ON THE QUALITY OF AUTOMATIC RELATIVE ORIENTATION PROCEDURES

Thomas Läbe, Timo Dickscheid and Wolfgang Förstner

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

Commission III/1

KEY WORDS: Sensor orientation, comparative analysis, accuracy assessment, matching

ABSTRACT:

This paper presents an empirical investigation into the quality of automatic relative orientation procedures. The results of an in-house developed automatic orientation software called AURELO (Läbe and Förstner, 2006) are evaluated. For this evaluation a recently proposed consistency measure for two sets of orientation parameters (Dickscheid et al., 2008) and the ratio of two covariances matrices is used. Thus we evaluate the consistency of bundle block adjustments and the precision level achievable. We use different sets of orientation results related to the same set of images but computed under differing conditions. As reference datasets results on a much higher image resolution and ground truth data from artificial images rendered with computer graphics software are used. Six different effects are analysed: varying results due to random procedures in AURELO, computations on different image pyramid levels and with or without points with only two or three observations, the effect of replacing the used SIFT operator with an approximation of SIFT features, called SURF, repetitive patterns in the scene and remaining non-linear distortions. These experiments show under which conditions the bundle adjustment results reflect the true errors and thus give valuable hints for the use of automatic relative orientation procedures and possible improvements of the software.

1 INTRODUCTION

Orienting images is one of the basic tasks in photogrammetry. For aerial images the full automation of this task can be considered as solved, but for close range photogrammetry full automatic orientation still is a vital research area (cf. the work of Pollefeys (Sinha and Pollefeys, 2004), Nister (Engels et al., 2006), Zisserman and Schaffalitzky (Schaffalitzky and Zisserman, 2002)). Due to errors introduced while solving the recognition task, special investigations into the factors influencing the accuracy which are not relevant in classical non-automatic bundle adjustment have to be considered, e.g. texture, repetitive structures, rotation of the images, use of all or less points in the adjustment. Depending on the used algorithms, there are various parameters which have an influence on the accuracy of the result. This paper investigates the quality of automatic bundle orientation software, especially the effect of various factors onto the quality of the result. As there is no general method for automatic mensuration of control points, we investigate the relative orientation of multiple images instead. The findings can be generalized for a wide range of scenarios and applications.

2 METHOD FOR QUALITY CHECK

To evaluate the result of an automatic orientation procedure we need to compare the orientation parameters between two datasets directly. Automatic methods may be stochastic or deterministic, depending on whether they call a random number generator, e. g. for a RANSAC procedure, or not. In case the algorithm contains a stochastic component, we cannot use the 3D object points, as different runs may choose different points and we investigate in relative orientation procedures without a subsequent absolute orientation. Therefore we follow the approach of (Dickscheid et al., 2008) which compares the orientation parameters directly. The measures may be used (1) for evaluating the variations due to the randomness of the algorithm and (2) for comparing a bundle adjustment result with a reference dataset or with each other. A reference dataset in this context is a dataset with superior accuracy and thus may be a result of just the same algorithm, e.g. on an image pyramid level with a higher resolution.

The approach introduces a so-called *consistency measure* c: The two parameter sets should differ by a spatial similarity transformation. This transformation is estimated and the square root of the variance factor of the final estimation is the consistency measure

$$c = \sqrt{\Omega/R}$$

with the Mahalanobis distance Ω of the two datasets in a common coordinate system and with the redundancy R, which is $6N - 7$ for sets with N images. The full possibly singular covariance matrices of the orientation parameters of both datasets are rigorously taken into account in this congruence test. The advantage of this approach is that the comparison is summarized in a single scalar value. A value of $c = 1$ means that the differences of the two sets are (on average) consistent with the precision given by the orientation procedures. If $c < 1$, the accuracies are too pessimistic, if $c > 1$, the accuracies are too optimistic. Note that the reliability of the consistency measure depends on the redundancy in the estimation of the transformation parameters. Therefore we expect reliable consistency estimates for datasets with more than three or four images. We give the consistency measure for every experiment we describe in this paper.

