20th International Conference on the Application of Computer
Science and Mathematics in Architecture and Civil Engineering
K. Gürlebeck and T. Lahmer (eds.)
Weimar, Germany, 20–22 July 2015

EVALUATION OF THE METRIC TRIFOCAL TENSOR FOR RELATIVE THREE-VIEW ORIENTATION

V. Rodehorst

Bauhaus-Universität Weimar
Faculties media and civil engineering
Bauhausstr. 11, 99423 Weimar, Germany
E-mail: volker.rodehorst@uni-weimar.de

Keywords: Relative Orientation, Calibrated Camera, Metric Trifocal Tensor, Evaluation

Abstract. *In photogrammetry and computer vision the trifocal tensor is used to describe the geometric relation between projections of points in three views. In this paper we analyze the stability and accuracy of the metric trifocal tensor for calibrated cameras. Since a minimal parameterization of the metric trifocal tensor is challenging, the additional constraints of the interior orientation are applied to the well-known projective 6-point and 7-point algorithms for three images. The experimental results show that the linear 7-point algorithm fails for some noise-free degenerated cases, whereas the minimal 6-point algorithm seems to be competitive even with realistic noise.*

1 INTRODUCTION

The automatic and reliable calculation of the relative camera pose and orientation from image correspondences is one of the challenging tasks in photogrammetry and computer vision. If cameras view an arbitrary 3D scene from two distinct positions, the mapping of each point depends on its unknown spatial depth. However, the projection ray of a point in the first view reprojects onto the second view as a line on which the corresponding point must be located. This relation which is independent of the scene structure is called *epipolar geometry* (see Figure 1). The *baseline* $\mathbf{b}$ joining the two projection centers $\mathbf{C}$ and $\mathbf{C}'$ intersect with the image planes at the *epipoles* $\mathbf{e}$ and $\mathbf{e}'$. Thus, for each image the epipole is the projection of the other camera center. An image point $\mathbf{x}$ in one view generates an *epipolar line* $\mathbf{l}'$ in the other view and all these lines pass through the particular epipole. The plane defined by an object point $\mathbf{X}$ and the two projection centers is called *epipolar plane*.

Algebraically, we can describe this relation between points and lines in different planar coordinate systems with help of the *fundamental matrix* [1, 2]. Given at least $n \geq 8$ image point pairs $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ in the image planes which are distributed in general position (i.e. are not collinear), we can linearly compute the 3×3 fundamental matrix $\mathbf{F}$ with seven degrees of freedom, so that all epipolar lines $\mathbf{l}'_i = \mathbf{F}\mathbf{x}_i$ can be derived. The *essential matrix* describes the same relation in the calibrated case where the interior orientation is known. An interesting aspect of the minimal 5-point algorithms is their stability, even if points from coplanar objects are observed. An overview and evaluation of relative pose estimation methods for image pairs can be found in [3].

