freedom (DOF) are used. Consequently certain non-linear constraints among the parameters must be satisfied; and the direct linear solution does not satisfy them in general.
- The linearity is further achieved by considering the images to be **uncalibrated**, i.e. the information of an a-priori given interior orientation can not be used directly (or the linearity gets lost).
- The linearity of these representations can be used effectively only if the so-called **algebraic error** is minimized instead of the so-called *reprojection error* (i.e. the error in the original image measurements).
- The solution of the respective linear system of equations for determining the alternative parameters fails if the image correspondences originate from **exactly planar object** points or lines.
- In projective geometry only linear mappings are dealt with. Consequently, unknown **non-linear image distortion** can not be handled directly in this framework; known non-linear image distortion could be removed from the images in a preprocessing step.

In contrast to all these drawbacks, however, we have the **linearity** of these alternative representations, which is a huge benefit, as it liberates us from the requirement to provide approximate values for the exterior (and sometimes also interior) orientation parameters, which are inevitable for a conventional bundle block adjustment.

With the benefit and drawbacks mentioned above, a reasonable **strategy** would consist of two steps: a) perform the orientation using a convenient alternative representation and b) to use the orientation parameters obtained thereby as approximate values for a subsequent bundle block adjustment to further refine the result by considering a-priori known interior orientation and modelling the non-linear image distortion.

To successfully apply this strategy it is essential, that the linear orientation in step a) does not fail. The main reason for the linear orientation to fail are coplanar object points. Due to the noise in the images this will not only happen for mathematically exact coplanar object points, but already for *nearly* coplanar points. The required deviation from a common plane is further increased by the other drawbacks; i.e. the negligence of the internal constraints, the unknown interior orientation and the unknown non-linear image distortion.

Therefore, it is interesting to investigate the effects of the drawbacks mentioned above on the linearly obtained orientation parameters in the case of nearly coplanar object points; e.g. what minimum deviation from a common plane is necessary for a successful solution if the internal constraints are considered or neglected and if algebraic error or reprojection error is minimized? For this investigation we will consider the trifocal tensor, which is made up of 27 elements (with 18 DOF) and linearly describes the relative orientation of three uncalibrated images. Compared with the other linear representations of two (the fundamental matrix with 9 elements and 7 DOF, [Luong and Faugeras 1996]) and four images (the quadfocal tensor with 81 elements and 29 DOF, [Hartley 1998]), the trifocal tensor is more robust than the fundamental matrix, due to the third image, and not as complex as the quadfocal tensor (9 vs. 52 internal constraints).

This article is structured in the following way. Section 2 gives an short overview on the properties of the trifocal tensor. In section 3 the results of synthetic experiments are presented, followed by an example using real data in section 4. The findings are summarized in section 5.

2 THE TRIFOCAL TENSOR

The trifocal tensor is made up of 27 homogenous elements, thus can be visualized as a $3 \times 3 \times 3$ cube of numbers. Slices in every direction of this cube return 3×3 matrices with special properties; e.g. [Ressl 2003]. These slices allow the determination of the six epipoles in the three images and the determination of the three respective fundamental matrices. In case of unknown interior orientation, the latter can be further used to derive a common interior orientation for the three images using the so-called *Kruppa equations*; e.g. [Hartley and Zisserman 2001]. Finally the fundamental matrices and the interior orientation can be used to obtain the projection centers and rotation matrices of the relative orientation of the three images; [Ressl 2003].

The trifocal tensor can be computed from corresponding points and/or lines across the three images. Each triple of points gives 4 independent homogenous equations, so-called *trilinearities*, and each triple of lines gives 2 independent equations - all being linear in the tensor's elements. Consequently, at least 7 point- or 13 line-correspondences, or a proper combination, are needed for the direct linear solution of the trifocal tensor minimizing algebraic error.

The relative orientation of three uncalibrated images has only 18 DOF. Consequently 9 constraints must be satisfied by the 27 tensor elements, one of the constraints is the fixing of the tensor's homogenous scale. Various sets of constraints were proposed in the past; see [Ressl 2003] for an overview.

For computing a valid trifocal tensor, which satisfies the constraints, preferably by minimizing reprojection error instead of algebraic error, we have to use the so-called *Gauss-Helmert model*, [Koch 1999], also called *general case of least squares adjustment.* This non-linear iterative method requires approximate values for the tensor elements, which could be obtained from the direct linear solution.

