AUTOMATED END-TO-END WORKFLOW FOR PRECISE AND GEO-ACCURATE RECONSTRUCTIONS USING FIDUCIAL MARKERS

Markus Rumpler[a]*, Shreyansh Daftry[ab], Alexander Tscharf[c], Rudolf Prettenthaler[a], Christof Hoppe[a], Gerhard Mayer[c] and Horst Bischof[a]

[a] Institute for Computer Graphics and Vision, Graz University of Technology, Austria
(rumpler, hoppe, bischof)@icg.tugraz.at, rudolf.prettenthaler@tugraz.at

[b] Robotics Institute, Carnegie Mellon University, USA
daftry@cmu.edu

[c] Chair of Mining Engineering and Mineral Economics, Montanuniversität Leoben, Austria
(alexander.tscharf, gerhard.mayer)@unileoben.ac.at

Commission III/3

KEY WORDS: Photogrammetric Computer Vision, Unmanned Aerial Vehicles, Image-based 3D Reconstruction, Mapping, Image Acquisition, Calibration, Online Feedback, Structure-from-Motion, Georeferencing, Fiducial Markers, Accuracy Evaluation

ABSTRACT:

Photogrammetric computer vision systems have been well established in many scientific and commercial fields during the last decades. Recent developments in image-based 3D reconstruction systems in conjunction with the availability of affordable high quality digital consumer grade cameras have resulted in an easy way of creating visually appealing 3D models. However, many of these methods require manual steps in the processing chain and for many photogrammetric applications such as mapping, recurrent topographic surveys or architectural and archaeological 3D documentations, high accuracy in a geo-coordinate system is required which often cannot be guaranteed. Hence, in this paper we present and advocate a fully automated end-to-end workflow for precise and geo-accurate 3D reconstructions using fiducial markers. We integrate an automatic camera calibration and georeferencing method into our image-based reconstruction pipeline based on binary-coded fiducial markers as artificial, individually identifiable landmarks in the scene. Additionally, we facilitate the use of these markers in conjunction with known ground control points (GCP) in the bundle adjustment, and use an online feedback method that allows assessment of the final reconstruction quality in terms of image overlap, ground sampling distance (GSD) and completeness, and thus provides flexibility to adopt the image acquisition strategy already during image recording. An extensive set of experiments is presented which demonstrate the accuracy benefits to obtain a highly accurate and geographically aligned reconstruction with an absolute point position uncertainty of about 1.5 times the ground sampling distance.

1 INTRODUCTION

Photogrammetric methods and image-based measurement systems have been increasingly used in recent years in different areas of surveying to acquire spatial information. They have become more popular due to their inherent flexibility as compared to traditional surveying equipment such as total stations and laser scanners (Leberl et al., 2010). Traditional aerial photogrammetry demands resources and occasions high costs for manned, specialized aircrafts and is therefore only economical for very large survey areas. In contrast, terrestrial photogrammetry is cheaper and more flexible, but is limited by the ground based camera positions. Hence, scene coverage is limited as visibility problems may arise depending on the scene geometry in certain areas which are not visible in images taken from ground level. Photogrammetry with Unmanned Aerial Vehicles (UAVs) has recently emerged as a promising platform which closes the gap and combines the advantages of aerial and terrestrial photogrammetry and also serves as low-cost alternative to the classical manned surveying.

The availability of affordable high quality digital consumer grade cameras combined with the use of lightweight, low-cost and easy to use UAVs offers new ways and diverse capabilities for aerial data acquisition to perform close range surveying tasks in a more flexible, faster and cheaper way. In conjunction with an automated multi-image processing pipeline, 3D reconstructions and dense point clouds from images can be generated on demand, cost-efficient and fast. Fully automated methods for image-based 3D reconstructions originate in the field of image processing (Hartley and Zisserman, 2004) and have been integrated in many, partly freely available software packages (e.g. VisualSfM, Acute3D, Pix4D, Agisoft PhotoScan, PhotoModeler, etc.). The methods are able to calculate the camera orientations and scene structure represented as a (sparse) 3D point cloud from an unordered set of images. In subsequent steps, the model gets refined to generate a more dense point cloud (Furukawa and Ponce, 2009, Hirschmueller, 2005).