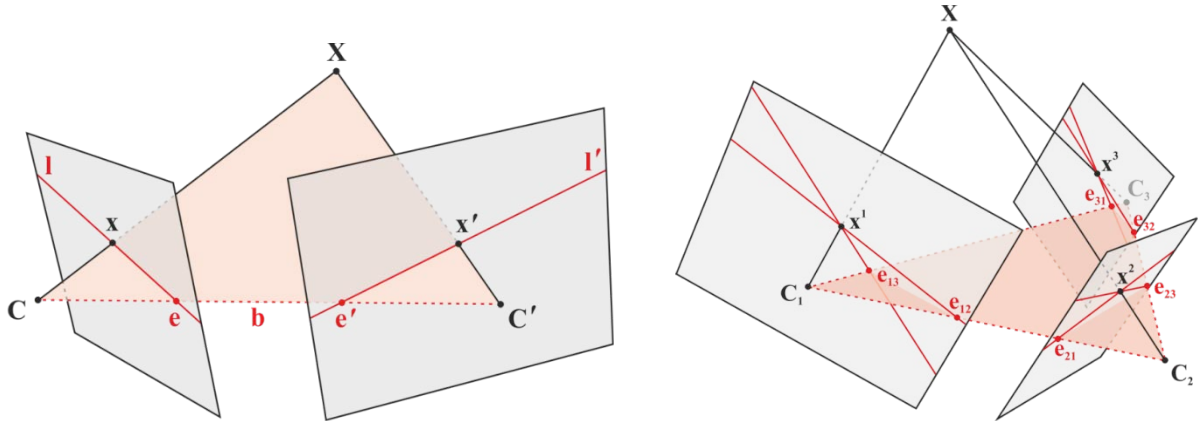


Figure 1: Epipolar geometry (left) and trifocal geometry (right)

2 TRIFOCAL TENSOR

Although two views are enough to determine depth, the point correspondence between two views is not unique, since a point is related to a line instead of a point. By extending the geometry to three views, the point correspondence ambiguity can be removed using the third view. If an object point is identified in two images and the projection centers are not collinear, the position in the third image can be predicted unambiguously by intersection of the corresponding epipolar lines (see Figure 1). The three projection centers $\mathbf{C}_j$ define a so-called *trifocal plane* which introduces three additional constraints on the epipols

$$\mathbf{e}_{23}^{\mathrm{T}}\mathbf{F}_{12}\mathbf{e}_{13} = \mathbf{e}_{31}^{\mathrm{T}}\mathbf{F}_{23}\mathbf{e}_{21} = \mathbf{e}_{32}^{\mathrm{T}}\mathbf{F}_{13}\mathbf{e}_{12} = 0\,. \tag{1}$$

Algebraically, the trifocal tensor relates image points or lines seen in three views in a similar way like the fundamental matrix relates points in two views. The tensor depends on image coordinates only without an explicit spatial point or line. In this paper we concentrate on point coordinates only. Given at least $n \geq 7$ image point correspondences across three images $\mathbf{x}_i^1 \leftrightarrow \mathbf{x}_i^2 \leftrightarrow \mathbf{x}_i^3$ which are distributed in general position (i.e. are not collinear), we can compute the $3\times3\times3$ trifocal tensor $\mathsf{T} = \left[\mathsf{T}_1, \mathsf{T}_2, \mathsf{T}_3\right]$ with 18 degrees of freedom, where each submatrix T_j is singular with rank two. At first, the image points $\mathbf{x}_i^1$, $\mathbf{x}_i^2$ and $\mathbf{x}_i^3$ must be conditioned separately by translation to the origin and scaling to a mean distance of $\sqrt{2}$ using the 3×3 similarity transformations $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$. Then, the trifocal tensor for the conditioned image points $\tilde{\mathbf{x}}_i$ can be derived from the *point-point-point* relation [5]

$$\left[\tilde{\mathbf{x}}^2\right]_\times \left(\sum_{j=1}^{3} \tilde{x}_j^1 \tilde{\mathsf{T}}_j\right) \left[\tilde{\mathbf{x}}^3\right]_\times = \mathbf{0}\,. \tag{2}$$

where $[\ ]_\times$ denotes a 3×3 skew-symmetric matrix. For each point triplet we get four constraints on the elements of $\tilde{\mathsf{T}}$ which can be combined into a $4n\times27$ design matrix $\mathbf{A}$. Its derivation is omitted for simplicity but the complete matrix can be found in [4]. The linear homogeneous equation system of the form $\mathbf{At} = \mathbf{0}$ can be solved for the 27 elements of vector $\mathbf{t}$ with help of *singular value decomposition* (SVD). After reshaping the solution vector row-wise into the $3\times3\times3$ tensorial slices, we have a basic trifocal tensor $\tilde{\mathsf{T}}$ for the conditioned points.

