EXPERIMENTS WITH METADATA-DERIVED INITIAL VALUES AND LINESEARCH BUNDLE ADJUSTMENT IN ARCHITECTURAL PHOTOGRAMMETRY

Niclas Börlin*[a] and Pierre Grussenmeyer[b]

[a] Department of Computing Science, Umeå University, Sweden, niclas.borlin@cs.umu.se
[b] ICube Laboratory UMR 7357, Photogrammetry and Geomatics Group, INSA Strasbourg, France

KEY WORDS: Bundle adjustment, convergence, initial values, metadata, architectural photogrammetry

ABSTRACT:

According to the Waldhäusl and Ogleby (1994) "3x3 rules", a well-designed close-range architectural photogrammetric project should include a sketch of the project site with the approximate position and viewing direction of each image. This orientation metadata is important to determine which part of the object each image covers. In principle, the metadata could be used as initial values for the camera external orientation (EO) parameters. However, this has rarely been used, partly due to convergence problem for the bundle adjustment procedure.
In this paper we present a photogrammetric reconstruction pipeline based on classical methods and investigate if and how the linesearch bundle algorithm of Börlin et al. (2004) and/or metadata can be used to aid the reconstruction process in architectural photogrammetry when the classical methods fail. The primary initial values for the bundle are calculated by the five-point algorithm by Nistér (Stewénius et al., 2006). Should the bundle fail, initial values derived from metadata are calculated and used for a second bundle attempt.
The pipeline was evaluated on an image set of the INSA building in Strasbourg. The data set includes mixed convex and non-convex subnetworks and a combination of manual and automatic measurements.
The results show that, in general, the classical bundle algorithm with five-point initial values worked well. However, in cases where it did fail, linesearch bundle and/or metadata initial values did help. The presented approach is interesting for solving EO problems when the automatic orientation processes fail as well as to simplify keeping a link between the metadata containing the plan of how the project should have become and the actual reconstructed network as it turned out to be.

1 INTRODUCTION

1.1 Metadata

Compared to aerial photogrammetry, close-range photogrammetry is not limited to blocks of vertical images based on parallel strips with a predefined overlap. In architectural photogrammetry, image blocks are often made of horizontal stereopairs or strips of parallel images along the façades, complemented with closely spaced oblique images when turning around building corners. The well-known Waldhäusl and Ogleby (1994) "3x3 rules" have evolved to dense strips of images to better handle sets of unoriented and markerless terrestrial images by feature detection (Barazzetti et al., 2010).

In aerial photogrammetry, the position of the aircraft cameras is usually recorded by Global Navigation Satellite Systems (GNSS). Increasingly, positions and orientation angles of terrestrial images are recorded by internal GNSS and gravity sensors of cameras or smartphones. Smartphone applications allow the drawing of paths, standpoints and statistics. This metadata, combined with a sketch of the project site and the approximate position and viewing direction of each image could be used as initial values for the camera external orientation (EO) parameters.

Various metadata, methods and documentation standards for the collection and processing of Close Range Photogrammetry have been published (Bryan et al., 2009; Geometaverse, 2012). The metadata considered in this paper is the approximate position and viewing direction of each camera relative to the object, the roll and pitch angles of each camera, and the relative left-to-right ordering of the cameras at each position. This information should ideally be available in the project documentation. If not, rough estimates of e.g. the camera roll and pitch angle and Z coordinate (often approximately 2m above ground level for hand-held cameras), should be possible to infer from visual inspection of the images. Furthermore, in conjuntion with a 2D sketch of the scene, a rough estimate of the XY coordinate and aiming direction should be possible to infer.

1.2 Automatic processing

Today, automatic processing of images is available by e.g. PhotoSynth[1]. The processing is fully automatic; after uploading the images, the software does feature point detection, matching, camera calibration, relative orientation, and bundle adjustment, automatically. The result is presented as one or more groups of images that the software managed to process. The results are often impressive, but when the automation fails and e.g. only part of an image set is included into the reconstruction, it is difficult to continue and complete the reconstruction. Furthermore, it is not possible to supply guidance to the software, e.g. camera information or the relationship between images. The same limitation apply to e.g. the Bundler software[2] (Snavely et al., 2008) although it is of course technically possible to modify the source code.