In addition to c, we use p and r_{max} from (Dickscheid et al., 2008): The two values compare two covariance matrices and both depend on the generalized eigenvalues λ_i. We compare the two covariance matrices of the two datasets after applying the estimated spatial similarity transformation. Thus both covariance matrices are in the same coordinate system and can be compared. p is the average ratio of the covariance matrices. If $p = 1$, the covariance matrices are identical, if $p > 1$ there is an (average) deviation in precision. Note that, as p is a metric, $p >= 1$ and p is symmetrically. Thus p gives no information about which covariance matrix is smaller, in other words, which accuracies are better. The second value r_{max} is the maximum ratio of the standard deviations: $r_{max} \leq \sqrt{\max_i \lambda_i}$ and thus indicates the maximum difference between the matrices. Note that r_{max} is not symmetric: if $r_{max} < 1$ then the precision of the second covari-

ance matrix is better, otherwise it is worse for some function of the parameters.

If $c \neq 1$ the covariance matrices are not consistent with the differences of the two datasets. When arguementing with p and r_{max} in this case it is reasonable not only to compare the given covariance matrices, but to also take the observed differences into account. Thus $\widehat{p} = \widehat{c}\, p$ and $\widehat{r}_{max} = \widehat{c}\, r_{max}$ are also used in these cases. These values represent the actual occurred accuracies. $\widehat{r}_{max}$ is maximum standard deviation of arbitrary function of parameters assuming the estimate $\widehat{c}$ for c to be realistic.

3 ORIENTATION SOFTWARE

For our investigations we use an in-house developed orientation software for the wide-baseline case of calibrated cameras called AURELO. The following paragraph gives a brief overview of the software. Details can be found in (Läbe and Förstner, 2006).

The first step in AURELO is the extraction of feature points on all given images. The SIFT-descriptor (Lowe, 2004) or the SURF-descriptor (Bay et al., 2006) can be used to describe and match the points of an image pair. The SURF descriptor is a fast adaption of the SIFT-descriptor. We use the implementation of D. Lowe for the SIFT descriptor [1] and the implementation of the ETH Zurich, Switzerland, for SURF [2]. Because the descriptors are rotation- and scale-invariant, a large number of possible configurations can be handled. No prior information about the overlapping parts or the sequence of the images need to be given. Therefore a matching is done between all possible image pairs. Relative orientations for each pair of images are computed with the help of a RANSAC procedure based on the 5-point solution proposed by D. Nister (Nister, 2004). It should be noted that, as RANSAC uses a random number generator, multiple runs of the software may lead to different results. The best pairwise relative orientations are linked together to generate an input for a concluding bundle adjustment. Thus the resulting orientation parameters are given with a full covariance matrix. Note that by default all points are used in this adjustment, independent of the number of observations per object point.

4 EXPERIMENTAL RESULTS

4.1 Description of the Datasets and Experiments

We use a wide range of datasets in our experiments because many effects are highly dependent on the geometric configuration and on the amount of texture in the images. Due to the limited space, only a part of the results can be presented in this paper, but we verified the observed effects in a number of further, unpublished experiments. The datasets we used for the data here are summarized in Table 1, see also Figure 1. Two cameras were used to take the images: Nikon D70s with a 20mm lens (6 Megapixel images, datasets A,B,D) and HP photosmart 435 (3 Megapixel, datasets C,E).

We calibrated the cameras and corrected the nonlinear distortion by rectifying the images. All experiments are done with the same settings for the control parameters, unless otherwise noted. The standard parameters imply that the feature extraction is done with the operator by D. Lowe.

Figure 1: Sample images of datasets B,E,F (from left to right).

[1] http://www.cs.ubc.ca/∼lowe/keypoints
[2] http://www.vision.ee.ethz.ch/∼surf

dataset	# img.	description
A	6	entrance of a large building with stairs, medium texture
B	6	outdoor stairs with rich texture, matches in background which are far away from cameras
C	4	poster and box on a table, different perspectives and rotations of the camera
D1	12	facade with rich texture, one strip with nearly parallel viewing directions and high overlap (ca. 90%)
D2	9	same configuration as D1, but with less overlap (ca.75%)
E	3	indoor scene with a regular grid on the wall
F	5	rendered images with 1 Megapixel image size: two walls with texture on it. The texture varies between artificial repetitive patterns and real texture.

Table 1: Overview of the datasets used in the experiments.