Note: In projective geometry every entity is represented as an homogenous vector, e.g. a 2D point $\mathbf{x}$ as $\mathbf{x} = (x, y, 1)^\top$. Now suppose the point $\mathbf{x}$ is measured in a digital image with 2000×3000 pixels and is located far away from the origin of the coordinate frame. In this case the coordinates x and y will be in the order of 1000, whereas the homogenous extension still is 1. This difference in order between the Euclidian and the homogenous part will cause enormous numerical problems if such projective points are used to compute other quantities; e.g. the trifocal tensor from several point correspondences. These problems can be avoided easily if the projective entities are shifted and scaled prior to the computations. This procedure is due to Hartley, who used this for computing the fundamental matrix; [Hartley 1995]. He proposes to translate the set of image points in the way that their centroid $\boldsymbol{x}_C$ is moved to the origin and then to scale the translated points isotropically by $m = \sqrt{2}/s$, where s is the average distance of the points from $\boldsymbol{x}_C$.

3 EXPERIMENTAL RESULTS FROM SYNTHETIC DATA

In [Ressl 2003] the trifocal tensor is computed by different methods for different image configurations and for a varying number of point correspondences; all with regard to nearly coplanar object points. These examples are based on synthetic data and shall demonstrate

- the differences between minimizing algebraic and reprojection error,
- the effects of considering or neglecting the internal constraints of the tensor, and
- the impact of critical configurations; i.e. how close must points lie to the same plane so that the computation fails? To answer this question the object points are placed inside a cuboid, which is then incrementally compressed in one direction till the computation fails; the compression for which the computation is still possible will be referred to as **minimum thickness** of the cuboid.

Five different image configurations: 'Tetra', 'Air1' and 'Air2' (with strong image geometry), and 'Street1' and 'Street2' (with weak image geometry), see figure 1, are summarized below. For each image configuration the trifocal tensor was computed in five different ways:

'UCA': The direct linear solution or in other words the *unconstrained solution* (with 26 DOF) *minimizing algebraic error.*

'UCR': The *unconstrained solution* (with 26 DOF) *minimizing reprojection error* realized in the Gauss-Helmert model. This iterative estimation is initialized by the 'UCA' solution.

'CR': The *constrained solution* (with 18 DOF) *minimizing reprojection error.* This iterative estimation is initialized by the 'UCA' solution.

'CR*': This iterative estimation is identical to 'CR' but it is initialized by the known true trifocal tensor.

'CA': This is a *projection method*, which returns that valid trifocal tensor TFT (with 18 DOF), represented by the vector $\boldsymbol{q}$, which lies closest to the 'UCA' solution $\boldsymbol{t}$; i.e. $|\mathrm{TFT}(\boldsymbol{t}) - \mathrm{TFT}(\boldsymbol{q})| \rightarrow \min$.

For each image configuration the object cuboid is filled with $n_0 = 512$ points, which are then projected into the images and Gaussian noise with 1 pixel standard deviation is added. From these image points a small sample of k correspondences is selected, starting from $k = 7$ (the minimum number) up to $k = 15$. For these samples the tensor