1.3 Initial values

Before the bundle adjustment can be run, initial values of EO parameters and object points (OP) need to be calculated. Given measurements in images taken by a calibrated camera, the five-point method by Nistér (Stewénius et al., 2006) is today considered the gold standard for relative orientation (RO) and used for that purpose. In principle, metadata could also be used as initial values for the EO parameters. However, this has rarely been used due to convergence problem for the bundle adjustment procedure.

1.4 Aim of paper

The aim of this paper is to present a photogrammetric reconstruction pipeline based on classical methods and to investigate if and

[1] http://www.photosynth.com
[2] http://phototour.cs.washington.edu/bundler

how the linesearch bundle algorithm presented in Börlin et al. (2004) and/or metadata can be used to aid the reconstruction process in architectural photogrammetry when the classical methods fail. The paper focuses on the RO and bundle adjustment problems, and assumes that the measurement and matching problems have been solved.

2 ALGORITHMS

2.1 High-level algorithm

The high-level reconstruction algorithm is loosely based on Snavely et al. (2008) and is formulated as follows:

Given an image set with measured and matched points:

1. Identify which image pair i, j to start with.
2. Initialize the set of trusted images to contain image i.
3. Repeat while there are untrusted images left
   - 3.1. Identify which untrusted image j to add.
   - 3.2. Solve the RO problem between a trusted image i and image j to obtain initial EO values for image j.
   - 3.3. Estimate object points by forward intersection.
   - 3.4. Run the bundle adjustment.
   - 3.5. If the bundle succeeds
     - 3.5.1. Add image j to the set of trusted images.

     otherwise
     - 3.5.2.a Stop and report failure.

     or
     - 3.5.2.b Optionally: Try another RO algorithm in the next iteration. If the bundle fails again for the new initial EO values, stop and report failure.

To avoid initial bundle convergence problems due to a very short camera baseline, the bundle adjustment is deferred until the longest baseline is at least 25% of the average camera-OP distance. Which of the branches 3.5.2.a and 3.5.2.b is to be used depends on the experiment.

2.2 Image selection

The selection in step 3.1. of which image to add is based on the area covered by (the convex hull of) points common to images i and j. The pair is chosen as the pair with the largest covered area between a trusted image i and an untrusted image j, measured in image j. In the initial step 1., i and j are chosen among all images.

2.3 Initial values

Initial values for the EO parameters of the new camera j are calculated in step 3.2. by solving an RO problem given two images i and j, either by the five-point method (Stewénius et al., 2006) or from metadata. The five-point algorithm is as follows:

1. Use the algorithm in Stewénius et al. (2006) to solve for the essential matrix. There may be multiple solutions.
2. For each potential essential matrix, compute the four possible RO as described in Förstner et al. (2004, Ch. 11.1.4.3, 11.1.4.6). In each RO, camera i is the reference camera at the canonical position.
3. For each RO, perform forward intersection (Förstner et al., 2004, direct solution in Ch. 11.1.5.1) to establish OP coordinates of the points common to images i and j.
4. Keep the RO that has the largest number of OPs in front of both cameras. In case of a tie, keep the RO that has the smallest average image residual (reprojection error).
5. Given the optimal RO, scale the baseline based on the average scale of the OPs common to the i-j camera pair (in local coordinates) and the OPs created by the trusted camera network (in global coordinates). The scale is measured by $\mathrm{sqrt}(\mathrm{trace}(C))$, where C is the covariance matrix of the object points.
6. Align the canonical camera i of the scaled optimal RO with the EO of camera i in the trusted network. Use the aligned EO of camera j as initial values for camera j in the trusted network.

The metadata algorithm is as follows:

1. Align the i-j camera pair in the metadata with camera i in the trusted network.
2. Pick a scaling camera $k \neq i$ as the trusted camera closest to camera i in the metadata.
3. Scale the aligned i-j baseline based on the i-k distances in the trusted network and in the metadata.
4. Use the scaled EO of camera j as initial values for camera j in the trusted network.