In the tables we give additional results as given by the orientation procedure:

- the number of object and image points of the final bundle block adjustment.
- the outlier rate of the matches: Due to the relative orientation of an image pair computed in the RANSAC loop, outliers in the matched points of this image pair are detected. The average rate of these outliers over all image pairs is given in the tables.
- $\sigma_{x'}$: average precision of the image coordinates in pixel. As all image observations currently are used with the same weight and thus the weight matrix for the observations is the unit matrix and $\sigma_{apriori} = 1$, $\sigma_{x'} = \widehat{\sigma}_0$, with $\widehat{\sigma}_0$ being the square root of the variance factor of the bundle adjustment.
- σ_{X_O}: The average precision of the projection centres over all N images (using $\widehat{\sigma}_0$ and the inverse of the normal equation matrix): $\sigma^2_{X_0} = \sum_n \left(\sigma^2_{X_n} + \sigma^2_{Y_n} + \sigma^2_{Z_n}\right)/N$. Note that all datasets have an arbitrary scale, because no absolute orientation has been done. The unit of the object coordinate system is defined by the distance from the first to the second camera. All datasets of one investigation have been scaled to the same scale to ensure that the values can be compared to each other.
- $\sigma_{\omega\phi\kappa}$: average precision of the orientation angles in gon.
- σ_{OBJ}, $\sigma_{OBJ,MAX}$: average and maximal precision of the object points due to the bundle adjustment result. Also scaled to the same scale in one table.

4.2 Consistency over Multiple Runs

Due to the fact that a random procedure for the generation of approximate orientation values is used in AURELO, the user gets different results when starting the program multiple times with identical parameter settings and input images. The variations of the orientation parameters should be small compared to their accuracies.

To test the influence of the random procedure, we run 5 datasets yielding 10-42 samples. The results of these tests can be found in Table 2. The variations of the orientation parameters over multiple runs are shown by their standard deviation ϵ_{X_0} and $\epsilon_{\omega\phi\kappa}$ with

$$\epsilon^2_{X_0} = \frac{1}{N(S-1)} \sum_n \sum_s \left| X_{sn} - \bar{X}_n \right|^2$$

and

$$\epsilon^2_{\omega\phi\kappa} = \frac{1}{N(S-1)} \sum_n \sum_s \left| \begin{pmatrix} \omega_{sn} - \bar{\omega}_n \\ \phi_{sn} - \bar{\phi}_n \\ \kappa_{sn} - \bar{\kappa}_n \end{pmatrix} \right|^2$$

dataset	plev	#runs	R	ϵ_{X_0}	$\epsilon_{\omega\phi\kappa}$ [gon]	$\bar{\sigma}_{X_0}$	$\bar{\sigma}_{\omega\phi\kappa}$ [gon]	$\epsilon_{X_0}/\bar{\sigma}_{X_0}$	$\epsilon_{\omega\phi\kappa}/\bar{\sigma}_{\omega\phi\kappa}$	max. c	mean c
A	0	42	29	0.000074	0.00162	0.000537	0.00564	0.14	0.29	1.39	0.63
B	0	10	29	0.000029	0.00404	0.000015	0.00247	1.91	1.63	5.34	2.84
	1	10	29	0.000014	0.00137	0.000027	0.00429	0.53	0.32	0.95	0.56
	2	10	29	0.000036	0.00271	0.000081	0.01144	0.45	0.24	1.11	0.47
C	1	10	17	0.000101	0.02196	0.000075	0.02529	1.34	0.87	3.28	1.66
D1	2	10	65	0.000022	0.00220	0.001019	0.03895	0.02	0.06	0.21	0.09
D2	2	10	47	0.000023	0.00484	0.000144	0.03621	0.16	0.13	0.44	0.28

Table 2: Comparison of multiple runs of AURELO with identical parameter settings. Plev: pyramid level on which the orientation has been computed (0: original images, 1: first pyramid level, ...). R: redundancy for the similarity transformation between the samples, $R = 6N - 7$ with N images. ϵ_{X_0} ,$\epsilon_{\omega\phi\kappa}$: standard deviation of the projection centres and the orientation angles calculated over the multiple runs. $\bar{\sigma}_{X_0}$, $\bar{\sigma}_{\omega\phi\kappa}$: average precision for projection centres and orientation angles due to the covariance matrix of the orientation parameters over all samples. Max. c, mean c: Maximal and mean consistency measure of all consistency measures between all samples (pairwise).

computed over S multiple runs. Note that all results of a dataset are transformed to the same datum in order allow a comparison. The ratio of the standard deviations ϵ_{X_0}, $\epsilon_{\omega\phi\kappa}$ and the mean accuracies $\bar{\sigma}_{X_0}$,$\bar{\sigma}_{\omega\phi\kappa}$ should be small, at least less than 1.0. (These two values can be combined to one value c_s, see (Dickscheid et al., 2008).) This is fulfilled for most of the tests, except for dataset B on pyramid level 0 and dataset C. These unexpected results can be explained with the weak treatment of outliers in AURELO: The decision about inlier and outlier is done only with the epipolar geometry calculated from the approximate orientation values. These values are computed in the RANSAC loop with only 5 homologous points and thus a wrong decision concerning outlier/inlier is possible. No robust adjustment is used, so that small outliers still may present in the final result. This result thus reveals a deficiency of the algorithm: it does not always yield reliable results due to its randomness, especially on a high precision level. This needs a clarification of the causes, which e. g. could be the non-robustness of the final bundle adjustment.