However, the trifocal tensor has 27 elements but only 18 degrees of freedom. This basic linear approach does not satisfy the eight remaining algebraic constraints. In [5] a method is proposed that takes the basic tensor as initial estimate and finds a geometrically valid tensor that satisfies all internal constraints. At first, the two epipols $\mathbf{e}_{21}$ and $\mathbf{e}_{31}$ are extracted from $\tilde{\mathsf{T}}$ as the left and right null-vectors according to

$$\left[\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\right]^{\mathrm{T}} \mathbf{e}_{21} = \mathbf{0} \quad \text{with} \quad \mathbf{u}_j^{\mathrm{T}} \mathsf{T}_j = \mathbf{0}^{\mathrm{T}} \quad \text{and} \quad \left[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\right]^{\mathrm{T}} \mathbf{e}_{31} = \mathbf{0} \quad \text{with} \quad \mathsf{T}_j \mathbf{v}_j = \mathbf{0}\,. \tag{3}$$

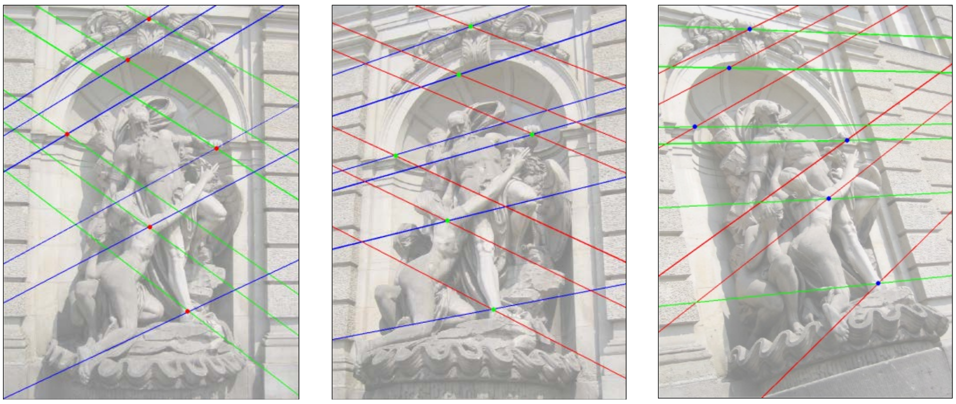


Figure 2: Trifocal geometry computed from of six corresponding image point triplets. The example shows the Prometheus fountain at Berlin University of the Arts

Then, a 27×18 matrix $\mathbf{E}$ is assembled such that $\mathbf{t} = \mathbf{Ea}$, where $\mathbf{E}$ expresses the linear relationship of the unknown projection matrices $\mathbf{P}_2$ and $\mathbf{P}_3$ with known epipols

$$\tilde{\mathsf{T}}_{jkl} = a_{jl} e_{31_k} - e_{21_j} b_{kl} \quad \text{for} \quad j,k,l = 1,2,3 \tag{4}$$

and vector $\mathbf{a}$ contains the remaining 18 elements a_j^k and b_j^l. Now, we have to minimize the algebraic error of $\|\mathbf{AEa}\|$ subject to $\|\mathbf{Ea}\| = 1$ using a constrained *direct linear transformation* (DLT) algorithm. After reshaping the 27 elements of the solution vector into tensor form, the final T for the original points can be obtained by deconditioning

$$\mathsf{T}_j = \mathbf{T}_2^{-1} \left(\sum_{k=1}^{3} \mathbf{T}_{1_{kj}} \tilde{\mathsf{T}}_k \right) \mathbf{T}_3^{-\mathsf{T}} \quad \text{for} \quad j = 1,2,3\ . \tag{5}$$

Similarly to the fundamental matrix, there exists also a *minimal* algorithm to determine the trifocal tensor from exactly six point correspondences across three images [6]. An example of the computed trifocal geometry is given in Figure 2.

