BENCHMARKING AUTOMATIC BUNDLE ADJUSTMENT RESULTS

Timo Dickscheid, Thomas Läbe and Wolfgang Förstner

Institute of Geodesy and Geoinformation, Department of Photogrammetry
University of Bonn
Nussallee 15, 53115 Bonn, Germany
dickscheid@uni-bonn.de, laebe@ipb.uni-bonn.de, wf@ipb.uni-bonn.de

Commission III/1

KEY WORDS: Digital photogrammetry, bundle adjustment, accuracy analysis, quality control, sensor orientation

ABSTRACT:

In classical photogrammetry, point observations are manually determined by an operator for performing the bundle adjustment of a sequence of images. In such cases, a comparison of different estimates is usually carried out with respect to the estimated 3D object points. Today, a broad range of automatic methods are available for extracting and matching point features across images, even in the case of widely separated views and under strong deformations. This allows for fully automatic solutions to the relative orientation problem, and even to the bundle triangulation in case that manually measured control points are available. However, such systems often contain random subprocedures like RANSAC for eliminating wrong correspondences, yielding different 3D points but hopefully similar orientation parameters. This causes two problems for the evaluation: First, the randomness of the algorithm has an influence on its stability, and second, we are constrained to compare the orientation parameters instead of the 3D points. We propose a method for benchmarking automatic bundle adjustments which takes these constraints into account and uses the orientation parameters directly. Given sets of corresponding orientation parameters, we require our benchmark test to address their *consistency* of the form deviation and the internal precision and their *precision level* related to the precision of a reference data set. Besides comparing different bundle adjustment methods, the approach may be used to safely evaluate effects of feature operators, matching strategies, control parameters and other design decisions for a particular method. The goal of this paper is to derive appropriate measures to cover these aspects, describe a coherent benchmarking scheme and show the feasibility of the approach using real data.

1 INTRODUCTION

1.1 Motivation

Automatic bundle adjustment is a widely used tool for photogrammetric applications. A lot of work has been done on different bundle parametrizations, error models, on linearization, robustification methods and optimization strategies. An excellent overview is given in Triggs et al. (2000). Both proprietary and open source (Lourakis and Argyros, 2004) bundle adjustment implementations are available. In the last years, a growing number of feature detectors and descriptors have been proposed which allow for automatic matching under widely separated views and a considerable degree of distortions, allowing us to compute the relative orientation of image sequences fully automatically. The question arises how to verify, and especially how to benchmark such automatic bundle adjustment methods in a statistically sound manner. It would be desirable to compare the precision of such systems under varying conditions and against each other.

Benchmark tests are usually carried out by processing datasets with ground truth. In fact though, reliable ground truth data for camera orientations is only available when using artificially rendered images to our knowledge. In a real scenario, it is very difficult to measure camera orientations with an accuracy which is clearly superior to the expected bundle adjustment. Validation of course needs to take into account both uncertainties, that of the test data set and that of the reference data set. Furthermore, the use of data from external sensors like IMU's lacks a proper theoretical model for the evaluation.

The classical error model for bundle adjustment is the squared reprojection error of observed homologous points, in case of homogeneous precision actually realizing a Maximum-Likelihood estimate. As stated above however, automatic systems often contain random components for filtering the correspondences and hence yield different 3D points, disqualifying the points as a basis for the evaluation. Our aim is thus to directly use the different estimates of camera orientations as a foundation for the evaluation.

1.2 Problem statement

Our goal is to develop measures suitable for a benchmark test of automatic bundle adjustments based on the estimated frames and their theoretical covariance matrices. With frame we denote the parameters of the exterior orientation of a camera, following Pennec and Thirion (1995). For corresponding sets of frames, we want to evaluate (1) the *consistency* c of the form deviation and the internal precision and (2) the *precision level* p w. r. t. the precision of a reference data set. Both measures should be related to a metric in order to be able to rank different methods.

We are primarily interested in comparing corresponding sets of frames estimated by different methods. If one dataset can be used as a reference with clearly superior precision, the remaining datasets may be benchmarked by computing consistency and precision level with respect to the reference for each of them.

Furthermore, if a particular method uses a random subprocedure such as RANSAC, it may be meaningful to compare repeated estimates of this method in order to evaluate the repeatability of the results. We will thus present suitable measures for comparing results from different methods as well as results from repeated runs of the same method.

1.3 Outline of the paper

We will give a brief summary on the state of the art of automatic bundle adjustment methods and benchmark tests in section

2. Section 3 gives an overview of the proposed benchmarking scheme, starting with a formalization of the problem and introducing the respective metrics. In section 4, we will detail the parameter transformations necessary for computing these metrics. The exact approaches for computing the consistency and precision measures are described in sections 5 and 6, respectively. We show the feasibility of the proposed approach with an experiment based on real data in section 7 and conclude with a short summary and outlook.

2 RELATED WORK

Fully automatic solutions to the relative orientation problem are available since several years. The algorithms are based on automatic correspondence detection techniques, ranging from simple correlation-based methods to rotation and scale invariant (Lowe, 2004) or even fully affine invariant solutions (Mikolajczyk and Schmid, 2004; Matas et al., 2004). Schaffalitzky and Zisserman (2002) were probably the first to automatically find overlapping pairs in unordered image sets under widely separated views. They use affine invariant feature descriptors for finding corresponding points in image pairs. Another solution has been proposed early by Martinec and Pajdla (2002) who make use of "tracks" to increase the amount of feature correspondences.