4.3 Orientation Parameters on Different Pyramid Levels

An important factor for the computation time and memory consumption of the orientation procedure is the size of the input images. Therefore we investigate the decrease of accuracy when calculating the whole orientation procedure on a higher image pyramid level instead on the original image. All thresholds used in AURELO remain constant during the tests. Table 3 shows the results of two datasets which can be regarded representative for the conclusions drawn.

All pyramid levels where we got a useful result are listed. In these two datasets the number of images which could be connected by the orientation procedure nearly remains the same over all levels. If a dataset includes parts with low overlap of the images, a decrease in the number of images can occur in lower pyramid levels as in the two given examples.

The number of object points decreases from pyramid level to pyramid level with the same factor as the number of image points (not listed in the table). Thus, as to be expected due to the identical geometry, the distribution of the number of observations per object point is independent of the pyramid level.

The consistency measure between the first and all other image pyramid levels is not higher than 1.8, the consistency values between all the levels (not listed in the table) are in the same range. Thus the change of the orientation parameters from level to level are in the order of the assumed precision or slightly higher. There is no evidence not to trust the internal precision values, especially when comparing them to each other.

The evaluation of the change in precision when changing the pyramid level needs to be compared with the theoretical expectation: In higher levels we expect a loss in precision for two factors: 1. loss of points and 2. less accurate points due to the reduced resolution. We thus expect the standard deviation of the orientation parameters to increase with the same factor as the measurement accuracy (given with the same pixel size!) and to decrease with the square root of the ratio of the number of object-/image points. So a reasonable theoretical factor between two image pyramid levels l and m is

$$f_{o\sigma} = (\sigma_{0,l}/\sigma_{0,m})\sqrt{N_m/N_l)}$$

with N_l , N_m points in pyramid levels l and m. This factor can be compared to the average ratio p of the two covariances matrices of the orientation parameters. The table shows that the two values are nearly the same. This validates the use of p for practical cases.

Without taking into account the type of point operator, one could assume a factor of 2 in the measurement accuracy between level l and $l+1$ and a factor of 1/4 in the number of observations. Thus a reasonable theoretical factor between levels l and $l+1$ is $2\sqrt{4} = 4$. The table shows that there are sometimes significant differences to that value. The increase in the standard deviations is smaller between the original image and the first pyramid image (max. 2.1). This is also the smallest value. This phenomenon can be observed with other datasets, too. From a practical point of view this is an important finding, because it allows the use of the first image pyramid level without a big loss in accuracy. The large value of p=10.1 of dataset D1 from level 1 to level 2 may be explained by a loss of texture: A large area of the image has the same texture. If on level 2 this texture is not visible any more, the is a high loss in the number of extracted image points.

4.4 Influence of Twofold and Threefold Points on the Result

To reduce the computation time of the bundle adjustment, we may consider to reduce the number of observations used for the reconstruction. One simple possibility would be to ignore object points with a small number of observations, especially object points with only two observations. The consistency of twofold points can't be tested very well, because the second observation may lie at every position on the epipolar line induced by the first observation. Thus there may be undetected false matches, especially when repeated patterns occur along the epipolar lines. These false matches may lead to unstable or at least to less accurate solutions in the bundle adjustment. Thus, if the object points are used in further applications, twofold points may be deleted after the adjustment or not used at all in the adjustment. This is of course only possible if there are enough other n-fold points, $n > 2$, available.