2.4 Bundle adjustment

In the classical bundle adjustment algorithm (see e.g. Mikhail et al., 2001), at iteration k, an update vector $\mathbf{p}_k$ is estimated from the normal equations of the Gauss-Markov model evaluated at current estimate $\mathbf{x}_k$ of the unknowns. The new estimate is accepted as $\mathbf{x}_k + \mathbf{p}_k$ without any check that the new estimate is better than the current. This may lead to failure to converge, especially if the initial values are far from the minimum.

The Levenberg-Marquardt (L-M) method (Levenberg, 1944; Marquardt, 1963) has been suggested as a method to dampen bundle adjustment and reduce the number of convergence failures (Triggs et al., 2000). In the L-M method, a candidate update $\mathbf{p}_k$ is calculated from the normal equations modified by a bias $\lambda_k > 0$. If the residuals at the new point $\mathbf{x}_k + \mathbf{p}_k$ is small enough, the new point is accepted and the bias is reduced in the next iteration. Otherwise, the update is rejected, the bias is increased, and a new update is calculated. This selection scheme ensures that successive accepted points have decreasing residuals. The L-M method is in popular use in Computer Vision but has received criticism within the photogrammetric community, see e.g. Gruen and Akca (2005, sec. 4.6).

In this paper, we instead chose the method presented in Börlin et al. (2004), where the update $\mathbf{p}_k$ is calculated from the unmodified normal equations. A number of candidate points $\mathbf{x}_k + \alpha_k \mathbf{p}_k$ are calculated for $\alpha_k = 1, 1/2, 1/4, \ldots$ etc. The first candidate point

that produces a reduced residual sum is accepted as the new estimate. A minimal $\alpha_{\min} = 10^{-6}$ is used to avoid an infinite loop. This technique is well-known in the Optimization community as *line search*, see e.g. Nocedal and Wright (2006), and shares several convergence advantages with the L-M method.

2.5 Tested algorithms

Six versions of the high-level algorithm were implemented in the Matlab programming language[3] and used in the experiments:

C5 The classical bundle with five-point initial values.

L5 The linesearch bundle with five-point initial values.

CM The classical bundle with metadata initial values.

LM The linesearch bundle with metadata initial values.

C5M The classical bundle with five-point initial value as primary and metadata initial values as backup (in branch 3.5.2.b).

L5M The linesearch bundle with five-point initial value as primary and metadata initial values as backup.

3 DATA SET

The dataset used in this paper consisted of 42 images covering part of the INSA building in Strasbourg (Figure 1). The data set included parallel image strips along some of the façades, convergent subgroups of images around corners, as well as divergent images, especially near a concave corner of the building (near image 24 in Figure 2).

The images were taken with a Canon EOS 5D camera (full frame image chip, 4368x2912 pixels) equipped with a 20mm calibrated lens. Initial attempts to use bundler to obtain point measurements failed. Instead, PhotoModeler Scanner 2011 (PMS11) was used for point measurements. A total of 98 object points (OPs) were measured manually. Furthermore, the SmartPoint feature detection algorithm of PMS11 was used to add a further 936 automatic OPs to the project. After residual inspection, the measured data was exported from PMS11 and imported into Matlab.

The metadata protocol was designed to be easy to specify. A sketch of the object was presented on screen, and the approximate planimetric position and viewing direction was input. The position was input by user clicks whereas the viewing direction could either be input in multiplies of 30 or 45 degrees, or calculated by clicking an aim point in the sketch, see Figure 2. Multiple cameras could be specified to e.g. be parallel, have the same position, or have the same aim point. Furthermore, each image was inspected to determine the camera roll angle (in multiples of 90 degrees) and pitch offset (in multiplies of 30 degrees). From the metadata, the actual EO parameters were calculated, see Figure 3.

4 EXPERIMENTS AND RESULTS

4.1 Automatic

Two attempts were made to process the images automatically. The images were uploaded to PhotoSynth and also processed locally by Bundler. Neither method managed to reconstruct more than half the images. The largest image groups were 18 images (images 1–17, 25), and 9 images (images 6–14) , respectivelly.

[3] http://www.mathworks.com

Figure 1: Four images from the INSA dataset. From left to right, image 7, 17, 21, and 27. The image 21 roll angle was given as -90 degrees, the image 17 pitch angle was given as 30 degrees. The remaining roll and pitch angles were given as 0 degrees for these images.