3 CALIBRATED TRIFOCAL TENSOR

In many applications we are able to calibrate the cameras in beforehand. Taking advantage of the known interior orientation stored in the upper-triangular 3×3 *calibration matrix* $\mathbf{K}$ and the metric parameterization of the transformation functions should stabilize the reconstruction process. For the metric 3D reconstruction from three calibrated views we can parameterize the trifocal tensor with five parameters for the relative orientation of the first two cameras and six parameters for the exterior orientation of the third camera.

Since each image point correspondence $\mathbf{x}_i^1 \leftrightarrow \mathbf{x}_i^2 \leftrightarrow \mathbf{x}_i^3$ provides three constraints, we need at least $n \geq 3\tfrac{2}{3}$ point triplets to determine the eleven unknown parameters. However, the problem is still not solved satisfactorily due to its algebraic complexity and lack of efficient algorithms. For example [7] parameterizes the relative orientation by the two translation vectors $\mathbf{t}_2$ and $\mathbf{t}_3$ and the two quaternions $\mathbf{q}_2$ and $\mathbf{q}_3$ for the rotations. These 14 parameters are further reduced by fixing the first baseline to unit length using the constraint $\mathbf{t}_2^{\mathsf{T}}\mathbf{t}_2 = 1$ and normalizing the quaternions with $\mathbf{q}_2^{\mathsf{T}}\mathbf{q}_2 = 1$ and $\mathbf{q}_3^{\mathsf{T}}\mathbf{q}_3 = 1$. The detailed estimation procedure for this metric trifocal tensor is studied in [8].

A minimal algorithm for the metric trifocal tensor is sketched in [9] where only four points in three calibrated views are used. However, in case of noisy point correspondences, the reprojected fourth object point will not exactly coincide with the observed image point and therefore the authors demonstrates the enormous complexity by visualizing the remarkable solution space. The minimal parameterization of the trifocal tensor in [10] uses projective line geometry in order to provide necessary and sufficient constraints. Finally, in [11] a new meaningful parametrization with non-colinear pinholes from a quotient Riemannian manifold is proposed.

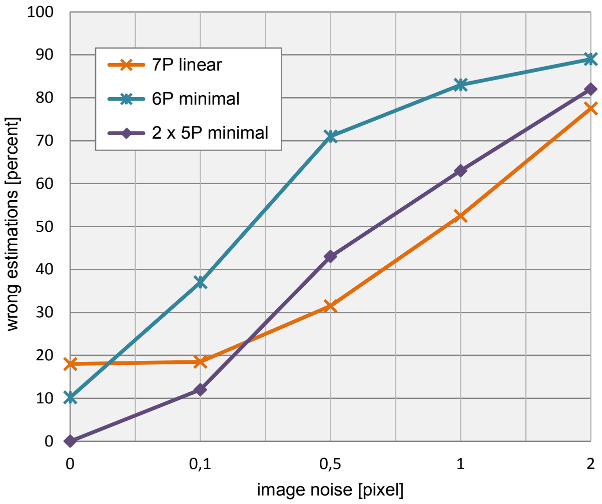


Figure 3: Comparison of the relative orientation of three views using synthetic data