To test the influence of n-fold points, we have computed the bundle adjustment with all points and with all but the 2- and 3-fold points, respectively. The results of three datasets are shown in Table 4. First we want to compare datasets D1 and D2, as they consists of images of the same object. The orientation parame-

dataset	pyramid level	#images in result [%]	outliers of matches [%]	# object points	$\sigma_{x'}$	σ_{X_O}	$\sigma_{\omega\phi\kappa}$	c	$f_{o\sigma}$	p	r_{max}
rich texture	0	12	15.4	19186	0.25	0.0007	0.0026	0.0	-	-	-
(D1)	1	12	11.2	15470	0.15	0.0011	0.0033	1.0	1.3	1.3	1.5
	2	12	12.0	2028	0.30	0.0077	0.0321	1.2	10.9	10.1	19.5
	3	12	23.4	676	0.35	0.0391	0.1708	1.3	4.0	4.3	11.8
	4	12	14.9	294	0.37	0.1419	0.6263	1.5	3.2	3.7	6.7
medium texture	0	6	20.8	2753	0.38	0.0008	0.0050	0.0	-	-	-
(A)	1	6	17.6	1597	0.28	0.0016	0.0107	1.8	2.0	2.1	3.0
	2	6	19.9	686	0.29	0.0063	0.0398	1.6	3.2	3.4	5.6
	3	6	12.2	293	0.41	0.0232	0.1429	1.3	4.3	3.9	6.3
	4	4	8.5	113	0.35	0.0844	0.4833	1.0	2.8	3.0	4.8

Table 3: Quality of the orientation on different image pyramid levels. Consistency measure c: Measure between image pyramid level 0 (original images) and all other levels. p, r_{max}: average/maximal ratio of standard deviations between level l and level $l-1$. $f_{o\sigma}$: theoretical factor for the decrease of accuracy, $f_{o\sigma} = (2\sigma_{0,l}/\sigma_{0,l-1})\sqrt{N_{l-1}/N_l)}$, l =pyramid level, N_l object points on level l.

dataset	points used	# object points	# image points	$\sigma_{x'}$	σ_{X_O}	$\sigma_{\omega\phi\kappa}$	σ_{OBJ}	$\sigma_{OBJ,MAX}$	c	p	r_{max}
D1	all	19186	57554	0.25	0.00071	0.0026	0.00209	0.23867	0.0	-	-
	without 2-fold	9374	37925	0.25	0.00067	0.0028	0.00114	0.00686	0.7	1.2	1.5
	without 2,3-fold	5137	25214	0.25	0.00075	0.0032	0.00079	0.00265	1.0	1.5	3.3
D2	all	15565	42707	0.21	0.00036	0.0023	0.00083	0.04141	0.0	-	-
	without 2-fold	7053	25683	0.21	0.00043	0.0029	0.00052	0.00162	1.2	1.4	4.4
	without 2,3-fold	3393	14703	0.21	0.00059	0.0041	0.00039	0.00096	1.5	2.1	14.8
B	all	37176	96563	0.28	0.00991	0.0021	0.00583	1.53229	0.0	-	-
	without 2-fold	12894	47999	0.25	0.01412	0.0025	0.00374	1.30997	10.6	1.4	2.6
	without 2,3-fold	5029	24404	0.24	0.02185	0.0033	0.00318	0.71910	7.9	1.9	4.3

Table 4: Quality of the Orientation without points with two or three observations. All orientations have been done on the original images. Consistency measure: Measure between orientation with all points and with less points. p, r_{max}: average and maximal factor between covariance matrices of orientation with all points and with less points.

ters of D1 and D2 do not vary significantly with respect to their accuracy: the consistency measure is below 1.5. The accuracy of the image coordinates ($\sigma_{x'}$) remains constant: All points lie on a facade with very good texture and thus the observations of 2-fold and 3-fold points have the same accuracy as the other points. The number of object and image points and hence the computation time for the bundle adjustment decreases significantly, e.g. by a factor of 4 when comparing 'all points' and 'without 2-fold points' of D2. The average ratio of the covariance matrices between 'all points' and 'without 2-fold points' is 1.2 (dataset D1) and 1.4 (dataset D2) which may be an acceptable accuracy loss in most applications. The ratio for 'without 2,3-fold points' of D1 may be also acceptable, but for D2 the value shows a loss in accuracy. Here the maximal ratio has a large value (14.8). This in an indication that without 2- and 3-fold points some images could not be oriented with good accuracy. As the images consist of one strip and D2 has less overlap, the orientation of the images at the beginning and end of the strip are difficult without 3-fold points. We conclude from this observation that for single strips a general deletion of 2- and 3-fold tiepoints may be problematic, for circular arrangements of images this problem would not occur.