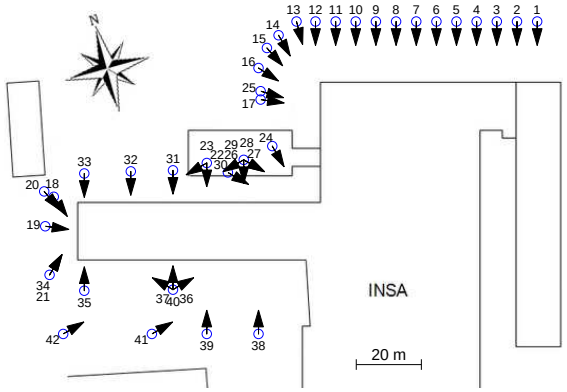


Figure 2: Planimetric metadata for the INSA dataset. The position and viewing direction was input by user clicks on the sketch of the object. Some cameras were grouped, e.g. cameras 1–12 were specified to be equidistant with parallel viewing directions, cameras 13–17 and 25 were specified to have the same aim point, and cameras 26–29 were specified to be coincident. Note that cameras 26, 29 and 21, 34 have identical planimetric metadata.

4.2 Full image set, five-point either/or metadata

The **C5**, **L5**, **CM**, and **LM**, algorithms were applied to the full image set. A maximum of 20 iterations was allowed for each bundle. The **C5** and **L5** algorithms using five-point initial values reconstructed the complete network with at most 4 iterations per bundle and a final $\sigma_0 = 0.96$, indicating a coordinate measurement error of 0.96 pixels. The final result is shown in Figure 4. With metadata initial values, the **CM** algorithm failed immediately whereas the **LM** algorithm reconstructed 17 images (1–13, 15–16, 24, 26) before failing to add image 14.

4.3 Less-one-image, metadata as backup

To generate more test problems, 42 data sets were generated by removing one of the images each from the full data set $\mathcal{S}_0$. Each image subset $\mathcal{S}_i$ contained all images except image i. This corresponds to the situation where the removed image was never taken. It could also be seen as an attempt to determine which image is the most important for the reconstruction to succeed. The **C5** algorithms was applied to each subset $\mathcal{S}_i$ to investigate if the removed image was critical for reconstruction or not.

The reconstruction of the full subsets $\mathcal{S}_i$ succeeded in 35 cases with a final σ_0 below 1.01 and with 5 iterations or less per bundle. Of the remaining 7 subsets, 3 failures (subsets $\mathcal{S}_6$, $\mathcal{S}_7$, $\mathcal{S}_9$) were due to no overlap between the remaining measured points. Thus, by removing e.g. image 7, the image network was effectively split in two disjoint parts.

The algorithms **L5**, **C5M**, and **L5M** were applied to the other 4 cases (subsets $\mathcal{S}_1$, $\mathcal{S}_{22}$, $\mathcal{S}_{27}$, and $\mathcal{S}_{42}$) to determine if the line-

Table 1: Reconstruction results for the image subsets where the algorithms produced differing results. Image set $\mathcal{S}_i$ contains all images except image i. The image numbers that were successfully reconstructed are listed, followed by the image where the algorithm failed, if any. If the complete reconstruction was successful, the largest number of bundle iterations that were needed is listed together with the σ_0 for the final network. Finally, if a metadata backup algorithm was successful, the number of times when the bundle failed for five-point initial values is listed.

Image subset	Algorithm	Largest reconstructed network	Failure when adding image	Largest number of required bundle iterations	Final σ_0	Number of failed bundles
$\mathcal{S}_1$	**C5**	2–13, 15–16, 26	24	-	-	-
	L5	all	-	8 (image 41)	3.4	-
	C5M	2–13, 15–16, 26	24	-	-	-
	L5M	all	-	8 (image 41)	3.4	0
$\mathcal{S}_{22}$	**C5**	1–17, 23–31	32	-	-	-
	L5	1–17, 23–31	32	-	-	-
	C5M	all	-	10 (image 32)	0.95	1 (image 32)
	L5M	all	-	6 (image 32)	0.95	1 (image 32)
$\mathcal{S}_{27}$	**C5**	1–17, 22–26, 28–31	32	-	-	-
	L5	1–17, 22–26, 28–31	32	-	-	-
	C5M	1–17, 22–26, 28–31	32	-	-	-
	L5M	1–17, 22–26, 28–31	32	-	-	-
$\mathcal{S}_{42}$	**C5**	1–40	41	-	-	-
	L5	1–40	41	-	-	-
	C5M	1–40	41	-	-	-
	L5M	all	-	11 (image 37)	0.95	1 (image 41)