Pollefeys et al. (2000) started early to compute 3D structure from uncalibrated image sequences and were probably the first to implement a fully automatic approach to the relative orientation problem. Mayer (2005) has implemented an automatic system for the precise estimation of relative orientations from an image sequence for acquiring 3D building models. The system matches interest points (Förstner and Gülch, 1987) across images and performs adjustments of intermediate results whenever possible. Roth (2004) describes a fully automatic solution to the relative orientation problem for the uncalibrated case using SIFT features (Lowe, 2004) for automatic correspondence detection. He evaluates his implementation based on the reprojection errors of object points on a typical image sequence. Läbe and Förstner (2006) proposed another solution based on SIFT features, which uses calibrated cameras.

Lourakis and Argyros (2005) benchmarked different bundle adjustment implementations based on reprojections errors, using datasets with ground truth observations. Their aim was to compare different optimization strategies for the adjustment. Our goal is to deploy a method that works in the absence of ground truth and without the need for common object points, which makes the task a challenging problem. Pennec and Thirion (1995) addressed a similar problem and proposed means for dealing with frame-to-frame correspondences using an Extended Kalman Filter. They argue that errors in estimation problems of rigid transforms are not sufficiently modeled, as often applied, by an additive bias. This is due to the fact that the addition of transforms is not generally defined. Instead, they state that errors should be represented as transformations. We agree with this statement and follow a similar approach here, but aim at a closed solution for all frames instead of a recursive scheme.

Comparing the result of free networks is confronted with the *gauge* or *datum* problem: Only parameters and covariance matrices can be compared which refer to the same coordinate system. This problem has been identified and solved for sets of 2- and 3-D points by Baarda (1967) using S-Transformations, representing a differential similarity transformation into a well defined coordinate system. The concept has been generalized for point sets in arbitrary orientation by Molenaar (1981), using a K-Transformation as a preprocessing step which guarantees the S-Transformation to be close to a unit transformation. Transferring the concept of K- and S-transformations to sets of frames with six degrees of freedom is not worked out to our knowledge and marks the core of this paper.

Besides the mentioned gauge or datum problem, one has to cope with the specification of a reference covariance matrix when evaluating the precision of parameters. This problem has not been tackled in depth for sets of frames, as it has been for sets of points using covariance functions. However, we can generally determine a reference covariance matrix by specifying a measurement design and then using the resulting theoretical covariance matrix. This is the way we want to follow.

3 OVERVIEW OF THE APPROACH

3.1 Problem formalization

Input data For our discussion, we first assume that two corresponding sets of N estimated frames are given by their parameters and their joint covariance matrices:

$$\{ {}^a\boldsymbol{d}_1, {}^a\Sigma_{d_1 d_1}\} \qquad \{ {}^b\boldsymbol{d}_2, {}^b\Sigma_{d_2 d_2}\} \tag{1}$$

We use right subscripts to number the datasets, and left superscripts to denote the coordinate system in which the parameters are given. The complete vectors consist of N subvectors each, where N is the number of cameras:

$${}^a\boldsymbol{d}_1 = \begin{bmatrix} {}^a\boldsymbol{d}_{11} \\ {}^a\boldsymbol{d}_{12} \\ \vdots \\ {}^a\boldsymbol{d}_{1N} \end{bmatrix} \quad \text{and} \quad {}^b\boldsymbol{d}_2 = \begin{bmatrix} {}^b\boldsymbol{d}_{21} \\ {}^b\boldsymbol{d}_{22} \\ \vdots \\ {}^b\boldsymbol{d}_{2N} \end{bmatrix}. \tag{2}$$

The representation of the frame has to capture both translation and rotation. In our context, each frame $\boldsymbol{d}_{in}$ is a 7-dimensional vector consisting of the camera's projection center $\boldsymbol{x}_{in}$ and the camera's rotation represented by a unit quaternion $\mathbf{q}_{in}$:

$$\boldsymbol{d}_{in} = \begin{bmatrix} \boldsymbol{x}_{in} \\ \mathbf{q}_{in} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_{in} \\ q_{in} \\ \vec{\boldsymbol{q}}_{in} \end{bmatrix} \qquad i = \{1, 2\} \tag{3}$$

We consider ${}^a\Sigma_{d_1 d_1}$ and ${}^b\Sigma_{d_2 d_2}$, which represent the uncertainty of the N frames, to be two full $7N \times 7N$ covariance matrices. In general they may have a rank deficiency, due to the normalization of the quaternions and possibly due to the definition of the gauge or datum, e. g. by fixing one frame and the coordinate difference to another frame.