Dataset B shows a different behaviour. The consistency measures between the orientation with all points and without 2-fold or 2- and 3-fold points is 10.6 and 7.9, thus the orientation parameters change with respect to their high accuracy level. The average accuracy of the image points becomes slightly better ($\sigma_{x'}$ decreases from 0.28 to 0.25 and 0.24). These two aspects allow the assumption that most outliers were among the 2- and 3-fold points. A visualization of the object points (not shown here) indicates that many twofold points which are far away and lie on twigs of a tree have been deleted. There are points remaining which are very far away, explaining the high maximal value for the standard deviation of the object points. Dataset B has also shown high consistency values for multiple runs (see Table 2). Here again we can draw the conclusion that for this dataset a better outlier detection has to be implemented. The results of the orientation without twofold points may be more accurate due to outliers even though the average ratio p of the covariance matrices is 1.4.

4.5 Using Different Feature Extractors

If the input images are large, e.g. greater than 5 Megapixel, the time for point extraction is an important part of the overall computation time. Therefore we integrated SURF as an alternative to the SIFT operator by D.Lowe. SURF speeds up also the matching by a factor of 2, because the length of the descriptor of a feature point is only 64 compared to 128 when using the original operator. Both operators double the image size before extracting the points. Both operators were used with their standard parameters. Table 5 summarizes a test which shows the influence of the operators.

The examples show that SURF delivers less points in all cases, sometimes only 1/3 of the number of Lowe points. The matched SURF points also contain more outliers than the Lowe points in all cases. The accuracy of the points is worse in datasets A and B when using SURF, so that the average accuracy decreases by a factor of about 4 in these two datasets. This is theoretically the same loss which occurs when using the next image pyramid level (see chapter 4.3). This loss in accuracy must be compared to the computation time: In our tests SURF is faster with a factor of about 4-8, 4-5 on the smaller images of datasets A and E. As the same factor (4-5) can be expected when calculating the Lowe SIFT-points on the next pyramid level, an alternative for these two datasets with the same loss of accuracy to SURF would be to use the next pyramid level with the SIFT operator by D. Lowe. However, there is still the speedup by a factor of 2 during

dataset	Features	#images in result	# object points	outliers of matches [%]	$\sigma_{x'}$	σ_{X_O}	$\sigma_{\omega\phi\kappa}$	c	p	computation time [sec]
B (plevel=0)	Lowe	6	37176	6.1	0.28	0.0099	0.0021	-	-	648
	SURF	6	9181	12.2	0.43	0.0324	0.0066	4.8	3.8	77.3
A (plevel=1)	Lowe	6	1597	17.6	0.28	0.0019	0.0107	-	-	131.6
	SURF	6	555	39.2	0.65	0.0086	0.0438	1.1	4.4	21.3
E (plevel=0)	Lowe	3	196	46.1	0.78	0.0071	0.1090	-	-	105.2
	SURF	3	169	57.7	0.66	0.0145	0.1828	10.0	2.4	23.5

Table 5: Quality of the Orientation calculated with different point extractors. Consistency measure: Measure between orientations with the two point extractors. Computation time: computation time for feature point extractions in seconds.

matching which may be advantageous when choosing SURF.

Dataset E (repetitive patterns) shows a different behaviour: The measurement accuracy of SURF is slightly better, the number of object points is only 13 % smaller. In contrast, the average distance of the covariance matrices of the orientation parameters increases by a factor of 2.4 when using SURF. This may be explained by the fact that the distribution of the observations over the images is better when using the Lowe SIFT operator. But the consistency measure between these two orientations is 10, the orientation parameters show big differences that cannot be explained by gaussian noise in the image coordinates. So far the source of this difference is unknown. Further investigations have to be done to find out which result is more reliable.

4.6 Results with with Repetitive Patterns and Ground Truth

To guarantee that there is no unknown bias in the results of AURELO which is not even visible when comparing results on different image pyramid levels, a comparison with ground truth data is necessary. Therefore we rendered artificial images with an OpenGL program. We combined this experiment with an investigation of different matching strategies and their performance with repetitive patterns. Therefore the scene (dataset F) shows two walls (Figure 1). The walls had been textured partly with an artificial repetitive pattern and partly with rich texture from real images. We varied the amount of the repetitive patterns.

Table 6 shows experiments with two different matching criteria: The first experiments are done with the standard criterion: The ratio of the feature vector distance between the second best and the best match must be below a certain threshold, e.g. 70% in these tests. If there are multiple matches due to repetitive patterns, non of the matches will be used in general. The second part of the Table shows experiments with threshold matching: The distance of the feature vector must be below a defined threshold. Here multiple matches will be used in the RANSAC loop to compute the relative orientation of an image pair. Thus the relative orientation is able to use matches of repetitive patterns, but must detect the false matches as outliers.