Most of the UAVs used as photogrammetric sensor platforms and even some of today's cameras are equipped with a global navigation satellite system (GNSS) such as GPS, an electronic compass, barometric pressure sensors for altitude and an inertial measurement unit (IMU) to estimate the platform orientation within 1-2 meters in position and 1-2° orientation accuracy (Pfeifer et al., 2012) for direct georeferencing (Nilosek and Salvaggio, 2012). Nevertheless, these parameters are just an approximation for metric and automated applications. In general, the uncertainty in the position estimation and camera orientation by GNSS and IMU on-board sensors does not allow for sufficient accuracy (due to attenuation/blocking, shadowing or reflection of GPS signals near buildings, steep slopes, special materials, etc.) that is necessary for fully automated image-based reconstructions and measurements.

Many of these afore mentioned 3D vision methods demonstrate increasing robustness and result in high quality and visually appealing models. However, the model uncertainty of the reconstructions is not always clear and so they are often not directly suited for photogrammetric applications. Many methods either use a fixed given calibration or try to estimate camera parameters during the processing, but nearly all of them include manual

*Corresponding author.

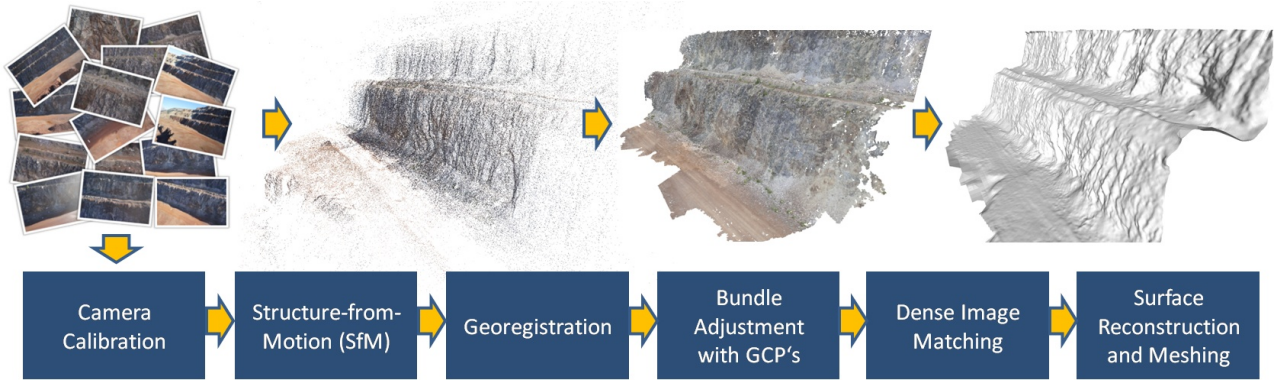

Figure 1: Automated processing workflow for precise and geo-accurate reconstructions. Top row: Image set, sparse reconstruction, dense point cloud and triangle-based surface mesh of a quarry wall in open pit mining.

steps e.g. for indirect georeferencing to establish tie point correspondences and aligning the reconstructions in a world coordinate frame. In this context, we see the need for a user-friendly, fully automated processing pipeline including user guidance during image acquisition, an easy to use camera calibration procedure and accurate georeferencing of reconstructed models in a world coordinate system having absolute geographic position and orientation with predictable reconstruction accuracy.

In this paper we propose the use of uniquely perceptible fiducial markers such that they can be automatically detected, reconstructed and integrated into the different steps of a fully automated end-to-end workflow to obtain precise and geo-accurate reconstructions (Figure 1). Our contribution is three-fold. Firstly, we present and advocate the use of planar printed paper based fiducial markers as a target pattern to obtain accurate and reliable camera calibration. Secondly, we integrate online feedback (Hoppe et al., 2012) during the data acquisition step to ensure that acquired images are suited for subsequent automated photogrammetric processing and satisfy predefined quality requirements. Automated processes impose high demands on the quality and especially on the geometric configuration of the images. Photogrammetric methods need to handle and cope robustly with convergent, oblique and unordered sets of images, where scale, depth and ground sampling distance (GSD) changes are immanent. This step thus ensures that the final Structure-from-Motion (SfM) reconstruction meets the accuracy requirements and results in a complete reconstruction of the object. Lastly, with known ground control point (GCP) positions of the markers, we are able to automatically set the generated 3D models into its geographic reference coordinate frame by subsequent indirect geo-referencing. By additionally integrating GCPs and optimization for camera calibration parameters in the bundle adjustment, we show that we are able to create precise and geo-accurate 3D reconstructions needing no manual interaction.