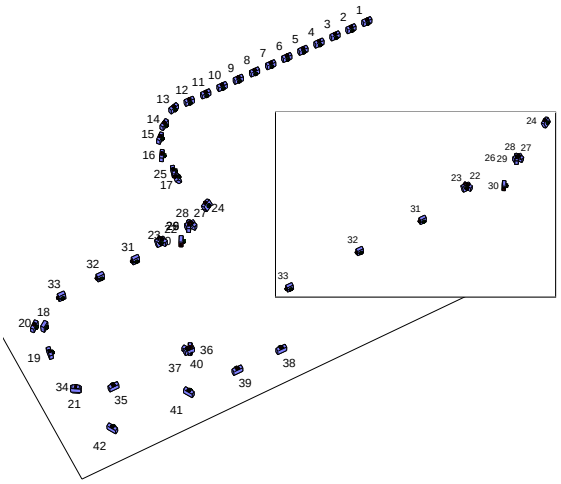


Figure 3: 3D representation of the EO metadata for the INSA dataset. Labels are generally placed behind the cameras.

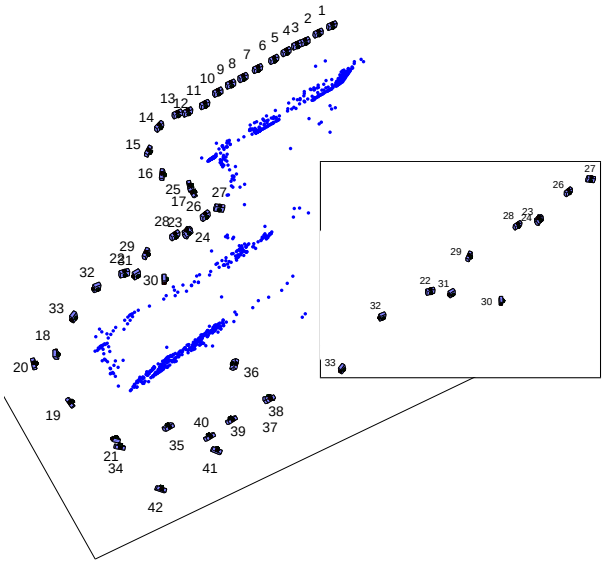


Figure 4: The full image network after convergence.

search bundle and/or metadata initial values as backup could reconstruct the network when the **C5** algorithm could not. The results are detailed in Table 1. In summary, for subset $\mathcal{S}_1$, the linesearch bundle helped convergence, though with an elevated σ_0, whereas metadata as backup did not help. The opposite was true for subset $\mathcal{S}_{22}$. Subset $\mathcal{S}_{42}$ required both linesearch and metadata to succeed whereas neither linesearch bundle, metadata as backup, nor both, improved the situation for subset $\mathcal{S}_{27}$. An illustration of the differences between the algorithms and initial values is shown in Figure 5