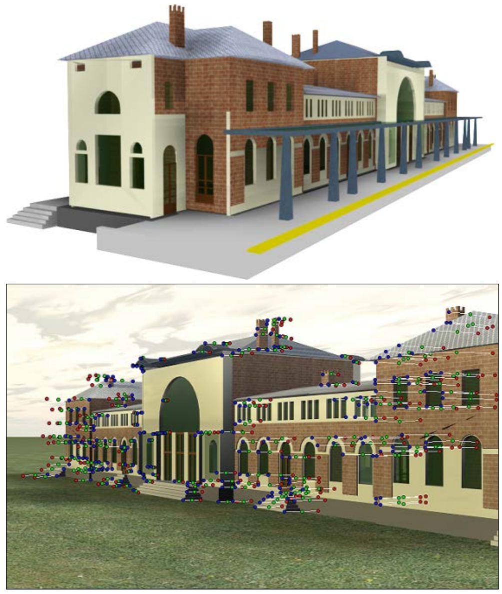


Figure 4: Textured 3D model with matched image feature triplets

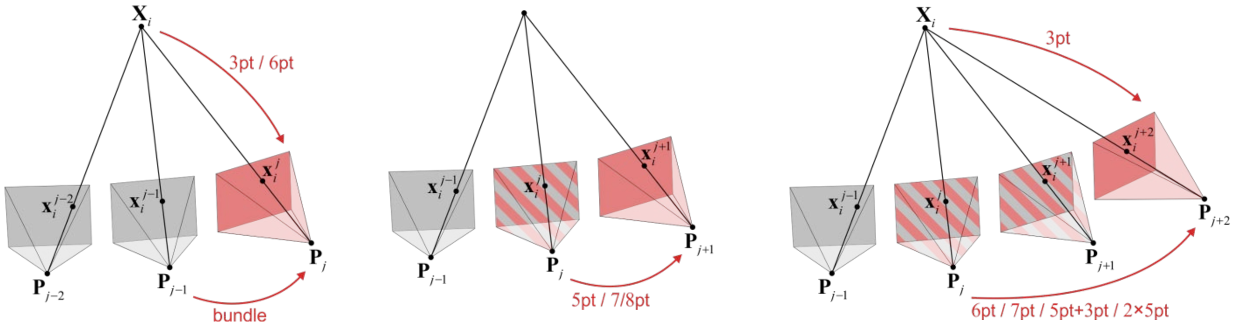
However, in this paper we simply compute the linear trifocal tensor from the normalized image triplets $\tilde{\mathbf{x}}_i^1 \leftrightarrow \tilde{\mathbf{x}}_i^2 \leftrightarrow \tilde{\mathbf{x}}_i^3$ by applying the inverted calibration matrices $\tilde{\mathbf{x}}_i = \mathbf{K}^{-1}\mathbf{x}_i$. The fundamental matrices contained in the trifocal slices should now describe the individual rotations and translations. Since both baselines are thereby fixed to unit length, we have to find a proper scaling for the second baseline. At first spatial coordinates are triangulated using the camera pairs $(\mathbf{P}_1, \mathbf{P}_2)$ and $(\mathbf{P}_1, \mathbf{P}_3)$.
Then, the mean distance of the object points to the common camera $\mathbf{P}_1$ defines the scale ratio. Note that this simple linear version does not incorporate all the partly unknown and complicated polynomial constraints that are required for an optimal metric trifocal tensor. Nevertheless, the computation of the direct linear solution is fast and unique. In contrast to the fundamental matrix this version of the metric trifocal tensor works for planar point configurations too.

4 EXPERIMENTAL RESULTS

To evaluate the relative orientation of three images, we analyzed the following approaches

- The metric trifocal tensor which is based on the *linear 7-point algorithm* [5,4]
- The *minimal 6-point algorithm* [6,4] with normalized image coordinates
- The *minimal 5-point algorithm* [3] for two image pairs with recovery of different scaling

using synthetically generated data with ground truth. At first 100 spatial object points are randomly generated in general position and projected into the images using simulated cameras. The camera positions are also selected arbitrarily, but the view directions are restricted to the approximate location of the point cloud. The image coordinates range from 1 to 1024 and are displaced with Gaussian noise. Every technique is examined 10,000 times to obtain statistically significant results. Critical point configurations or camera motions are not explicitly checked for these synthetic experiments and may lead to false estimations.
The experimental results are summarized in Figure 3. In the tested metric case, the minimal 5-point algorithm for calibrated image pairs performs always better than the minimal 6-point algorithm for three uncalibrated images. It seems that because of some missing constraints the linear 7-point algorithm fails for some degenerated cases although the input data is noise-free. However, with realistic noise the two additional point triples stabilize the approach and the linear 7-point algorithm overtakes the minimal 5-point algorithm.