In the following sections, we outline the workflow of our automated multi-image reconstruction pipeline from image acquisition and camera calibration to processing and georeferencing in detail. We show in an extensive evaluation that our method allows geo-accurate position measurements with accuracies in the range of a few centimeters. In two typical scenarios and datasets in open pit mining (Figure 7) and architectural facade reconstruction (Figure 3) we show that a measurement uncertainty of about 1.5 times the ground sampling distance is reached.

2 RECONSTRUCTION PIPELINE

In this section, we describe a fully automated processing pipeline for the reconstruction of camera positions, geo-accurate 3D object geometry and the generation of surface models. Our image-based reconstruction pipeline can be roughly divided into four parts: camera calibration; determination of the exterior parameters of camera positions and orientations together with the reconstruction of sparse 3D object points; geo-registration by transforming the 3D model into a geographic reference coordinate frame; densification of the object points and generation of a polygonal surface model.

Figure 2: Fiducial markers, typical calibration image with printed marker sheets arranged on the planar surface of a floor and reliably detected markers with center position and ID.

2.1 Calibration

Accurate intrinsic camera calibration is critical to most computer vision methods that involve image based measurements, and often a prerequisite for multi-view stereo (MVS) tasks (Stretcha et al., 2008). In particular, accuracy of Structure-from-Motion computation is expected to be higher with an accurately calibrated setup as shown in the work of (Irschara et al., 2007). Given its crucial role with respect to overall precision, camera calibration has been a well-studied topic over the last two decades in both photogrammetry and computer vision. However, in most of the calibration literature, a strict requirement on the target geometry and a constraint to acquire the entire calibration pattern has been enforced. This is often a source of inaccuracy when calibration is performed by a typical end-user. Hence, aiming at the accuracy of target calibration techniques without the requirement for a precise calibration pattern, we advocate the use of a recently proposed fiducial marker based camera calibration method from (Daftry et al., 2013).

The calibration routine follows the basic principles of planar target based calibration (Zhang, 2000) and thus requires simple printed markers to be imaged in several views. Each marker includes a unique identification number as a machine-readable black and white circular binary code, arranged rotationally invariant around the marker center. The marker patterns are printed on several sheets of paper and laid on the floor in an approximate grid pattern (see Figure 2). There is no strict requirement for all markers to be visible in the captured images. To ensure that each marker is classified with a high degree of confidence and to eliminate any false positives in case of blurred or otherwise low quality images, a robust marker detection is employed (see also Section 2.3.1). An initial estimate of lens distortion parameters

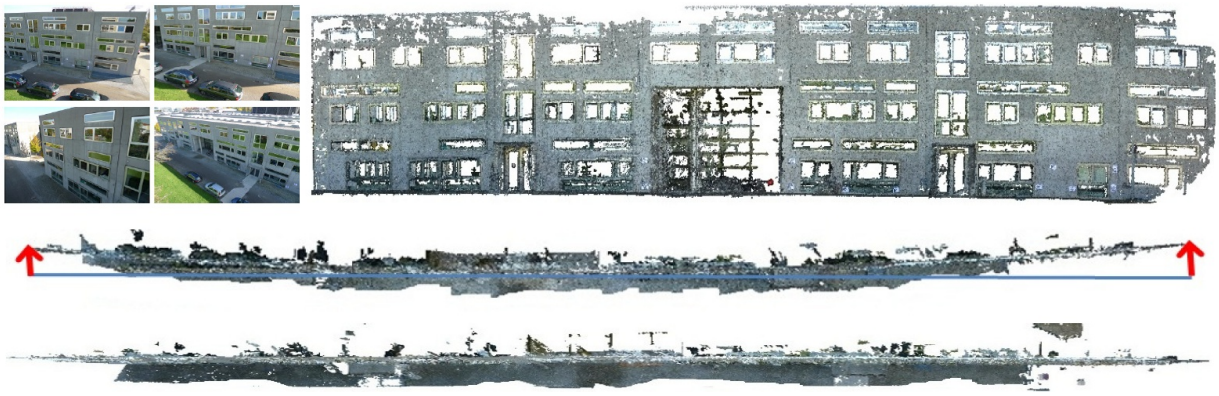

Figure 3: In the facade reconstruction using OpenCV calibration significant bending is prevalent (middle). In contrast, accurate camera parameters delivered by the method from (Daftry et al., 2013) results in a straight facade reconstruction (bottom).

attempts to minimize the reprojection error of extracted feature points based on homographies between the individual views. After determining the constant, but unknown focal length f and determining the calibration matrix K, bundle adjustment is applied in a subsequent step to perform non linear optimization of the intrinsics (f,u_o,v_o) with u_o,v_o the principal point and radial distortion θ. For details on the calibration routine please refer to the original paper.