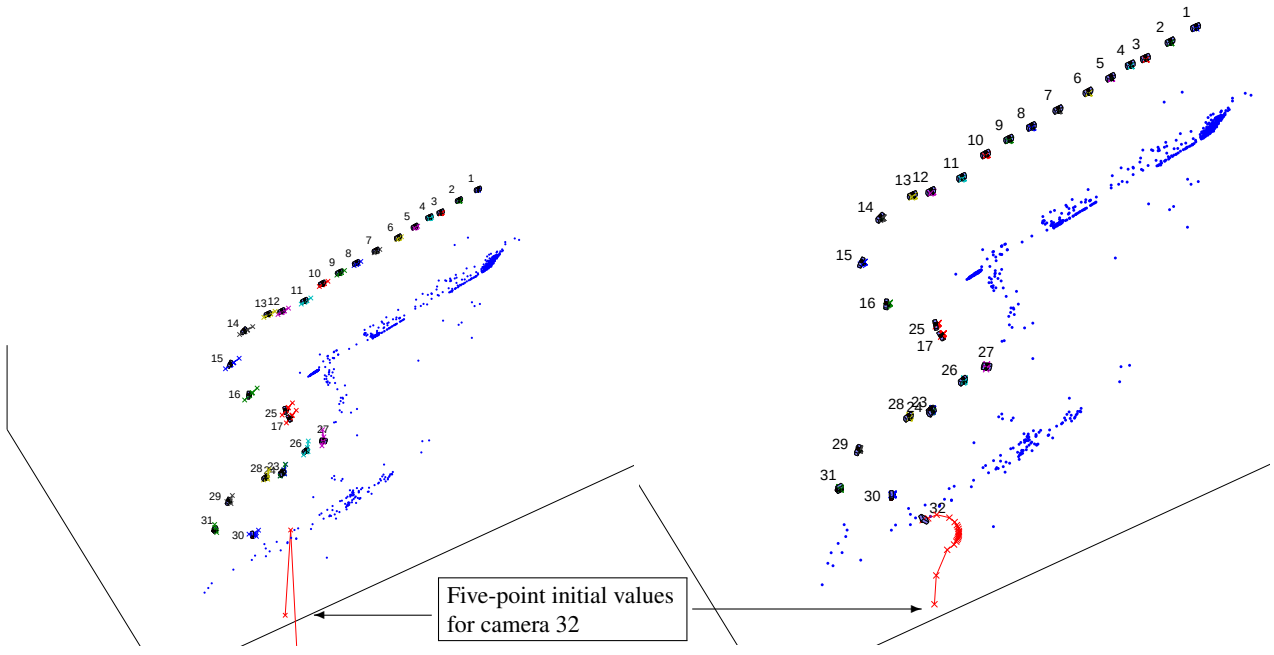
To investigate if some reconstruction failures were due to slow convergence, the above experiments were rerun with the maximum number of allowed bundle iterations increased to 50. This enabled the **L5** and **L5M** algorithms to succeed for subsets $\mathcal{S}_{22}$ and $\mathcal{S}_{27}$ with a final $\sigma_0 = 0.95$ and a maximum required bundle iteration count of 32 and 47, respectively.

4.4 Summary

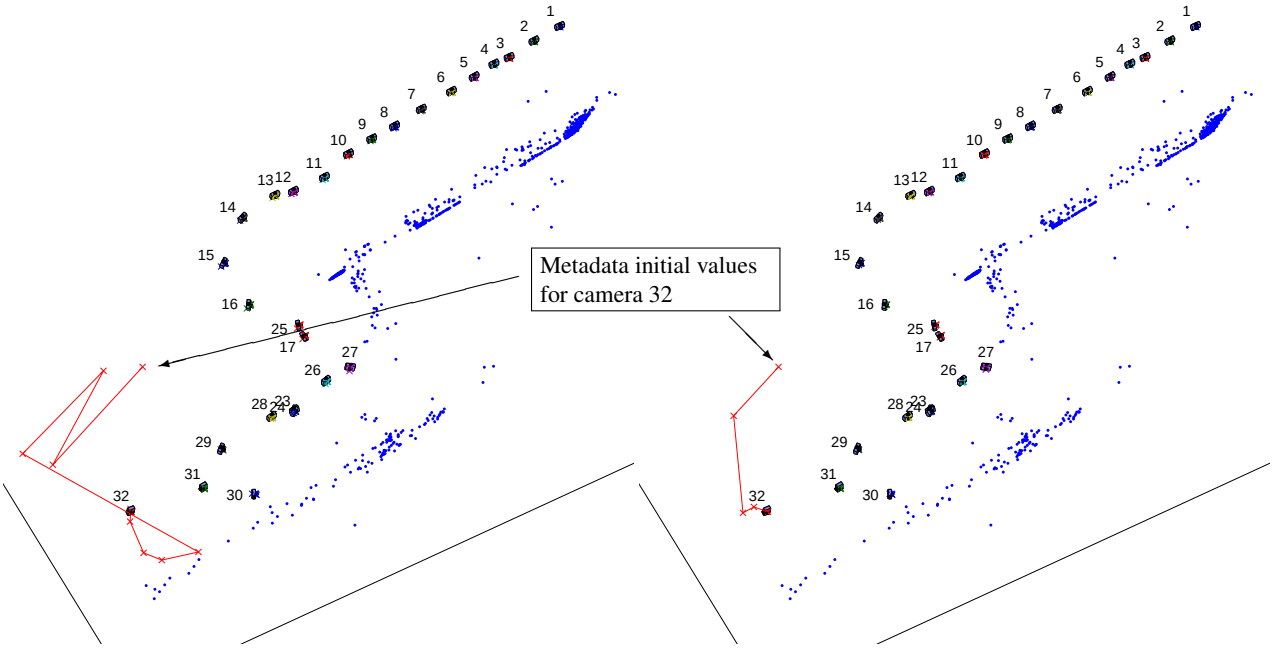
With only metadata as initial values, neither algorithm was able to reconstruct the whole data set. With five-point initial values, the **C5** algorithm worked on the complete data set and 35 subsets. All algorithms failed on 3 subsets due to lack of overlap. Of the final 4 subsets where **C5** failed, 1 required linesearch to work, 1 required metadata, and 1 required both. With a maximum allowed bundle iteration count of 20, 1 subset failed for all algorithms. If the maximum iteration count was increased to 50, the **L5** algorithm managed to reconstruct all subsets but one and **L5M** managed to reconstruct all subsets. The reconstructions of $\mathcal{S}_1$ showed an elevated σ_0, independent of iteration count.