In all cases listed the orientation was able to connect all five images. Tests were also done with 100% repetitive patterns. The orientation failed completely in this case or give a connection of two images with very bad accuracies.

Best Matching: The consistency measure c of the experiments with 'best matching' is below 1.8. This shows that there is no bias between the ground truth data and the results of AURELO that is not in the order of the standard deviations of the result. Thus the accuracies given represent the true accuracy situation when using Lowe features and best matching in this image setup (large overlap between the images, no textured background that is very far away).

We can conclude that the orientation procedure is able to cope with up to 40% repetitive patterns with nearly the same accuracy ($\widehat{p}$ between 1.6 and 1.8). As the repetitive patterns are equally distributed in the scene, there are enough reliable point correspondences left. Due to the matching criteria no additional remaining outliers for the bundle adjustment are to be expected. Even with 80% repetitive patterns the accuracy loss is only about a factor of 2 compared to the experiments without repetitive patterns. In real world scenes repetitive patterns may cause more problems, because they are often not equally distributed.

Threshold Matching: The experiments with threshold matching show worse results: Although σ_{X_O} and $\sigma_{\omega\phi\kappa}$ have higher values, the errors are larger than indicated by the accuracies: The consistency measure is always greater than 2.9 and one experiment is clearly wrong: With 40% repetitive patterns the consistency measure is 840. Tests show that multiple runs in this case show different results, often better ones. As expected, the outlier rate of the matches is higher than for the best matching: at least 59% outliers. The results show, that this outlier rate is too high for the current algorithm to provide reliable results. As $\widehat{p}$ is always much higher than in the experiments with 'best matching', we do not recommend the use of 'threshold matching'. With the current parameter settings and simple outlier detection methods no advantage can be observed when there is a high amount of repetitive patterns in the observed scene.

The same experiments were carried out with the SURF operator. The results of the SURF operator were always worse than with the Lowe operator. But the SURF results become better with more repetitive patterns: The comparison with 80% repetitive patterns shows that $\widehat{p}_{SURF}/\widehat{p}_{Lowe}$ is 1.5. Again, 'threshold matching' is always worse than with 'best matching' when using SURF.

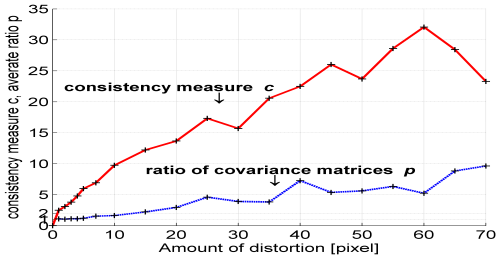

Figure 2: Quality of the orientation with images with non-linear distortion. X-Axis: Amount of maximal distortion in the image in pixel. Y-Axis: consistency measure c and average ratio of covariance matrices between dataset without distortion and dataset with distortion. Used images: dataset A, Nikon D70s camera with 20mm lens.

4.7 Results with Different Amounts of Distortions

In many practical applications, especially when using consumer cameras, the question whether to correct the non-linear distortions of the images arises. We used existing lookup tables for distortion and scaled the distortion offsets. These scaled distortion lookup tables were then applied to the already rectified (and

Matching	rep. parts %	# object points	# image points	outliers in matches [%]	$\sigma_{x'}$	σ_{X_O}	$\sigma_{\omega\phi\kappa}$	c	p	$\widehat{p}$
Best	0	1791	5369	3.8	0.20	0.00055	0.0077	1.6	1.0	1.6
	20	2030	5941	7.4	0.23	0.00059	0.0084	1.5	1.2	1.8
	40	1561	4761	14.3	0.22	0.00222	0.0085	1.2	1.4	1.7
	60	1484	4368	21.9	0.23	0.00446	0.0103	1.5	1.5	2.3
	80	1133	3316	35.5	0.27	0.01648	0.0141	1.8	1.9	3.4
Threshold	0	1404	4161	59.2	0.22	0.00542	0.0101	3.6	1.7	6.1
	20	1219	4179	68.3	1.56	0.00785	0.0570	2.9	8.6	24.5
	40	1254	3676	71.0	1.04	0.00618	0.0710	840.5	11.4	9581.7
	60	892	2685	75.3	0.97	0.02894	0.0490	30.7	11.6	243.6
	80	861	2005	84.2	1.49	0.02055	0.2303	56.4	27.3	1539.7