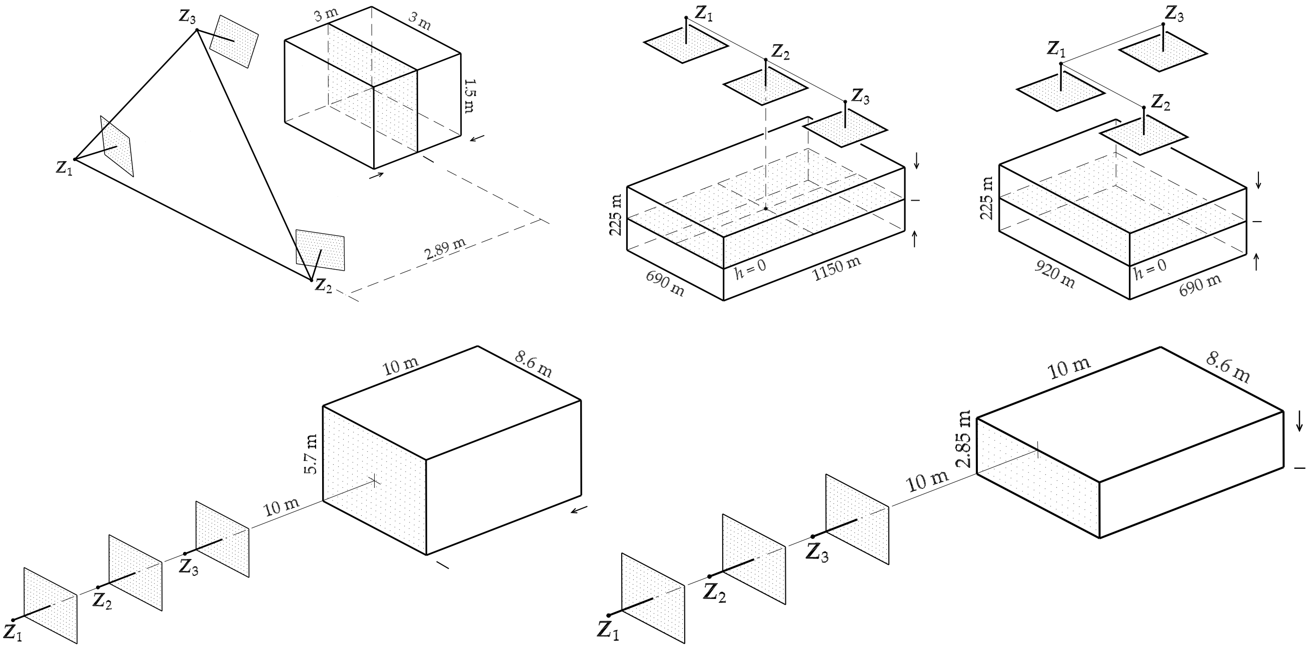

Figure 1: *Sketches of the five image configuration.* ***Upper row*** *- configurations with strong image geometry: 'Tetra', 'Air1' and 'Air2'.* ***Lower row*** *- configurations with weak image geometry: 'Street1' and 'Street2'. The arrows point in direction of the cuboid compression, the limit of which is represented by the small bar. The camera geometry for 'Tetra', 'Street1' and 'Street2' is based on the Nikon DCS460; i.e.* 3000×2000 *pixels (9 μm) with principal distance of 3500 pixel. The camera geometry for 'Air1' and 'Air2' is based on the aerial normal angle case; i.e. image size* $23{\times}23$ *cm*^{2}*, assumed to be scanned with 15 μm, and principal distance 300 mm.*

is computed by the five methods mentioned above. From the computed tensor projection matrices are extracted following [Hartley 1997], which are then used to determine all n_0 points into object space by spatial intersection. This results in a projective, not a Euclidian, reconstruction of all n_0 points. Using the known Euclidian object coordinates of the k sample points a transformation M between the projective reconstruction and the true Euclidian space can be computed. With M the other $n_0 - k$ points are also transformed from the projective reconstruction to the Euclidian space and the mean and maximum differences (termed *ground errors*) between the transformed and known Euclidian positions are determined.

These steps are repeated for 1000 samples and the overall mean of the mean and maximum ground errors are stored. Then the cuboid is compressed in a certain direction and the whole process is repeated for the thinner cuboid. The overall means of the mean and maximum ground errors are then plotted against the compression rate of the cuboid; see figure 2 (left part) which shows the plot for the 'Tetra'-configuration and 8 point correspondences.

For each image configuration a threshold for the mean ground error was set. If this threshold is hurt by the mean ground error of a certain sample, the *number of failures* for the respective pair (image configuration and cuboid compression) is increased by 1. After all 1000 samples the percentage of failures are also plotted against the cuboid compression; see figure 2 (right part).

Due to the space limitation only the plot for the image configuration 'Tetra' and 8 corresponding points is included in this article. For a detailed description of these synthetic experiments with all respective plots see [Ressl 2003]. The results found there can be summarized in the following way:

- The *ground errors* obtained for the different computation methods tend to be the same
  - for a particular image configuration, if the *number of point correspondences increases*, and
  - for a particular number of point correspondence, if the *stability of the image geometry increases.*

  Therefore, from 10 points onwards for 'Tetra', 'Air1' and 'Air2', and from 25 points onwards for 'Street1' and 'Street2', all computation methods return practically the same result; i.e. then *the simple direct linear solution ('UCA') is equivalent to the rigorous constrained computation ('CR') minimizing reprojection error.*

- For small numbers of point correspondences, ≤ 8 for the stable configurations 'Tetra', 'Air1' and 'Air2' and for the unstable configurations 'Street1' and 'Street2' in general, the unconstrained minimization of reprojection error generally performed better than any algebraic (constrained or unconstrained) minimization. Therefore, *the benefit of minimizing only reprojection error, is larger than of considering only the constraints.*

- Concerning the impact of the *minimum thickness of the cuboid* on the various computation methods, we can say that