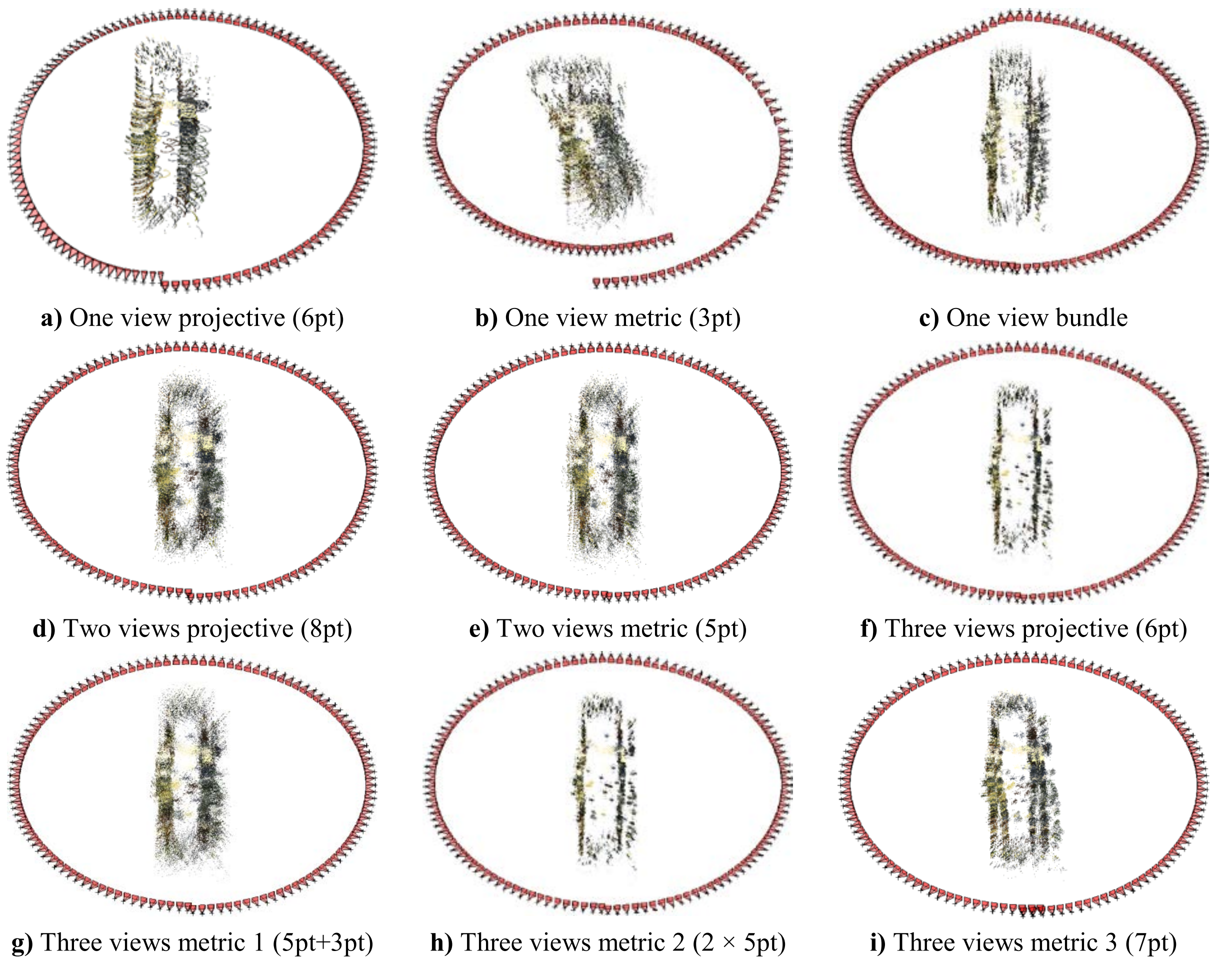
a) Single-view orientation **b)** Relative two-view orientation **c)** Relative three-view orientation

Figure 5: Tested strategies for camera path estimation

The following tests demonstrate the performance of the proposed multi-view algorithms. Figure 4 shows our synthetic 3D model of the railway station in Bonn together with the results of the automatically detected and matched image features. The camera generates 120 images while moving around the building and keeping it in view. On average 960 corner features are extracted from each image which typically yields 310 correspondences over three successive views. Additionally, we add some normal distributed noise with a standard deviation of 0.5 pixels to the measured image coordinates in order to emphasize the differences between the tested strategies (see Figure 5).

To compare the general projective methods, we removed the known calibration information of the synthetic camera from the images points and recovered the camera motion according to the metric algorithms [3,4].

- **One view**: if an initial reconstruction from two or three images is available, the subsequent camera positions may be recovered from the spatial coordinates using single-view orientation methods:
  - **Projective**: the new camera position can be determined from at least six object coordinate projections using the over-determined *spatial resection* [5]. The matching object coordinates may be identified by analyzing the image point correspondences to the existing reconstruction.
  - **Metric**: in order to prevent that the reconstruction error is shifted to the interior orientation, the camera motion can be computed with help of the *minimal 3-point* pose estimation algorithm [4]. For this experiment the required three points are selected randomly and the method is repeated 100 times. Finally, the solution with the lowest reprojection error is taken as result.
  - **Bundle**: if the camera motion is somewhat smooth with a relative small baseline, the new camera position may be computed with incremental *bundle adjustment* [5,4] only, where the initial values are taken from the last camera of the already processed sequence.
- **Two views**: many techniques for pose or ego-motion estimation rely on two succesive views. Subsequently, we compare the following relative orientation methods for camera pairs [3]:
  - **Projective**: very simple and computational effective is the over-determined *8-point algorithm* for the fundamental matrix.
  - **Metric**: a little more robust is the *minimal 5-point algorithm* for computing the essential matrix. In our implementation the five points are selected randomly and the algorithm is applied 100 times. Finally, the solution with the lowest sampson distance is taken as result.
- **Three views**: Finally we present the strategies for camera path estimation which are based on the relative orientation of camera triplets:
  - **Projective**: we apply the *minimal 6-point* algorithm for the general projective trifocal tensor. The required six points are selected randomly and the method is repeated 100 times. Finally, the solution with the lowest reprojection error is taken as result.
  - **Metric 1**: the first metric method consists of the *minimal 5-point* algorithm to compute the essential matrix for the first two cameras followed by the *minimal 3-point* algorithm for the pose of the third camera.
  - **Metric 2**: the second alternative uses the *minimal 5-point* algorithm to compute the essential matrix for the first and second camera pair. Then we triangulate spatial coordinates and take the



a) One view projective (6pt) **b)** One view metric (3pt) **c)** One view bundle

d) Two views projective (8pt) **e)** Two views metric (5pt) **f)** Three views projective (6pt)

g) Three views metric 1 (5pt+3pt) **h)** Three views metric 2 (2 × 5pt) **i)** Three views metric 3 (7pt)

Figure 6: Strategies to reconstruct the object structure and camera motion using different image orientation techniques

mean distance of the object points to the common camera to find a proper scale ratio for the baselines.

- **Metric 3**: as last method we demonstrate the results of the proposed over-determined *7-point algorithm* for the metric trifocal tensor.

In a next step these initial reconstructions should be improved by extending the baselines for better triangulation. At first the image features must be tracked as long as possible to maximize the camera distance. For the exemplary sequence of the railway station an image point is on average visible in 4.5 subsequent views. However, some features can be tracked up to 22 frames which extend the triangulation angle from 3 to 66 degrees. For the following experiment we started from the two-view metric reconstruction. Then we selected all image points which are observable in at least four images and triangulated the object coordinates using all involved views. Thereby, the number of points is reduced to approximately the half while the accuracy of the spatial coordinates is apparently better. As already mentioned, the final step of a reconstruction algorithm should be the global optimization using bundle adjustment in order to obtain maximum accuracy. In contrast to the preceding step, which modifies the spatial objects points only, improves the metric bundle adjustment the cameras as well.