Consistency of corresponding datasets Consistency in our context means form deviation of two sets of corresponding frames related to the internal precision. It should be based on the average deviation of corresponding camera parameters and will be denoted by the symbol c. We require c to take small values for small form deviations, and vice versa. A straightforward basis is the Mahalanobis distance

$$\Omega = ({}^s\boldsymbol{d}_1 - {}^s\boldsymbol{d}_2)({}^s\Sigma_{d_1 d_1} + {}^s\Sigma_{d_2 d_2})^{-1}({}^s\boldsymbol{d}_1 - {}^s\boldsymbol{d}_2)^{\mathsf{T}} \tag{4}$$

which together with the redundancy $R = 6N - 7$ yields

$$c = \sqrt{\frac{\Omega}{R}} \quad . \tag{5}$$

c is F-distributed with R and ∞ degrees of freedom in case the internal covariances are correct. It holds for uncorrelated datasets given in the same coordinate system s. In section 5.1, we will give two theoretically equivalent derivations of c based on (4). The transformation of corresponding sets of frames into a common coordinate system s is explained in section 4.

Precision level of corresponding datasets When computing c, we assume the bundle adjustment procedures to know their own precision level, i. e. ${}^{s}\Sigma_{d_i d_i}$. Thus we are left with comparing their precision levels if they show consistency. We identify the difference in precision level p of corresponding datasets with the average ratio of their standard deviations:

$$p = p\left({}^{s}\Sigma_{d_1 d_1}, {}^{s}\Sigma_{d_2 d_2}\right) \tag{6}$$

The precision level p should be related to a metric of the covariance matrices involved. Again we require this quantity to take small values for small differences in precision level, and vice versa. Clearly, such a direct comparison of covariance matrices is only possible if the datasets share the same gauge and coordinate system, say s. The main part thus consists in making the K- and S-transformation according to Baarda (1967) and Molenaar (1981) explicit for our specific problem. These transformations are derived for sets of frames, not only for sets of points in section 4.

3.2 Evaluating methods containing random subprocedures

If the automatic orientation procedure that we want to evaluate has a random component, we can only trust the benchmark metrics if repeatability of the results is ensured for the same input data. Then we need to evaluate the consistency of repeated estimates from the same method, i. e. the consistency of multiple datasets $\{{}^{a}\boldsymbol{d}_k, {}^{a}\Sigma_{d_k d_k}\}$ or $\{{}^{b}\boldsymbol{d}_k, {}^{b}\Sigma_{d_k d_k}\}$, $2 \le k \le K$. More precisely, we want to address the variation of these repeated estimates related to their average internal precision. We will derive a particular consistency measure c_s for such sets of samples in section 5.2, and accept only datasets where c_s is below a threshold T_{c_s}.

Besides consistency, one may be interested in evaluating the variation of the precision level over repeated estimates, i. e. based on the variation of p. In our experiments it turned out that the precision level undergoes very small variations only, hence we will not give any detailed results in this paper. It should though be noted that a derivation may be useful when investigating procedures with very pronounced random components.

3.3 Comparing different bundle adjustment methods

When comparing different methods, we assume that the dataset $\{{}^{a}\boldsymbol{d}_1, {}^{a}\Sigma_{d_1 d_1}\}$ serves as a reference with clearly superior precision. We may then compute c and p w. r. t. this reference for a second dataset $\{{}^{b}\boldsymbol{d}_2, {}^{b}\Sigma_{d_2 d_2}\}$. Repeating this procedure for datasets from different methods but using the same reference, we end up with multiple pairs of measures (c_i, p_i). As c_i indicates consistency regarding the internal precision of a dataset, we require it not to exceed a threshold T_c in order for p_i to give meaningful evidence.

3.4 Benchmarking scheme

We come up with the following scheme for benchmarking automatic bundle adjustment methods, given a set of images:

1. Provide a reference dataset with clearly superior precision, i. e. by using a very large image resolution.

2. For every method m which contains a random subprocedure, compute its consistency c_{s_m} for a larger number of repeated estimates. Accept m for the benchmark test only if $c_{s_m} < T_{c_s}$, as described in section 5.2.

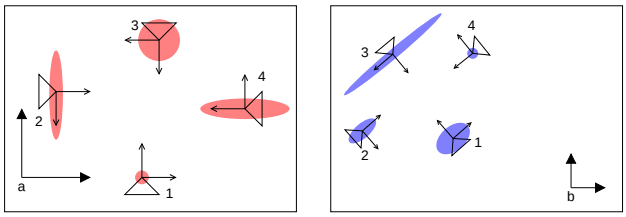

Figure 1: Birdseye view of an exemplary initial situation: Two sets of frames, left in reference system a, right in reference system b, represented by numbered tripods. The sets are approximately related by a spatial similarity transformation. The uncertainties of each camera, denoted as ellipses, as well as the gauge and origin may differ arbitrarily.

3. For every method i qualified for the test, compute c_i and r_i w. r. t. the reference dataset, as described in sections 5.1 and 6.

4. Report c_i and p_i for all methods with $c_i < T_c$.

4 TRANSFORMATION OF THE PARAMETERS

4.1 The K Transformation

The K-Transformation in our context is a similarity transformation between two sets of corresponding frames to accomplish approximate coverage in the same coordinate system. Consider the two sets of frames in (1). They generally differ by an unknown spatial similarity transformation $K(\boldsymbol{t}_K, \mathbf{q}_K, \lambda_K)$, where $\boldsymbol{t}_K$ is a translation between the projection centers, $\mathbf{q}_K$ is a 3D rotation in quaternion representation, and λ_K a scale factor. A sketch of such two parameter sets from a birdseye view is given in Figure 1. In the absence of any errors we expect

$${}^{a}\boldsymbol{d}_{1n} = \begin{bmatrix} {}^{a}\boldsymbol{x}_{1n} \\ {}^{a}\mathbf{q}_{1n} \end{bmatrix} = \begin{bmatrix} \lambda_K \cdot R(\mathbf{q}_K)\,{}^{b}\boldsymbol{x}_{2n} + \boldsymbol{t}_K \\ \mathbf{q}_K\,{}^{b}\mathbf{q}_{2n} \end{bmatrix} \tag{7}$$

We assume the similarity transformation K to be non-stochastic, therefore any method can be used to determine the parameters from a few or from all frames.[1] Applying K on ${}^{b}\boldsymbol{d}_2$ will bring the two sets of frames to approximate coverage in the coordinate system a, while the gauge however may still differ arbitrarily:

$${}^{a}\boldsymbol{d}_{2n} = K \circ {}^{b}\boldsymbol{d}_{2n} \tag{8}$$

We now need to update the covariance matrix ${}^{b}\Sigma_{d_2 d_2}$. As the quaternion multiplication $\mathbf{q}_1\mathbf{q}_2$ is bilinear, we may write

$$\mathbf{q}_1\mathbf{q}_2 = M(\mathbf{q}_1)\,\mathbf{q}_2 \tag{9}$$

with $M(\mathbf{q}_1) = \partial(\mathbf{q}_1\mathbf{q}_2)/\partial\mathbf{q}_2$. This enables us to convert (7) to

$$\begin{bmatrix} {}^{a}\boldsymbol{x}_{1n} \\ {}^{a}\mathbf{q}_{1n} \end{bmatrix} = \underbrace{\begin{bmatrix} \lambda_K R(\mathbf{q}_K) & 0_{3\times4} \\ 0_{4\times3} & M(\mathbf{q}_K) \end{bmatrix}}_{J_{K_n}} \cdot \begin{bmatrix} {}^{b}\boldsymbol{x}_{2n} \\ {}^{b}\mathbf{q}_{2n} \end{bmatrix} + \begin{bmatrix} \boldsymbol{t}_K \\ \mathbf{0} \end{bmatrix} \tag{10}$$

and use the $7N \times 7N$ block diagonal matrix $J_K = [J_{K_n}]$ to get

$${}^{a}\Sigma_{d_2 d_2} = J_K \cdot {}^{b}\Sigma_{d_2 d_2} \cdot J_K^{\mathsf{T}} \tag{11}$$

[1] Given at least two frames, we can compute a spatial similarity transformation by taking the translation from the two projection centers, the rotation from the two quaternions and the scale from the difference of two coordinates, for example. It should though be noted that direct solutions may lead to clearly non-optimal estimates, especially when the stochastic structure coded in the covariance matrices is complex and contains strong correlations between the cameras.

4.2 The S-transformation

In our context, the S-transformation is a differential non-stochastic similarity transformation of the datasets ${}^a\boldsymbol{d}_1$ and ${}^a\boldsymbol{d}_2$ into a prespecified coordinate system s, derived as a linearized formulation of the remaining spatial similarity transformation including the rotation parameters. To obtain a linearized model, we may parametrize the rotational parts of this transformation with three parameters representing small angles, which is admissible due to the preceding K-transformation. For each frame n, we thus have

$$ {}^s\boldsymbol{d}_{in} = \Delta S \circ {}^a\boldsymbol{d}_{in} \qquad i = \{1,2\} \tag{12}$$

where $\Delta S(\Delta\boldsymbol{t}, \Delta\boldsymbol{\omega}, \Delta\lambda)$ is the designated transformation applied to the K-transformed data, and ${}^s\boldsymbol{d}_{1n}$ and ${}^s\boldsymbol{d}_{2n}$ are given parameters. We use a full $7N \times 7$ Jacobian, given as the block diagonal matrix

$$ \boldsymbol{A} = [\boldsymbol{A}_n], \qquad \boldsymbol{A}_n = \begin{bmatrix} \boldsymbol{I}_3 & -\boldsymbol{S}({}^a\boldsymbol{x}_{2n}) & {}^a\boldsymbol{x}_{2n} \\ \boldsymbol{0}_{4\times 3} & \boldsymbol{A}_{\Delta\omega_n} & \boldsymbol{0}_{4\times 1} \end{bmatrix} \tag{13}$$

where

$$ \boldsymbol{A}_{\Delta\omega_n} = \begin{bmatrix} -\frac{1}{2}\,{}^a\vec{\boldsymbol{q}}_{2n}^{\mathsf{T}} \\ \frac{1}{2}\,[\,{}^aq_{2n}\boldsymbol{I}_3 - \boldsymbol{S}({}^a\vec{\boldsymbol{q}}_{2n})] \end{bmatrix} \tag{14}$$

with ${}^a\mathbf{q}_{2n} = [\ {}^aq_{2n} \quad {}^a\vec{\boldsymbol{q}}_{2n}^{\mathsf{T}}\ ]^{\mathsf{T}}$ and $\boldsymbol{S}(\boldsymbol{x})$ the skew-symmetric matrix of a 3D vector $\boldsymbol{x}$. A weight matrix $\boldsymbol{W}_s$ captures the weight of each element in the complete parameter sets to encode which of the prespecified frames are used for determining the transformation. This enables us to finally define the S-Matrix

$$ {}^s\boldsymbol{S} = \boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}^{\mathsf{T}}\boldsymbol{W}_s\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{W}_s \tag{15}$$

and obtain the S-transformation for both datasets $i = \{1,2\}$

$$ {}^s\boldsymbol{d}_i = {}^s\boldsymbol{S}\ {}^a\boldsymbol{d}_i \qquad {}^s\Sigma_{d_id_i} = {}^s\boldsymbol{S}\ {}^a\Sigma_{d_id_i}\ {}^s\boldsymbol{S}^{\mathsf{T}} \quad . \tag{16}$$

The result of the S-transformation is indicated with a left superscript s.

5 CONSISTENCY BETWEEN SETS OF FRAMES

5.1 Consistency between two corresponding datasets from different methods

After applying the S-transformation, the corresponding sets of orientation parameters are given in a common coordinate system and gauge. We can therefore directly compute c as introduced in eq. (4) and (5) by inserting the parameters resulting from (16). c is easy to interpret: It is expected to be one on average in case that the form deviation of the two datasets matches their internal precision, and its square is $F_{6N-7,\infty}$-distributed. As desired, higher form deviations yield higher values for c, indicating a lower consistency, and vice versa. We propose to discard datasets where c is clearly larger than 1, i. e. by using a threshold T_c. This threshold should be $\sqrt{F_{\alpha,R,\infty}}$ with α a small percentage, i. e. 0.001 in case the different data sets are statistically independent. Usually the stochastic component of an algorithm does not lead to a statistically independent solution, e. g. RANSAC only randomly selects observations out of a given fixed set.

Pennec and Thirion (1995) state that the error model for estimating rigid transforms should be a transformation itself. Transferring this concept to our particular problem, we may consider the two sets of frames introduced in (1) as consistent if they do not differ significantly by a similarity transformation. More precisely, in case of perfect consistency we may expect

$$ E\left(\begin{bmatrix} {}^a\boldsymbol{x}_{1n} \\ {}^a\mathbf{q}_{1n} \end{bmatrix}\right) = \begin{bmatrix} \lambda\boldsymbol{R}(\mathbf{q})E\left({}^b\boldsymbol{x}_{2n}\right) + \boldsymbol{t} \\ \mathbf{q}E\left({}^b\mathbf{q}_{2n}\right) \end{bmatrix} \tag{17}$$

with observations $\boldsymbol{l} = \left\{{}^a\boldsymbol{x}_{1n}, {}^a\mathbf{q}_{1n}, {}^b\boldsymbol{x}_{2n}, {}^b\mathbf{q}_{2n}\right\}$, unknowns λ, $\mathbf{q}$ and $\mathbf{t}$ and constraints on the length of the quaternion. Using this as an error model, we get as an alternative formulation of (5)

$$ c = \sqrt{\frac{\widehat{\boldsymbol{v}}^{\mathsf{T}}\Sigma_{ll}^{-1}\widehat{\boldsymbol{v}}}{R}} \tag{18}$$

with corrections $\widehat{\boldsymbol{v}}$ for the observations $\boldsymbol{l}$. Eq. (18) is equivalent to estimating a small transformation $\boldsymbol{S}$ from the parameters after the K-Transformation, with contradictions ${}^a\boldsymbol{v}_1$ and ${}^a\boldsymbol{v}_2$ due to the remaining geometric differences:

$$ ({}^a\boldsymbol{d}_1 + {}^a\boldsymbol{v}_1) - \boldsymbol{S} \circ ({}^a\boldsymbol{d}_2 + {}^a\boldsymbol{v}_2) = 0 \tag{19}$$

Due to the differential parameters, we can simplify (19) using the Jacobian introduced in (13) and get

$$ \underbrace{\begin{bmatrix} {}^a\boldsymbol{x}_{1n} - {}^a\boldsymbol{x}_{2n} \\ {}^a\mathbf{q}_{1n} - {}^a\mathbf{q}_{2n} \end{bmatrix}}_{\mathrm{d}\boldsymbol{l}_n} + \underbrace{\begin{bmatrix} {}^a\boldsymbol{v}_{x_{1n}} - {}^a\boldsymbol{v}_{x_{2n}} \\ {}^a\boldsymbol{v}_{\mathbf{q}_{1n}} - {}^a\boldsymbol{v}_{\mathbf{q}_{2n}} \end{bmatrix}}_{\widehat{\boldsymbol{v}}_n} = \boldsymbol{A}_n \underbrace{\begin{bmatrix} \mathrm{d}\boldsymbol{t} \\ \mathrm{d}\boldsymbol{r} \\ \mathrm{d}\lambda \end{bmatrix}}_{\mathrm{d}\widehat{\boldsymbol{x}}_n} \tag{20}$$

Note that while the rotational parts of the frames are represented as quaternions, the rotational part $\mathrm{d}\boldsymbol{r}$ of the transformation is a 3-parameter representation in small angles. This allows us to solve the Gauss-Helmert-model without additional constraints on the unknowns. With

$$ \boldsymbol{S} = \boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}^{\mathsf{T}}\Sigma_{ll}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathsf{T}}\Sigma_{ll} \tag{21}$$

we obtain the negative residuals $-\widehat{\boldsymbol{v}} = \mathrm{d}\boldsymbol{l} - \boldsymbol{A}\cdot\mathrm{d}\widehat{\boldsymbol{x}} = \boldsymbol{S}\boldsymbol{l}$. Comparing (21) and (15), we see that the S-transformation can obviously be interpreted as determining the negative residuals of a differential similarity transformation. The difference between $\boldsymbol{S}$ and ${}^s\boldsymbol{S}$ is the weight matrix, which here is based on the inverse of the covariance matrix $\Sigma_{ll} = {}^a\Sigma_{d_1d_1} + {}^a\Sigma_{d_2d_2}$.

Observe that Σ_{ll} may have a rank deficiency, due to the unknown length of the quaternions and possibly due to the gauge of the sets of cameras. We therefore need to enforce regularity in Σ_{ll} without affecting the result. To overcome singularities caused by the gauges, we add the term $\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}$ which according to Rao (1967) exhibits the desired properties. Furthermore we account for the missing constraints on the length of the quaternions by adding $\boldsymbol{H}\boldsymbol{H}^{\mathsf{T}}$ with $\boldsymbol{H} = \mathrm{null}(\Sigma_{ll})$ in case the quaternions are normalized to 1. $\boldsymbol{H}$ can be given explicitly as the block diagonal matrix

$$ \boldsymbol{H} = [\boldsymbol{H}_n], \quad \boldsymbol{H}_n = \begin{bmatrix} \boldsymbol{0}_{3\times 1} \\ {}^a\mathbf{q}_{2n} \end{bmatrix} \quad , \tag{22}$$

thus we actually use $\Sigma'_{ll} = \Sigma_{ll} + \boldsymbol{H}\boldsymbol{H}^{\mathsf{T}} + \boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}$ in (21).

5.2 Consistency between repeated estimates of a stochastic method containing a random subprocedure

For evaluating the consistency of repeated estimates from a method containing random components, it is reasonable to compare the variation of the estimated parameter vectors with the average internal precision, given by the covariance matrices. We assume the repeated estimates to be uncorrelated, otherwise the following derivation yields only an approximate measure.

The parameter variation is given by the empirical standard deviation of the projection centers and the quaternions. With the mean values $\overline{\boldsymbol{x}}_n$ and $\overline{\mathbf{q}}_n$ computed across all samples K, we get with N the number of cameras

$$ \widehat{\epsilon}^2_{\boldsymbol{x}} = \frac{1}{3K(N-1)}\sum_{k=1}^{K}\sum_{n=1}^{N} |\,{}^s\boldsymbol{x}_{kn} - \overline{\boldsymbol{x}}_n|^2 \tag{23}$$

$$ \widehat{\epsilon}^2_{\mathbf{q}} = \frac{1}{3K(N-1)}\sum_{k=1}^{K}\sum_{n=1}^{N} |\,{}^s\mathbf{q}_{kn} - \overline{\mathbf{q}}_n|^2 \tag{24}$$

Furthermore we compute the average theoretical standard deviations from the covariance matrices ${}^{s}\Sigma_{d_k d_k}$ as

$$\overline{\sigma}^2_{\boldsymbol{x}} = \frac{1}{3NK}\sum_{k=1}^{K}\sum_{n=1}^{N}\sigma^2_{x_{kn}} + \sigma^2_{y_{kn}} + \sigma^2_{z_{kn}} \quad (25)$$

$$\overline{\sigma}^2_{\mathbf{q}} = \frac{1}{3NK}\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{i=0}^{3}\sigma^2_{q_{i,kn}} \quad (26)$$

The consistency of K samples is then derived as the mean ratio

$$c_s^2 = \frac{1}{2}\left(\frac{\widehat{\epsilon}^2_{\boldsymbol{x}}}{\overline{\sigma}^2_{\boldsymbol{x}}} + \frac{\widehat{\epsilon}^2_{\mathbf{q}}}{\overline{\sigma}^2_{\mathbf{q}}}\right) \quad (27)$$

Again we assume the samples to be uncorrelated here, and require c_s to be lower than $T_{c_s} = \sqrt{F_{\alpha,6K(N-1),\infty}}$.

An alternative formulation of c_s may be directly derived from the consistency measure c introduced in section 5.1 by computing c for each pair of the samples and then measure the variation of these values. This should yield a comparable quantity.