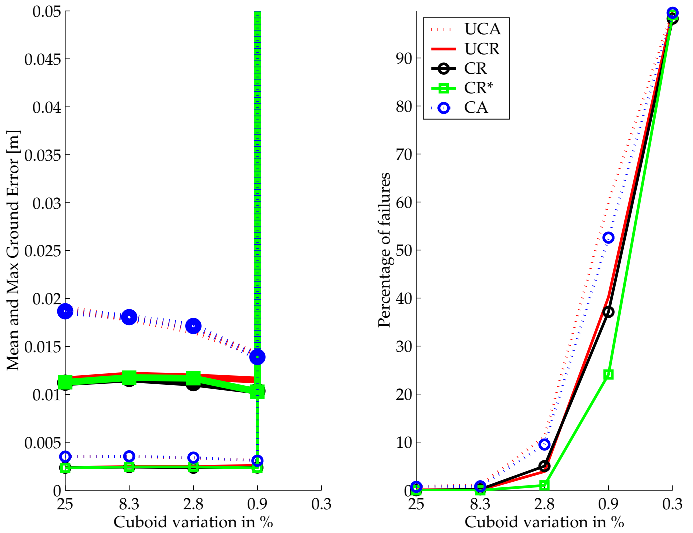

Figure 2: *Plot of the mean and maximum ground errors for the 'Tetra' configuration using: 8 points, 1 pixel noise, camera distance: 3 m, threshold for failures: 0.025 m. The mean ground error is plotted with thin lines, the maximum ground error with thick lines.*

  - for image configurations with strong geometry, i.e. 'Tetra', 'Air1' and 'Air2', the minimum thickness is practically independent on the computation method of the tensor, and becomes smaller for larger numbers of point correspondences.
  - for image configurations with weak geometry, i.e. 'Street1' and 'Street2', the minimum thickness is larger for the algebraic methods and smaller for the reprojection methods, where the consideration of the internal constraints adds a significant additional benefit; and for larger numbers of point correspondences the minimum thickness also gets smaller generally.

- Depending on the number of point correspondences, the computation of the trifocal tensor was *still successful* for the minimum thickness of the cuboid given in table 1; afterwards it failed. Because of the *discrete* cuboid compressions the actual minimum thickness is smaller than the presented values; i.e. if the percentage of failures for one compression is zero and for the next it is non-zero, then the actual minimum thickness lies somewhere in between.

4 EXPERIMENTAL RESULTS FROM REAL DATA

The findings of the previous section suggest that the practically most relevant image configurations 'Air1' and 'Tetra', could be used also for very flat objects, with a minimum thickness of about 1% of the camera distance, provided 10 point correspondences are available. One such object could be the facade of a building.

The Institute of Photogrammetry and Remote Sensing in Vienna created a test field for the calibration of terrestrial cameras by sticking retro reflecting targets on the facades of their inner courtyard; [Ballik 1989]. Due to the known object coordinates of the targets, these facades provide a suitable object to test the determination of the trifocal tensor for nearly coplanar object points using real data.

The right part of figure 3 shows one of three images from one of the facades, which were taken approx. 35 m away - from the third floor of the opposite part of the building using a calibrated Canon EOS 1Ds with a 20 mm objective[1]. Considering this camera distance the minimum thickness of about 1% would correspond to approx. 35 cm. Unfortunately the retro reflecting targets on this facade have a depth range of only 18 cm. Furthermore, the 20 mm objective used for this test has a significant amount of radial image distortion, see figure 3 left part, which was not removed in advance from the three images used for this test. Also the focal length of 20 mm is significantly different from the focal length of 32 mm, which was used for the synthetic 'Tetra' configuration.

Because of all these differences between the setup for the synthetic and the real data, the required thickness of the object points to successfully determine the trifocal tensor is larger than 1% of the camera distance and using only points on the facade fails. Therefore one point (#1) on the roof significantly behind the facade by 3.3 m has to be used in the test sample to successfully determine the trifocal tensor for the three images; see figure 3 right part. On the selected facade 121 points in total are visible in all three images and are extracted automatically with an accuracy of approx. 0.4 pixel[2].

The task of this experiment using real data is somewhat changed compared with the synthetic data: (i) compute the trifocal tensor for the three images using a subset of

[1] The digital camera Canon EOS 1Ds is equipped with a CMOS sensor of size 24 × 36 mm^2 and 4064 × 2704 pixels.

[2] This accuracy is better than the 1 pixel noise in the synthetic test, but due to the other prevailing differences (distortion and camera distance), the required thickness is still larger than 1%.

configuration	number of points		
	8	10	15
'Tetra': camera distance approx. 3 m	8.3 %	2.8 %	0.9 %
'Air1': camera distance approx. 1500 m	>15 %	1.7 %	0.6 %
'Air2': camera distance approx. 1500 m	5 %	1.7 %	0.6 %
configuration	number of points		
	15	20	25
'Street1': camera distance approx. 10 m	50 %	50 %	25 %
'Street2': reference distance approx. 2.85 m	50 %	25 %	12.5 %

Table 1: *Minimum thickness in percent of the camera distance for which the computation of the trifocal tensor was still successful. For the configurations 'Tetra', 'Air1' and 'Air2' the given values hold for any computation method, whereas for the two 'Street' configurations they hold only for the constrained method minimizing reprojection error ('CR').*

k points, (ii) retrieve the exterior orientation parameters for the relative orientation and since the camera is calibrated use the known calibration (2061.0, -1339.0, 2292.8) for this purpose. The extracted orientation parameters shall serve as approximate values for a subsequent bundle bock adjustment of the three images (with fixed interior orientation and fixed distortion) using the subset of the k points. After the block adjustment all 121 points in the three images are computed in object space by spatial intersection. This Euclidian reconstruction differs from the known positions of the retro reflecting targets by an absolute orientation A, i.e. shift, rotation and scale. This transformation A is computed using all 121 points. The remaining discrepancies after the absolute orientation are used to judge the quality of the initialization of the block adjustment.

The aim is to find the smallest possible subset of points for the computation of the tensor, which still provides good enough approximate values for the initialization of the bundle block adjustment. This task can already be solved with the minimum number of 7 points, which are shown in figure 3 right part. And as it turned out, for the creation of the approximate values it is not relevant whether the tensor is computed in the simple way 'UCA' (without the constraints and minimizing algebraic error) or in the rigorous way 'CR' (with the constraints and minimizing reprojection error). The remaining errors after finishing the bundle block adjustment, using the known calibrated interior orientation and the known non-linear distortion parameters, and the absolute orientation are:

reconstruction errors [m]		
	mean	max
x	0.009	-0.030
y	0.009	-0.033
z	0.006	0.030

For another task with the real data we could also neglect the known interior orientation of the camera and use Kruppa's equations to derive a common interior orientation for the three images. In this case, however, the depth range of the used subset of object points has to be expanded a lot, see figure 3 right part, and at least 15 points must be used: 11 points from within the facade, 2 points from the roof (lying 3.3 m behind the facade) and 2 points 3.1 m in front of it. For this point sample the rigorous computation of the tensor in the Gauss-Helmert model by minimizing reprojection error and considering the internal constraints (method 'CR') is necessary, because for the direct linear solution (method 'UCA') no valid interior orientation can be obtained using Kruppa's equations. The remaining errors after finishing the bundle block adjustment, using the determined interior orientation (2059.7, -1043.8, 2726.6) fixed and without any non-linear distortion parameters, and the absolute orientation are:

reconstruction errors [m]		
	mean	max
x	0.115	0.326
y	0.113	0.370
z	0.100	-0.429

5 SUMMARY

Concerning the computation of the trifocal tensor we discovered the following using different synthetic examples:

- the difference between minimizing algebraic and reprojection error in the computation is negligible the more point correspondences are used and the more the respective object points deviate from a common plane,
- minimization of reprojection error is more important than considering the internal constraints
- if the image geometry is not too bad and at least 10 point correspondences are used, a minimum thickness of the object points of about 1 % of the camera distance is already enough to allow a proper solution for the trifocal tensor.