5 SUMMARY AND OUTLOOK

Our experiments of the calibrated metric case have shown that for single-image orientation the minimal 3-point algorithm and for the relative orientation of image pairs the minimal 5-point algorithms perform best, especially in presence of noise. For the relative orientation of three calibrated images we also recommend applying the minimal 5-point algorithm independently to the image pairs and recovering the different scaling using the depth ratio of corresponding object coordinates. However, even though the metric trifocal tensor based on the linear 7-point algorithm does not enforce all necessary metric constraints, it can be more stable than the 5-point algorithms due to the possibility of over-determined estimation. Our detailed investigations of the relative orientation methods in [3] have shown that in general, the estimation of camera rotation is more reliable than the translation. The minimal 5-point algorithms often provide multiple solutions. The best selection criterion is a combination of a preceding cheirality test with the minimal number of points followed by the computation of the Sampson distance over all available points.

The registration of individual reconstructions based on spatial points suffers from critical surfaces and inaccurate triangulated object coordinates. Therefore we recommend determining a spatial similarity transformation using a single-view overlap for image pairs and a double-view camera consistency for image triplets. In case of a moving camera that captures video frames at high frequency the extension of the baseline is essential to improve the triangulation accuracy. However, the sequential approach accumulates the estimation errors over time and the reconstructed path tend to drift. To reduce this effect, the result necessarily should be improved with a global optimization technique like bundle adjustment.

Nevertheless, if the initial reconstruction is far from the correct solution, the non-linear optimization step is not able to work wonders. The general projective approaches are simple and quite helpful, if no calibration information is available so far. However, if coplanar points on dominant planes are observed the spatial resection with the 6-point algorithm, the relative two-image orientation with the minimal 7-point or linear 8-point algorithm and the relative orientation of image triplets with the minimal 6-point algorithm will fail. Unfortunately, all relative orientation algorithms produce wrong estimates from time to time. If a camera path over several hundred frames needs to be reconstructed only one miscalculation may corrupt the whole path. Thus, additional multi-camera constraints [3,4] are required to gives stable results for extensive camera path reconstructions.

REFERENCES

[1] R.I. Hartley: In defense of the eight-point algorithm, In T-PAMI, 19(6), 580—593, 1997.

[2] H.C. Longuet-Higgins: A computer algorithm for reconstructing a scene from two projections, Nature, vol. 293, 133—135, 1981.

[3] V. Rodehorst, M. Heinrichs and O. Hellwich: Evaluation of Relative Pose Estimation Methods for Multi-Camera Setups, IAPRS 37(B3b), 135—140, 2008.

[4] V. Rodehorst: Photogrammetric Computer Vision for spatio-temporal 3D reconstruction, habilitation thesis, Technische Universität Berlin, 2013.

[5] R.I. Hartley and A. Zisserman, 2004: Multiple view geometry in computer vision, Cambridge University Press, 2nd edition, 1—672, 2004.

[6] P.H.S. Torr and A. Zisserman: Robust parameterization and computation of the trifocal tensor, Image and Vision Computing, 15(8), 591—605, 1997.

[7] W. Förstner: New orientation procedures, In IAPRS, 33(3), 297—304, 2000.

[8] S. Abraham: Kamera Kalibrierung und metrische Auswertung monokularer Bildfolgen, Shaker Verlag, 1—170, 2000.

[9] D. Nistér and F. Schaffalitzky: Four points in two or three calibrated views: theory and practice, In IJCV, 67(2), 211—231, 2006

[10] J. Ponce, M. Hebert and M. Trager: Trinocular geometry revisited, submitted to IJCV, 2015.

[11] S. Leonardos, R. Tron and K. Daniilidis: A metric parametrization for trifocal tensors with non-colinear pinholes, In CVPR, 259—267, 2015.