5 DISCUSSION AND CONCLUSION

The aim of this paper was to present a photogrammetric reconstruction pipeline based on classical methods and to investigate



(a) **C5** (diverges) (b) **L5** (fails to converge)



(c) **C5M** (converges after 10 iterations) (d) **L5M** (converges after 6 iterations)

Figure 5: Iteration sequences for four algorithms when trying to add image 32. The successive positions of each camera are plotted with crosses connected by a line. The motion of most cameras are too small to be visible. The five-point initial EO values (top row) was on the wrong side of the building and facing the wrong direction. The **C5** bundle oscillated out of view before giving up after 8 iterations. Some oscillations induced in other cameras are visible, e.g. for cameras 15–17 and 25. On the same input, the **L5** bundle avoided oscillations by making small updates but failed to converge in the allowed 20 iterations. The metadata initial values (bottom row) were based on image 13/32 overlap and image 12/13 scaling and were also quite far from the true values but on the correct side of the building and facing the right direction. Given the metadata initial values, both the **C5M** and **L5M** bundles converged to the same position with a $\sigma_0 = 0.72$ in 10 and 6 iterations, respectively, the **C5M** bundle again showing larger oscillations of the two.

if and how the linesearch bundle algorithm presented in Börlin et al. (2004) and/or metadata could be used in architectural photogrammetry to aid the reconstruction process when the classical methods fail.

The results show that on the INSA data set, the classical **C5** algorithm worked well in most cases and neither linesearch nor metadata was needed for a successful reconstruction. However, in the cases where it did fail, linesearch bundle and/or metadata initial values did help. Indeed, when the allowed number of iterations was increased to 50, the linesearch/metadata combination succeeded in all cases. The difference in convergence results is consistent with Börlin et al. (2004) and suggests that the linesearch algorithm requires less precise initial values in order to converge. This, together with the simplicity to introduce e.g. GNSS and orientation metadata, has the potential to make photogrammetric reconstruction simpler and/or to reduce image dropouts in automatic reconstructions.

The attempts to reconstruct the images using the fully automated approaches by PhotoSynth and Bundler failed. It is unclear if the failures were due to feature point detection and matching problems, bundle convergence problems, or the lack of calibrated camera information. As the other experiments in this paper included some manual measurements, a comparison with other result is difficult, and the conclusion is limited to that the image set is non-trivial to reconstruct by fully automatic methods.

The algorithms evaluated in this study did not inlude any outlier detection. Instead, the data was inspected in PMS11 and thereafter assumed to be outlier-free. The small $\sigma_0 < 1.01$ of all reconstructions except one support that assumption. The exception was subset $\mathcal{S}_1$ with a $\sigma_0 = 3.4$. However, the fact that other reconstructions that included the same images showed a small σ_0 suggests that the reason for the elevated σ_0 was rather that the bundle got stuck in a non-global minimum for $\mathcal{S}_1$, something that neither algorithm managed to avoid.

It is a limitation of the study that all experiments are based on a single data set. However, we do believe that the data set is realistic for architectural photogrammetry and contains a representable variation of imaging geometries.

The selection of which image to be incrementally added to the network was based on the image area covered by points common with other images. This metric avoids the inclusion of images with little overlap to previously included images. However, as the measure is calculated from the "outermost" points (points forming the convex hull) only, it does not take the distribution of the interior points into account. Other selection criteria based on e.g. the RMS of the image points could be used instead. Furthermore, the algorithm only tries the "best" image to include before giving up or trying other initial values. While this would add to the algorithm complexity, a modification of the high-level algorithm to try multiple "second best" choices before giving up could increase the number of successful reconstructions.

Since the metadata was mainly used as backup to the measurement-based initial values, and the classical algorithms generally worked well, the amount of metadata actually used in this paper is small. On those grounds, it is difficult to argue for the addition of a processing step that is partly manual. However, at least if prepared before the photographic campaign, the metadata contains a description of the intention of the photogrammetrist. As such, it is linked to the photogrammetrists' mental picture of the project. If the reconstruction is based purely on measurements in the images and with little reference to the metadata, the link between the actual project and the metadata may weaken. If high-quality metadata is a goal of the project, it is likely that a higher level of interaction between the metadata and the reconstruction process is desired than what is the custom today. We believe that the tools presented in this paper will simplify maintaining the link between metadata and the reconstructed photogrammetric network.

Future work includes to create a complete reconstruction pipeline, including automatic feature point detection and matching and handling of outliers. Furthermore, the authors intend to make the source code freely available.

References

Barazzetti, L., Scaioni, M. and Remondino, F., 2010. Orientation and 3d modelling from markerless terrestrial images: Combining accuracy with automation. Photogramm Rec 25(132), pp. 356–381.

Börlin, N., Grussenmeyer, P., Eriksson, J. and Lindström, P., 2004. Pros and cons of constrained and unconstrained formulation of the bundle adjustment problem. IAPRS XXXV(B3), pp. 589–594.

Bryan, P., Blake, B., Bedford, J., Barber, D., Mills, J. and Andrews, D., 2009. Metric Survey Specifications for Cultural Heritage. English Heritage.

Förstner, W., Wrobel, B., Paderes, F., Craig, R., Fraser, C. and Dolloff, J., 2004. Analytical Photogrammetric Operations. 5 edn, IAPRS, chapter 11, pp. 763–948.

Geometaverse, 2012. Metadata and documentation requirements for close-range photogrammetry. http://gmv.cast.uark.edu/919/metadata-and-documentation-requirements-for-close-range-photogrammetry/.

Gruen, A. and Akca, D., 2005. Least squares 3d surface and curve matching. ISPRS J Photogramm 59(3), pp. 151 – 174.

Levenberg, K., 1944. A method for the solution of certain nonlinear problems in least squares. Quart. Appl. Math 2, pp. 164–168.

Marquardt, D. W., 1963. An algorithm for least squares estimation of nonlinear parameters. SIAM J Appl Math 11, pp. 431–441.

Mikhail, E. M., Bethel, J. S. and McGlone, J. C., 2001. Introduction to Modern Photogrammetry. Wiley.

Nocedal, J. and Wright, S. J., 2006. Numerical Optimization. 2 edn, Springer-Verlag.

Snavely, N., Seitz, S. M. and Szeliski, R., 2008. Modeling the world from internet photo collections. Int J Comp Vis 80(2), pp. 189–210.

Stewénius, H., Engels, C. and Nistér, D., 2006. Recent developments on direct relative orientation. ISPRS J Photogramm 60(4), pp. 284–294.

Triggs, B., McLauchlan, P., Hartley, R. and Fitzgibbon, A., 2000. Bundle adjustment — A modern synthesis. In: Vision Algorithms: Theory and Practice, Proceedings of the International Workshop on Vision Algorithms, Lecture Notes in Computer Science, Vol. 1883, Spring Verlag, pp. 298–372.

Waldhäusl, P. and Ogleby, C., 1994. 3x3 rules for simple photogrammetric documentation of architecture. IAPRS XXX(5), pp. 426–429.