6 PRECISION LEVEL BETWEEN DATASETS

Given two corresponding datasets, each of them with a valid consistency computed w. r. t. the same reference dataset, we finally finish the benchmark test by comparing their difference in precision level w. r. t. the reference dataset. Our aim is to use the ratio of standard deviations ${}^{a}\sigma_f$ and ${}^{b}\sigma_f$ of arbitrary functions $f(\boldsymbol{d})$ of the two sets of frames, once calculated with ${}^{a}\Sigma_{d_1 d_1}$ and once calculated with ${}^{b}\Sigma_{d_2 d_2}$. The ratios can be computed after a transformation into the same coordinate system s, so we choose the parameters resulting from (16) again and determine the generalized eigenvalues r^2 from $|{}^{s}\Sigma_{d_1 d_1} - r^2 \ {}^{s}\Sigma_{d_2 d_2}| = 0$. The maximum ratio of the standard deviations of an arbitrary function f of the parameters then is bounded (Baarda, 1967):

$$\frac{\sigma_{f(d)}^{\ {}^{s}\Sigma_{d_1 d_1}}}{\sigma_{f(d)}^{\ {}^{s}\Sigma_{d_2 d_2}}} \leq r_{\max} = \max_i r_i \quad (28)$$

As a measure for the precision level p, it would at first be reasonable to choose

$$p' = \overline{r} = \frac{\sum_{i=1}^{m} r_i}{m} \quad \text{or} \quad p'' = \mathrm{e}^{\overline{\ln r}} = \sqrt[m]{\prod_i r_i} \quad (29)$$

In fact though, (29) makes no distinction on the shape of the compared error ellipsoids, and may indicate similarity between clearly flat ellipsoids and a unit sphere. Following Förstner and Moonen (1999), the distance of the two covariance matrices can be defined as

$$d^2 = \frac{\sum_{i=1}^{m} \ln^2 r_i}{n} = \overline{\ln r^2} \geq 0 \quad (30)$$

It can be shown that d is a metric when comparing two covariance matrices (Förstner and Moonen, 1999). It is the mean quadratic deviation of r_i from 1, using the deviation $l_i = \ln r_i$ from 0. However we want to know the average quadratic deviation of the ratio r_i from 1, which is e^d, yielding the precision level

$$p := \mathrm{e}^{\sqrt{\overline{\ln r^2}}} \geq 1 \quad . \quad (31)$$

It can be seen that the ratios r_i and $1/r_i$ of standard deviations are weighted as equally deviating from ratio 1 if we use the logarithm before averaging. The precision level $p = 1$ is only reached in case the covariance matrices are equal, otherwise $p > 1$, e. g. $p = 1.05$ meaning the standard deviations calculated with Σ_2 on average deviate by 5% from those calculated with Σ_1. The formulation compares to the one given in (29) by a factor 2 in the exponent.

Figure 2: Image sequence used for the experiment

Pyramid Level	#Object points	c	p
0	15565	0.0	./.
1	12914	1.1	1.3
2	1250	1.2	14.3
3	384	1.7	66.5

Table 1: Four different results of our own software AURELO computed on different pyramid levels of the images in Figure 2.

7 EXPERIMENTAL RESULTS

The proposed approach can be used for different tasks: (1) comparing two methods, (2) comparing different versions of the same method, e. g. when changing control parameters and (3) comparing different results of an algorithm with a stochastic component. We present an experiment with our own software AURELO (Läbe and Förstner, 2006) to show the feasibility for evaluating the effect of image resolution on the quality of the results. An elaborate application of our benchmarking scheme for quality analysis can be found in Läbe et al. (2008).

The image sequence used is depicted in Figure 2. We present four series of estimates based on this sequence but using different image resolutions. The first series was computed on the original resolution (3008×2000) and will be used as a reference dataset. The three other series were computed on increasing pyramid levels, each level scaled down by a factor of two w. r. t. the previous one.

As our software uses a RANSAC scheme, we computed 10 repeated estimates with constant parameter settings for each series. For the second pyramid level, we obtained $\widehat{\epsilon}_{\boldsymbol{x}} = 7.67 \cdot 10^{-6}$, $\widehat{\epsilon}_{\mathbf{q}} = 8.1 \cdot 10^{-4}$, $\overline{\sigma}_{\boldsymbol{x}} = 4.8 \cdot 10^{-5}$ and $\overline{\sigma}_{\mathbf{q}} = 0.006$. Thus the consistency of the repeated estimates c_s is 0.1467, which is clearly smaller than the threshold T_{c_s}. We obtained similar results for the other series and thus reason that the datasets can be used for further evaluation.

The next step is the computation of the consistency c w. r. t. the reference dataset for each of the other series. If the theoretical covariance matrices computed by AURELO reflect the true accuracy situation related to the form similarity of estimated frames, we should obtain values not significantly deviating from 1. The results in Table 1 show that this is clearly true for the first pyramid levels. The smallest image resolution however yields $c = 1.7$, indicating that the software looses reliability on very small image resolutions. This either may be caused by a too optimistic covariance matrix for the observations or by outliers not detected by the algorithm. Regarding the proposed threshold T_c we can conclude that the consistency is satisfying for pyramid levels 1 and 2 and at least acceptable for pyramid level 3.

The final step consists of comparing the precision levels of the datasets, i. e. using p computed w. r. t. the reference to identify the most favorable estimate. As the number of pixels decreases by four and the accuracy of observed points by two on average, we expect the precision to decrease approximately by $\sqrt{4} \cdot 2 = 4$ between successive pyramid levels. Indeed this factor can be observed in Table 1: While the first pyramid level yields $p = 1.3$, thus better than 4 as expected, the successive series yield values close to 16 and 64, respectively. The particularly good behavior of pyramid level one has been observed in many of our previous experiments. Hence altogether the proposed benchmark test not only identified the expected winner, but also yielded measures that we are able to explain by design of the datasets.