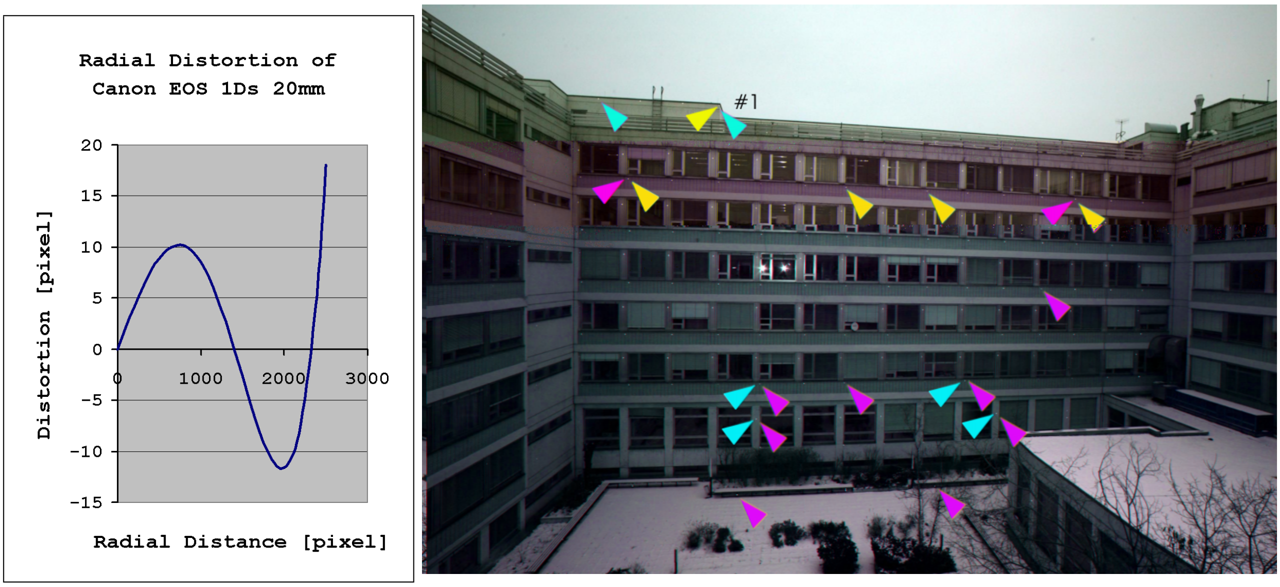

Figure 3: Left part: *The non-linear, i.e. radial , image distortion of the camera Canon EOS 1Ds with a 20 mm objective.* ***Right part***: *The second of three images of the facade used for the experiment. The 7 yellow arrows pointing up-right mark the points used to compute the trifocal tensor for the task with known interior orientation; #1 marks the one point on the roof significantly behind the facade by 2.5 m. The 15 blue arrows pointing up-left mark the points used to compute the trifocal tensor for the task with unknown interior orientation.*

Guided by this findings we also carried out an example using real images taken from a facade. From this example we see:

- if the interior orientation of the camera is known, then the exterior orientation derived from the trifocal tensor is sufficient to initialize a bundle block adjustment, even in the presence of significant radial distortion and even if the tensor is computed with the minimum number of 7 point correspondences (since the minimum thickness of the points on the facade was below 1 % of the camera distance, one point significantly away from the facade was necessary),
- if the interior orientation is unknown and the tensor is computed with at least 15 points with 4 points significantly away from the facade, reasonable approximate values for the interior orientation can be obtained using Kruppa's equations.

Although for most of the investigated examples, the simple direct linear solution ('UCA') for the tensor and the rigorous solution in the Gauss-Helmert model ('CR') return similar results, the latter solution is the recommended one, because for some situations the simple solution fails to yield a usable result.

Therefore the recommended strategy for image orientation is to first estimate the trifocal tensor rigorously in the Gauss-Helmert model, then to derive - if unknown - a common interior orientation, to extract the exterior orientation and finally to initialize with those orientation parameters a bundle block adjustment, which refines the orientation by additionally modelling the non-linear image distortion.

Acknowledgment

Part of this work was supported by the Austrian Science Fund FWF (P13901-INF).

References

Ballik, C. (1989). Signalisierung in der Präzisionsphotogrammetrie mit retroreflektierendem Material. Master's thesis, Technische Universität, Wien.

Hartley, R. (1995). In defence of the 8-point algorithm. In *Proceedings of the 5th International Conference on Computer Vision*, pp. 1064–1070. IEEE Computer Society Press.

Hartley, R. (1997). Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision 22*, 125–140.

Hartley, R. (1998). Computation of the quadrifocal tensor. In *Computer Vision - ECCV 1998, 5th European Conference on Computer Vision, Proceedings*, Volume 1406 of *Lecture Notes in Computer Science*, pp. 20–35. Springer.

Hartley, R. and A. Zisserman (2001). *Multiple View Geometry in Computer Vision, Reprinted Edition*. Cambridge, UK: Cambridge University Press.

Koch, K. R. (1999). *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer.

Luong, Q. and O. Faugeras (1996). The fundamental matrix: Theory, algorithms, and stability analysis. *International Journal of Computer Vision 17*, 43–76.

Ressl, C. (2003). *Geometry, Constraints and Computation of the Trifocal Tensor*. Ph. D. thesis, Vienna University of Technology.