Table 6: Quality of the Orientation according to repetitive patterns. Dataset used: dataset F (simulated data). The texture consists of repetitive parts (percentage given in column 2) and non repetitive parts. Lowe features with best matching (ratio between best and second best match as threshold) and threshold matching used. The consistency measure c is calculated between the ground truth data and the experiment, the average ratio of the covariances p and $\widehat{p} = cp$ are calculated between the best matching result with 0% repetitive patterns and all other experiments.

therefore declared distortion-free) images and the impact on the orientation results was observed. With the help of the distortion scaling we set the maximal distortion successively from 1 to 70 pixel. The result for a Nikon D70s camera (dataset A) is visualized in Figure 2. Here even with 70 pixel distortion in the images AURELO was able to connect all images.

The Figure shows the consistency measure and the average ratio of the covariance matrices with respect to the maximal amount of distortion. The consistency increases approximately linear with the amount of distortion. Even with small distortions less or equal to 5 pixels the consistency measure has values from 2 to 6. That means that (with low and with high distortions) the accuracies of the orientation parameters do *not* reflect the shift of the orientation parameters due to non-linear distortion. For distortions equal or less than 5 pixels the average ratio p between the covariances is up to 1.2, so the covariances in principle do not change. The consequence for practical applications is that (at least with this image setup) it is very difficult to draw conclusions about the remaining non-linear distortion in the images and their effect on the accuracy with the help of the statistical results of a bundle adjustment that does not use any calibration parameters. As the loss of accuracy can be large, e.g. in this experiment for 5 pixels distortion $\widehat{p}_5 = c_5\ p_5 = 7.2$ and for 25 pixels distortion, the actual distortion of the used lens, $\widehat{p}_{25} = c_{25}\ p_{25} = 80$, the non-linear distortions of the camera should always be used if available.

5 SUMMARY AND OUTLOOK

The experiences with the tests showed that the consistency measure in conjunction with the average and maximal ratio of covariance matrices is a useful tool to benchmark different results of automated relative orientation procedures. To summarize our findings, the following generalized conclusions can be drawn:

- Orientation procedures which use random number generators, especially RANSAC algorithms, may produce significantly different results when started multiple times. This is dataset dependent.
- The accuracy loss when using the first image pyramid level instead of the original images is smaller than the decrease when using the subsequent smaller levels.
- If there is enough overlap, points with only two or even with only three observations can be left out with only a very small loss in accuracy.
- The SURF feature extraction implementation used here runs significantly faster than the classical SIFT operator, but yields much worse results. If performance is not a main issue, we thus recommend using SIFT. If accuracy is no issue, we recommend to use SURF, which is approximately a factor 2 less accurate in standard deviation.
- Matching SIFT descriptors with a threshold delivers worse results than matching with the ratio of best to second best match. This is at least true when no other sophisticated outlier detection is used. Even for images with large areas of repetitive patterns threshold matching has no advantages.
- Unmodeled non-linear distortion has a significant influence on the result. A big amount of the errors is not reflected in the estimated accuracies of the orientation parameters.

Not all influences on the quality were taken under consideration in this paper, e.g. the influence of the geometry (rotated images), the reflectance properties of the object, the lighting conditions and so on. Nevertheless the findings can help to produce more reliable results and improve the orientation procedures.

For AURELO the need for the implementation of a more sophisticated outlier detection method and/or the use of a robust bundle adjustment became clearly visible. The comparison with other orientation procedures is also one objective for our future work.

REFERENCES

Bay, H., Tuytelaars, T. and Gool, L. v., 2006. Surf: Speeded up robust features. In: Proceedings of the 9. ECCV.

Dickscheid, T., Förstner, W. and Läbe, T., 2008. Benchmarking automatic bundle adjustment results. In: XXI. ISPRS congress, Beijing, submitted, accepted.

Engels, C., Stewénius, H. and Nistér, D., 2006. Bundle adjustment rules. In: Photogrammetric Computer Vision (PCV).

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

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. In: International Journal of Computer Vision, Vol. 20, pp. 91–110.

Nister, D., 2004. An efficient solution to the five-point relative pose problem. In: IEEE PAMI, Vol. 20(6), pp. 756–770.

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.

Sinha, S. and Pollefeys, M., 2004. Camera network calibration from dynamic silhouettes. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition.