Trajectory Reconstruction Using Long Sequences of Digital Images From an Omnidirectional Camera

BENNO SCHMEING[1], THOMAS LÄBE[2] & WOLFGANG FÖRSTNER[3]

Wir präsentieren einen Ansatz, um lange Bildfolgen einer omnidirektionalen Kamera mittels Bündelausgleichung auszuwerten. Wir nutzen Bilder des Multikamerasystems Ladybug3 von PointGrey, welches aus sechs Einzelkameras besteht. Die gegenseitige Überdeckung aufeinanderfolgender Bilder ist groß; Verknüpfungen zwischen weit entfernten Bildern kommen nur über Schleifenschlüsse zustande. Zwei Probleme sind zu lösen: (1) Die Bündelausgleichung muss Bilder einer omnidirektionalen Kamera verarbeiten und (2) Ausreissersuche und Näherungswertbestimmung müssen mit der speziellen Aufnahmegeometrie umgehen können. Wir lösen Problem (1) indem wir die Einzelkameras der Ladybug zu einer virtuellen Kamera zusammenfassen und für die Bündelausgleichung ein sphärisches Modell verwenden. Die Ausreisserdetektion (2) erfolgt über lokale Bündelausgleichungen benachbarter Bilder und anschließende robuste Gesamtbündelausgleichung. Ein Inertialnavigationssystem liefert die benötigten Näherungswerte für die Kamerapositionen.

We present a method to perform bundle adjustment using long sequences of digital images from an omnidirectional camera. We use the Ladybug3 camera from PointGrey, which consists of six individual cameras pointing in different directions. There is large overlap between successive images but only a few loop closures provide connections between distant camera positions. We face two challenges: (1) to perform a bundle adjustment with images of an omnidirectional camera and (2) implement outlier detection and estimation of initial parameters for the geometry described above. Our program combines the Ladybug's individual cameras to a single virtual camera and uses a spherical imaging model within the bundle adjustment, solving problem (1). Outlier detection (2) is done using bundle adjustments with small subsets of images followed by a robust adjustment of all images. Approximate values in our context are taken from an on-board inertial navigation system.

1 Introduction and Motivation

Trajectory reconstruction as well as *structure from motion (SfM)* are both intensively investigated. Many approaches rely on bundle adjustment, estimating both the camera positions and orientations as well the coordinates of 3D scene points.

Usually, the functional model of the bundle adjustment is based on a planar pinhole model. Hence, the bundle adjustment uses 2D image points as observations. We present a spherical camera model for bundle adjustment that uses 3D directions in the camera system as observations. Thus, we are able to model classical planar pinhole cameras as well omnidirectional cameras, e.g. multi-camera systems or catadioptric cameras having a single viewpoint.

[1] Benno Schmeing, University of Bonn, Department of Photogrammetry, E-Mail: benno.schmeing@gmx.de

[2] Thomas Läbe, University of Bonn, Department of Photogrammetry, E-Mail: laebe@ipb.uni-bonn.de

[3] Wolfgang Förstner, University of Bonn, Department of Photogrammetry, E-Mail: wf@ipb.uni-bonn.de

1.1 Related work

For pinhole cameras with planar sensor there are many existing packages for fully automatic bundle adjustment, e.g. the system of MAYER (2008), AURELO (LÄBE & FÖRSTNER, 2006) or the BUNDLER (SNAVELY ET AL., 2006). New developments focus on minimizing computational effort and handling large data sets. FRAHM ET AL. (2009) present an approach to process data sets up to the scale of millions of images.

However, although much work was done on the calibration of omnidirectional cameras, few publications address the topic of bundle adjustment for omnidirectional cameras. For multi-camera systems, the images from the system's single cameras are often processed separately, e.g. (LEE, 2009). Publications that present bundle adjustment for catadioptric cameras are for example (LHUILLIER, 2005) and (RITUERTO ET AL., 2010). Both consider for the cameras' special characteristic by sophisticated calibration functions.

1.2 Experimental setup

We have several sequences of images taken with the Ladybug3 multi-camera system (see Figure 1). The Ladybug3 is mounted on a hand-guided platform together with an Inertial Measurement Unit (IMU), odometer and GPS. Using the odometer data, the Ladybug3 is triggered once per meter. Since it is a multi-camera system, the Ladybug makes six images at a time. We use four (left rear, left front, right front and right rear) which have clear sight and can be expected to support the orientation process. We want to determine the camera trajectory from these images using a bundle adjustment.

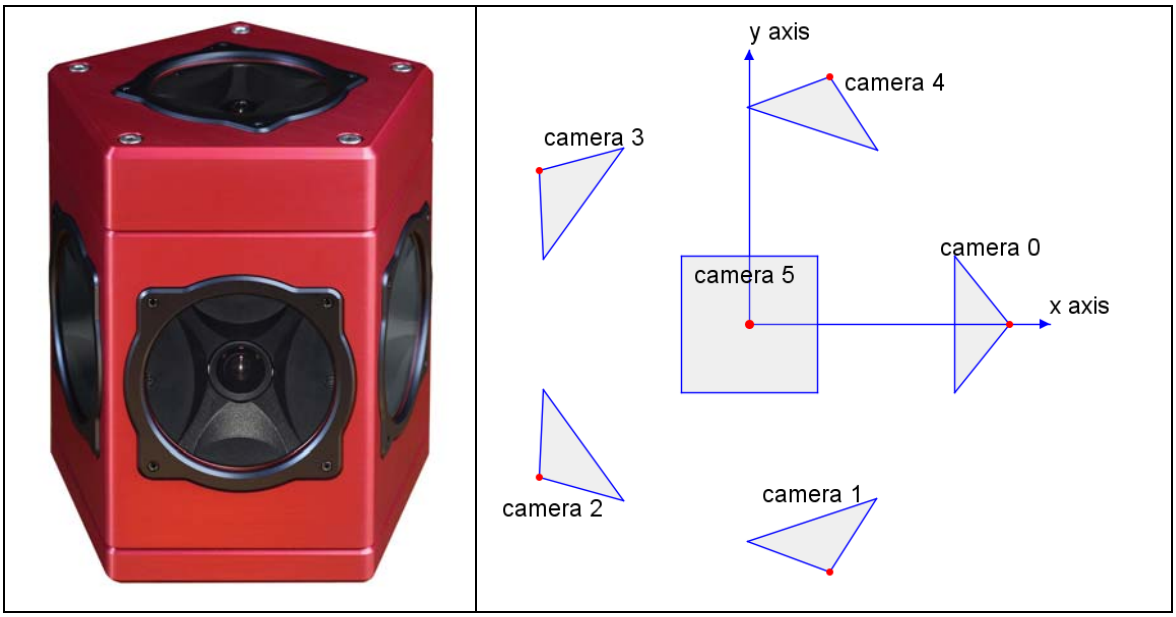

Figure 1: Ladybug3 camera arrangement. Left: picture of the Ladybug3. Right: Sketch of the camera arrangement: Five cameras are arranged horizontally, the sixth one points upwards. The x and y axis define the Ladybug reference frame.

1.3 Problems to solve

Since there is very little overlap between the view fields of the Ladybug's single cameras, a conventional bundle adjustment of the individual trajectories yields separate, diverging trajectories (see Section 3.2 for an example). Therefore, you must either add restrictions to enforce the stable mutual orientation of the single cameras within the bundle adjustment or combine the single cameras into a spherical virtual camera. We do the latter and transform the image coordinates from the cameras into directions within a common reference system, taking into account, that the projection centers of the individual cameras are distinct and not identical to the viewing center of the spherical camera. This has the advantage that the number of unknowns for the camera parameters decreases significantly in comparison to the approach of using restrictions. The model can also be used for omnidirectional cameras with a different mechanical setup. For a strict formulation of a multi-camera system, see (SCHNEIDER ET AL., 2011).
The geometry and size of the data set is a second problem. For long sequences of images with no or just very little observations between far away images, the bundle adjustment becomes numerically instable. Hence, outlier detection and provision of reliable approximate values is crucial. Usually, the generation of approximate values is based on relative orientations between pairs of images: Starting with a pair of images the other images are subsequently added. However, during long sequences of images, errors due to Random Walk and drift effects accumulate resulting in huge inconsistencies at loop closures. Therefore we currently use the data from the IMU, odometer and GPS to compute approximate values.

2 Description of our approach

2.1 Planar and spherical model for bundle adjustment

The functional model for bundle adjustment with a standard *planar* pinhole camera is based on the collinearity equation, which describes the projection of the points from the 3D photogrammetric system into the image:

$$\tilde{x} = \lambda K R \left[I - X_0 \right] \tilde{X} \tag{1}$$

Here, $\tilde{X}$ stands for the object points (in 3D homogeneous coordinates), $\tilde{x}$ stands for the image points (2D homogenous coordinates). The matrices K, R and X_0 are the calibration and rotation matrix of the camera respectively the position of the camera projection center. The unknown scale factor λ usually is eliminated by expressing (1) using the Euclidean coordinates. This does not allow to use viewing rays orthogonal or nearly orthogonal to the viewing direction. Eq. (1) is used for all image points observed in all cameras exposed at all times.
Unlike a pinhole camera, omnidirectional cameras generate observations in all directions. Therefore, we cannot use 2D image coordinates to describe these observations. Instead, we define the *spherical* model that uses normalized three dimensional directions:

$$m^s = N\left(R_w^s \left[I - X_0 \right] \tilde{X} \right) \tag{2}$$

with $N(...)$ normalizing a vector to unit length. In our context we split R into two parts: R_w^s describes the rotation from the world system into the spherical camera and R_s^v describes the

rotation from the spherical camera into the individual virtual cameras. In equation (2), we do not need R_s^v, as m^s is defined in the spherical camera system. Section 2.2 explains the computation of m^s from x^v.
In the bundle adjustment, we now minimize the Euclidian distances between the observed and computed directions:

$$\Phi = \min \sum \left| m - N\left(R_W^T \left[I - X_0\right]\tilde{X}\right)\right|^2$$

This is equivalent to minimizing the sum of squares of all directional differences between observed directions and adjusted directions, see MOURAGNON et al. (2009). Since equation (2) still represents the collinearity equation, we can use established and proven algorithms to solve the bundle adjustments subtasks, e.g. determining the relative orientation between cameras.

2.2 Combining the pinhole cameras to a spherical camera

We use the multi-camera system Ladybug3, which consists of six cameras[4] pointing in different directions thus observing a great deal of the sphere. However, the projection centers of the Ladybug's six cameras don't coincide but feature a radial shift of about 4 cm (see Figure 1). Therefore, we first transform the observations into virtual cameras whose projection centers coincide with the ladybug reference system's origin. In the second step, we merge the virtual (pinhole) cameras into a single spherical camera. The whole process is shown in Figure 2.

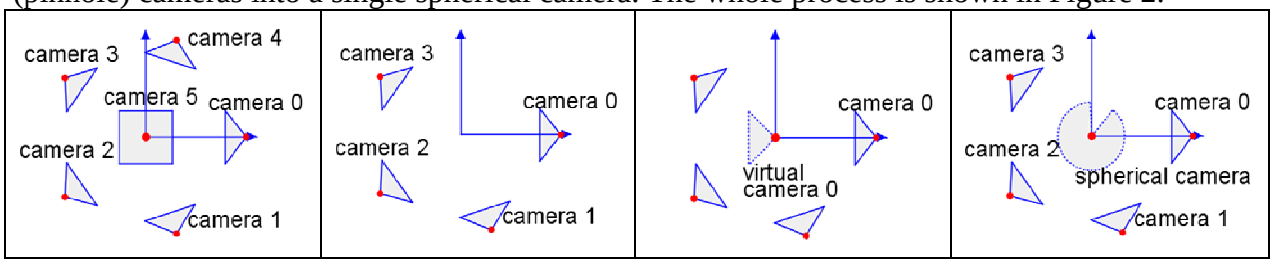

Figure 2: Combining the pinhole cameras to a spherical camera. The Ladybug consists of six cameras (left) of which we use four (middle left). Each camera is transformed into a virtual camera whose projection center coincides with the origin of the camera system (shown for camera 0, middle right). Finally the individual virtual planar cameras are fused into a virtual spherical camera, where the original image coordinates are represented as unit direction vectors (right).

2.2.1 Transformation from real to virtual pinhole cameras

When transforming the observations from real to virtual pinhole cameras, the dislocation of the cameras' projection centers results in a concentric point wise distortion. We implement this effect by applying equation (3):

$$x_v = \frac{d}{d + \Delta d} \cdot x_m \qquad (3)$$

Here, the Euclidean vectors x_m and x_v describe the *m*easured image positions respectively their projections into the *v*irtual cameras and Δd is the known offset between the individual real and

[4] We use only four cameras: The rear camera's view is blocked by equipment and the operator moving the camera platform. The top camera observes mainly sky.

the virtual camera. The distances parallel to the line of sight d are determined using forward intersection with other neighboring images.

2.2.2 Transformation from virtual cameras into the spherical camera

The transformation from the virtual cameras into the spherical camera is done by a rotation accounting for the virtual camera's orientation (compare equations (1) and (2)):

$$m = N\left(\left(R_c^r\right)^{-1} \cdot K^{-1} \cdot x_v\right) \tag{4}$$

To define R_c^r, we use calibration data from the manufacturer.

2.2.3 Accuracy of the directions in the ladybug reference system

Generally, the measurement accuracy of the point detector (Lowe 2004) is approximately 1/3 pixel. We need to clarify in how far errors occurring during the transformation of an image point into the virtual camera of the Ladybug reference system may downgrade the accuracy.

The accuracy of the transformation into the virtual camera depends on the accuracy σ_d of the distances parallel to the line of sight d determined by forward intersection. The uncertainty of the length of the baselines B directly influences the uncertainty of the distance d. Hence, σ_d depends both on the accuracy of the intersection and the relative accuracy of the baselines. For the recording situations faced here ($B = 1\,\text{m}, \sigma_B / B = 1\%, d > 1\,\text{m}$) we obtain standard deviations of the shift between x_v and x_m less than 0.1 pel. Thus, with the exception of gross errors, the transformation into virtual cameras does not downgrade the accuracy noticeable.

The effect of the transformation into the ladybug reference system is assessed using error propagation for equation (4). Since we miss reliable information about the accuracy of the manufacturer calibration, we assumed pessimistic values (projection center accuracy 1mm, camera orientation 0.1°) resulting in accuracies of about 1 pel. In Section 3, we will see that the actual accuracies are substantially better.

2.3 Workflow

We start with a data set consisting of:

- rectified images (compensated nonlinear distortion)
- approximate position and orientation of the Ladybug3 at every triggering point
- calibration data with arrangement of Ladybug's single cameras.

To determine the camera trajectory, we perform the steps listed below:

- Extraction of interest points according to (LOWE, 2004)
- Matching the interest points: Matching all images against each other is a massive task. Using the trajectory from the IMU, odometer and GPS, we can limit the number of pairs to be matched. Only image pairs whose computed fields of view provide enough overlap (here: 50%) are matched.
- Merging the pinhole cameras to a spherical camera: To merge the pinhole cameras, we follow the steps in Section 2.2. In order to do the transformation into the virtual cameras, we need to determine 3D coordinates in the camera system by forward intersection with

all linked images. The camera orientations are computed using the 5-point algorithm presented in (NISTER, 2004) and RANSAC (FISCHLER & BOLLES, 1981). The scale factor is set using the baseline length measured by the IMU and odometer.

- Generation of approximate values: Using the trajectory from the IMU, odometer and GPS (see Section 2.4) and the pairs of points from the matching, we compute approximate coordinates for the object points by intersection.
- Data cleaning: The pairs of points from the matching of images feature a significant number of outliers. Most – but not all – are detected during the relative orientation of the images. To clean the dataset from the remaining outliers, we execute local bundle adjustments with about 20 spherical cameras and use the resulting residuals for blunder detection.
- Global bundle adjustment: Depending on the size of the dataset, it may be necessary to call the final bundle adjustment (including all virtual cameras) more than once and do again a blunder detection in every iteration.

2.4 Generating approximate values

In Section 1.3, we have stated that random walk and drift effects are a problem when generating approximate values for long image sequences by adding up relative orientations. This problem can be solved by portioning the sequence into smaller subsequences, which are dealt separately and combined afterwards.

Here, we use a GPS sensor and an Inertial Measurement Unit (IMU) to determine approximate values for the Ladybug's exterior orientation. The data from the IMU is used to determine a precise trajectory by dead reckoning. GPS is used to fix the datum.

The resulting trajectory is very precise locally, but errors add up creating significant deviations both due to random walk and drift effects (see Table 1). However, the trajectory is well capable for generating approximate values.

Longitudinal component	Lateral component	Height component
-0.1647 m	1.0210 m	-0.2983 m

Table 1: Loop closure error from the IMU and GPS for a 778 m long trajectory. The data was recorded during the measurement for Section 3.3.

2.5 Implementation Issues

Since we use directions as observations (3 coordinates instead of 2 image coordinates) we cannot use a standard bundle adjustment package (without modifications). We therefore use the free available package SBA (source code open) (LOURAKIS, A. & ARGYROS, A.A, 2009). It allows to define the observation function together with the Jacobians w. r. t. the parameters. We implement the model described in Section 2.1 in SBA. Since SBA uses a sparse Levenberg-Marquardt implementation, it is also suitable for large datasets. In addition to our blunder elimination, we call SBA in an iterative scheme with a reweighting procedure in between to robustify the bundle adjustment.

3 Results

We use three data sets to evaluate the validity and performance of our approach:

- We demonstrate that the bundle adjustment yields the same results with the spherical camera model as with the planar model when applied to images taken with a classical pinhole camera.
- We demonstrate the higher robustness and consistency when modeling a spherical camera instead of several pinhole cameras to process data from an omnidirectional camera (here: the Ladybug3).
- We demonstrate that our approach is capable of dealing with large data sets.

3.1 Comparison of planar and spherical projection model

We first want to investigate the impact of the different projections models, planar and spherical model, on the result of the bundle adjustment. We perform bundle adjustment of a test dataset („City hall Leuven“) which is available at http://cvlab.epfl.ch/data/strechamvs. The dataset consists of 7 images with known calibration parameters that show a scene with detailed texture. We use the following three approaches:

1. Bundle adjustment with the free available software BUNDLER (see http://phototour.cs.washington.edu/bundler)
2. Bundle adjustment with our in-house software called AURELO (LÄBE & FÖRSTNER, 2006), which uses SBA for bundle adjustment.
3. Using exact the same set of observations as in the previous approach, but transforming the image coordinates into directions before minimizing the residuals of the directions in the bundle adjustment using the spherical model.

To compare the resulting orientation parameters, we use the approach of DICKSCHEID ET AL. (2008). We report the *consistency measure c* here, which can be interpreted as the average factor between the differences of the orientation parameters and their accuracies described by the full covariance matrix of the orientation parameters. The consistency measure between results (1) and (2) is c_{12} =6.7, assuming the same covariance matrix for BUNDLER as for AURELO, because BUNDLER does not deliver accuracies, between result (2) and (3) c_{23} =0.74. This shows that optimizing directions has a very small influence on the result ($c_{23} < 1$, so the differences are smaller than their assumed distribution indicated by the covariance matrices). Using another program and thus another set of observations has a much stronger influence ($c_{23} >> c_{12}$), which mainly is due to the different ways to eliminate blunders.

3.2 Using a single spherical camera instead of separate pinhole cameras

Processing *all* images of the Ladybug's single cameras in one bundle adjustment does not deliver reliable and consistent results, as to be expected. In order demonstrate this, we use a small test dataset of 15 equally spaced positions where images of 4 cameras of the Ladybug were taken. For the first test, we simply put all 60 images in our in-house software AURELO. As a second test case we used the new developed approach described above. The resulting trajectories of both tests are depicted in Figure 3. The single cameras' estimated projection centers are not centered

around the Ladybug's true positions but scatter strongly. The main problem is the small overlap between the Ladybug's cameras' fields of view. Thus there are few observations connecting images from different cameras. In case one camera's trajectory is broken, it can hardly be restored using the other cameras. Especially for narrow curves, this is a distinct possibility. Combining the images increases the probability of connecting all positions dramatically.
Summarizing, the spherical camera model is both more consistent, because the rigidity of the arrangement of the Ladybug's single cameras is considered, and robust, because there is a lower risk that a trajectory breaks.

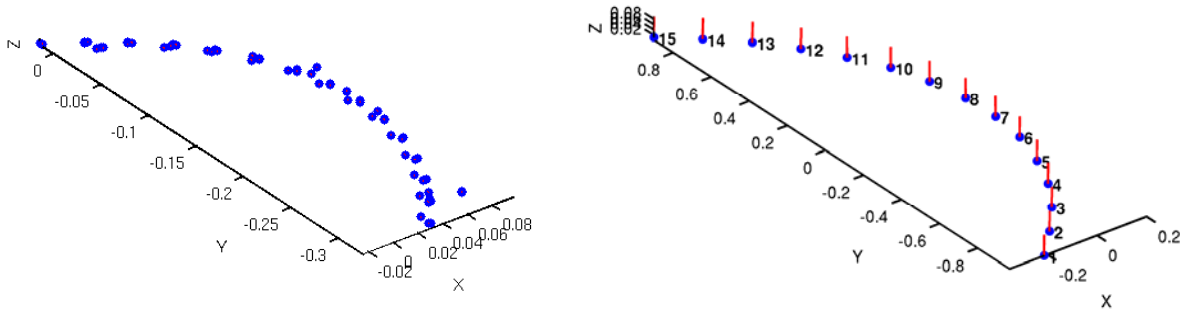

Figure 3: Visualization of a trajectory with 15 positions and 60 images. Left: Using each image separately without any constraints. Right: Combining the cameras to one virtual camera per position. Note the equally spaced result on the right which is much more accurate than the left one.

3.3 Processing a large data set

Here, we will show that our approach is capable of processing a large data set. The data set consists of three loops around a University Bonn library building (see Figure 4). There is a total of 778 triggering points (3112 pictures) geotagged by an Inertial Measurement Unit (IMU) and GPS. Before outlier detection, we start with observations to 130.466 putative object points. After outlier detection, 463.722 observations to 91.792 object points remain.

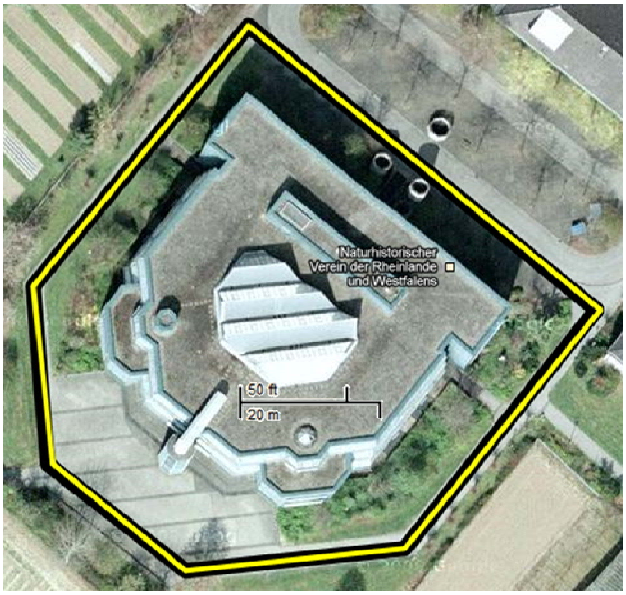

Figure 4: The data set for Section 3.3 was captured during three loops around the Branch Library for Medicine, Science and Agriculture of the University Bonn (drawn in yellow).

The bundle adjustment runs without problems and estimates the parameters for the exterior orientation for every triggering position.
The first parameter to assess the validity of the bundle adjustment results is the a posteriori variance factor. The bundle adjustment results in an a posteriori variance of 0.06° for the spherical camera system respectively 0.64 pel for the pinhole cameras, which compares well with the values from the tests performed above and with the value reported in (SCHNEIDER ET AL., 2011). Thus, there is no hint for undetected outliers or other problems, such as failed convergence, reaching a local optimum etc.
In the setup shown here, the standard deviations of the estimated projection centers increase up to 20 cm at the triggering positions furthermost from the point we used to fix the datum.

4 Conclusion and outlook

We have presented an algorithm to perform a bundle adjustment for omnidirectional cameras. The spherical camera model is both simple and easily adaptable to several kinds of cameras. With the procedures from Section 2.2 we expanded it so that we can process data from multi-camera systems like the Ladybug3.
Testing the spherical camera model on both publically available and own data sets, we have shown that:

- Using the spherical camera models yields the same results as standard bundle adjustment algorithms for images from pinhole cameras.
- When processing images from omnidirectional cameras like the multi-camera system Ladybug3, merging the different pinhole cameras into a single virtual spherical camera yields better robustness and consistency of the results.
- The workflow proposed here is also capable of processing large data sets.

Several problems have been identified, but not solved yet:

- Actually, the local bundle adjustments do not test all observations. Here, a scheme to improve the ratio between computational effort and completeness is preferable.
- Except for the bundle adjustment in SBA, the procedure is not optimized for low computation times.
- Depending on the quality of the approximate camera orientations and positions, the process of generating the virtual image needs to be iterated to yield accurate corrections when transforming the observations into the virtual cameras.
- There appears to be no generally method to obtain approximate values. The generation of approximate values is solved both for narrow (using a Kalman-Filter) and wide-baseline imagery (e.g. Snavely et al., 2006). However, the image sequence from Section 3.3 is a mixture of both where classical methods are likely to fail.

5 References

SNAVELY, N. ET AL., 2006: Photo Tourism: Exploring Photo Collections in 3D. In: ACM Transactions on Graphics, pp. 835-846.

FRAHM, J.M. ET AL., 2010: Building Rome on a Cloudless Day. In: ECCV 2010, pp. 368-381

LÄBE, T. & FÖRSTNER, F., 2006: Automatic Relative Orientation of Images. In: Proceedings of the 5th Turkish-German Joint Geodetic Days.

DICKSCHEID, T. & LÄBE, T. & FÖRSTNER, W., 2008: Benchmarking Automatic Bundle Adjustment Results. In: 21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS). Beijing, China 2008, pp. 7-12 Part B3a.

FISCHLER, M.A. & BOLLES, R.C., 1881: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. In: Communications of the ACM, volume 24, pp. 381-395.

LEE, T., 2009: Robust 3D street-view reconstruction using sky motion estimation. ICCV Workshops 2009, pp. 1840-1847.

LHUILLER, M., 2005: Automatic Structure and Motion using a Catadioptric Camera. In: OMNIVIS 2005.

LOURAKIS, A. & ARGYROS, A.A, 2009: SBA: A Software Package for Generic Sparse Bundle Adjustment. In: ACM Trans. Math. Software, volume 36, number 1, pp. 1-30, New York, NY, USA.

LOWE, D. (2004). Distinctive image features from scale-invariant keypoints. In: International Journal of Computer Vision, volume 20, p. 91–110.

Mayer H. (2008): Issues for Image Matching in Structure from Motion: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 37 (B3a), pp. 21-26.

MOURAGNON, E. LHUILLIER, M., DHOME, M., DEKEYSER, F., SAYD, P. (2009): Generic and real-time structure from motion using local bundle adjustment, *Image and Vision Computing*, Vol. 27, No. 8, pp. 1178-119.

NISTER, D, 2004: An Efficient Solution to the Five-Point Relative Pose Problem. IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 26, pp. 756-777.

RITUERTO, A., ET AL., 2010: Visual SLAM with an Omnidirectional Camera. ICPR 2010, pp. 348-351.

SCHNEIDER, J., SCHNINDLER, F. & FÖRSTNER, W., 2011: Bündelausgleichung für Multikamerasystemen, DGPF Tagungsband 20 / 2011.