8 CONCLUSION AND REMARKS

We defined statistically sound measures for comparing corresponding sets of orientation parameters estimated by automatic bundle adjustment procedures. They can be used to compare different methods, the same method with different control parameters and - if required - multiple samples of stochastic methods for bundle adjustment. The measures are computed after a transformation of the parameters into a well defined coordinate system, using the concept of K- and S-transformations that we made explicit for sets of camera frames.

The proposed measures are based on the consistency of the form deviation of corresponding camera frames with their theoretical precision as well as on the difference in precision level related to a reference dataset. We proposed a complete benchmarking scheme that addresses both aspects and also considers consistency of stochastic methods with random components over repeated estimates. It has been successfully applied to evaluate results from our own software and yielded meaningful results.

There are still some problems that we want to address. Using a simple direct solution for the K-transformation may sometimes yield really poor approximate values. Furthermore, numerical issues play an important role if the theoretical precision is very high. We are examining ways of conditioning the datasets to gain more reliable results with difficult input data. Finally, it may be meaningful to derive a consistency measure similar to c_s for evaluating the variation in precision level of repeated estimates when testing rather unstable bundle adjustment methods with very pronounced random components.

Benchmarking of bundle adjustment procedures of course needs to address other issues, such as computation time, the ability for selfdiagnosis w. r. t. systematic errors or the ease of including other information, such as external observations, e. g. GPS, into the estimations, which is out of the scope of the paper.

References

Baarda, W., 1967. S–Transformations and Criterion Matrices. Publications on Geodesy, Vol. 5(1), Netherlands Geodetic Commission, Delft, Netherlands.

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 Conference on Fast Processing of Photogrammetric Data, Interlaken, pp. 281–305.

Förstner, W. and Moonen, B., 1999. A metric for covariance matrices. In: V. S. S. F. Krumm (ed.), Quo vadis geodesia ...?, Festschrift for Erik W. Grafarend on the occasion of his 60th birthday, TR Department of Geodesy and Geoinformatics, Stuttgart.

Läbe, T. and Förstner, W., 2006. Automatic relative orientation of images. In: Proceedings of the 5th Turkish-German Joint Geodetic Days, Berlin.

Läbe, T., Dickscheid, T. and Förstner, W., 2008. On the quality of automatic relative orientation procedures. In: Proc. of the 21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS), Beijing, China.

Lourakis, M. I. and Argyros, A. A., 2004. Design and implementation of a sparse bundle adjustment software library based on the Levenberg-Marquardt algorithm. Technical report, Institute of Computer Science - FORTH, Heraklion, Crete, Greece.

Lourakis, M. I. and Argyros, A. A., 2005. Is Levenberg-Marquardt the most efficient optimization algorithm for implementing bundle adjustment? In: IEEE International Conference on Computer Vision, Vol. 2, Beijing, China, pp. 1526–1531.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. In: International Journal of Computer Vision, Vol. 60(2), pp. 91–110.

Martinec, D. and Pajdla, T., 2002. Structure from many perspective images with occlusions. In: 7th European Conference on Computer Vision, Copenhagen, Denmark, pp. 355–369.

Matas, J., Chum, O., Urban, M. and Pajdla, T., 2004. Robust wide baseline stereo from maximally stable extremal regions. Image and Vision Computing 22, pp. 761–767.

Mayer, H., 2005. Robust least-squares adjustment based orientation and auto-calibration of wide-baseline image sequences. In: ISPRS Workshop BenCOS 2005, Bejing, China, pp. 11–17.

Mikolajczyk, K. and Schmid, C., 2004. Scale & affine invariant interest point detectors. International Journal of Computer Vision 60(1), pp. 63–86.

Molenaar, M., 1981. A further inquiry into the theory of S–transformations and criterion matrices. Publications on Geodesy, Vol. 7(1), Netherlands Geodetic Commission, Delft, Netherlands.

Pennec, X. and Thirion, J.-P., 1995. Validation of 3D Registration Methods Based on Points and Frames. In: 5th International Conference on Computer Vision, Boston, MA, USA, pp. 557–562.

Pollefeys, M., Koch, R., Vergauwen, M. and Van Gool, L., 2000. Automated reconstruction of 3D scenes from sequences of images. In: ISPRS Journal Of Photogrammetry And Remote Sensing, Vol. 55(4), pp. 251–267.

Rao, C. R., 1967. Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In: Fifth Berkeley Symposium, Vol. 1, pp. 355–372.

Roth, D. G., 2004. Automatic correspondences for photogrammetric model building. In: Proceedings of the XXth ISPRS Congress, Istanbul, Turkey.

Schaffalitzky, F. and Zisserman, A., 2002. Multi-view matching for unordered image sets. In: Proceedings of the 7th European Conference on Computer Vision, London, UK, pp. 414–431.

Triggs, B., McLauchlan, P. F., Hartley, R. I. and Fitzgibbon, A. W., 2000. Bundle adjustment - a modern synthesis. In: ICCV '99: Proceedings of the International Workshop on Vision Algorithms, Springer-Verlag, London, UK, pp. 298–372.