Significant accuracy gains, both quantitative and qualitative, can be observed using the proposed method. Figure 3 shows a reconstruction of a facade that, although visually correct in appearance, suffers from geometric inconsistencies (significant bending) that is prevalent along the fringes when using standard calibration and undistortion results from OpenCV (Bradski, 2000). In contrast, using accurate camera parameters delivered by the method proposed in (Daftry et al., 2013) results in an almost straight wall.

2.2 Structure-from-Motion

Steps for the calculation of the exterior camera orientation include feature extraction and feature matching, estimation of relative camera poses from known point correspondences, incrementally adding new cameras by resection and computation of 3D object coordinates of the extracted feature points, followed by bundle adjustment to optimize camera orientations and 3D coordinates of the object points. A variety of methods exist for automated extraction and description of salient feature points. A well known method that is robust to illumination changes, scaling and rotation is the scale-invariant feature transform (SIFT) (Lowe, 2004). Since we assume that no further information about the images are known, feature matching requires an exhaustive comparison of all the extracted features in an unordered image set between all possible pairs of images. The expense related to correspondence search and matching is thus quadratic in terms of the number of extracted feature points in the scene, which can lead to a critical amount of time in data sets with several thousands of images. The complete correspondence analysis between all possible pairs of images is necessary in order to preserve as many image measurements as possible for an object point. The number of image measurements and a large triangulation angle is important for reconstruction accuracy. The theoretically optimal intersection angle is at 90°. Practically relevant is an angle between 20° and up to 40° when using the SIFT descriptor. By geometric verification of the found feature correspondences based on the five-point algorithm (Nistér, 2003), the relative camera orientations between image pairs can now be estimated and represented as an epipolar connectivity graph (Figure 4). Because the corresponding image measurements of the feature point matches may be corrupted by outliers and gross errors, the verification of the relative camera orientations is performed by means of a robust estimation method within a RANSAC loop (Fischler and Bolles, 1981). The determined epipolar graph expresses the spatial relations between images and represents the pairwise reconstructions and relative camera orientations, wherein the nodes of the graph represent the images and the edges correspond to the relationships and relative orientations between them. Starting from an initial image pair, new images are incrementally added to the reconstruction using the three-point algorithm (Haralick et al., 1991). Camera orientations and 3D point coordinates are then simultaneously optimized and refined using bundle adjustment (Triggs et al., 2000) by minimizing the reprojection error/residual between image measurements of the SIFT features and predicted 3D coordinates of the corresponding object point, formulated as a non-linear least squares problem. The results of the fully automated workflow so far are the outer orientations of the cameras and the reconstructed object geometry as a sparse 3D point cloud.

(a) (b)

Figure 4: Rows and columns of the epipolar graph represent individual cameras and their connections to each other based on shared feature matches and image overlap. (a) shows a traditional aerial survey with regular flight pattern, (b) the connections between cameras for an unordered oblique image data set.

2.3 Automated Marker-based Georeferencing

The reconstruction and external orientation of the cameras so far is initially in a local Euclidean coordinate system and only up to scale and therefore not metric. However, in most cases the absolute position accuracy of the measured object points is of interest. In addition, we want the created 3D model correctly stored and displayed in position and orientation in its specific geographic context. Based on known point correspondences between reconstructed object points and ground control points (GCPs), we first transform the 3D model from its local source coordinate frame into a desired metric, geographical target reference frame by shape-preserving similarity transform, also known as Helmert transformation (Watson, 2006). The transformation parameters for rotation, translation and scaling can be computed from a minimum number of three non-collinear point correspondences between reconstructed model and reference measurements. In prac-

tice, more than three point correspondences are used to allow for better accuracy in the registration of the model. Again, the method of least squares is used within a RANSAC loop to estimate a robust fit of the model to the reference points.

2.3.1 Marker-based Rigid Model Geo-Registration

To facilitate automation and to avoid erroneous point correspondences by manual control point selection, the association of point correspondences between image measurements and ground control points is encountered again using fiducial markers introduced for camera calibration. A requirement for full automation is that markers are detected robustly and stable in at least two images of the dataset and are clearly identified individually by their encoded ID. The detection also needs to work reliably from different flying altitudes and distances from the object. Instead of paper print outs, we make use of larger

Figure 5: Histogram for an unrolled circular marker and rotation invariant binning of the code stripe. The numbers from top to bottom indicate the probabilities for center, binary code and outer circle. The marker with ID 20 has been successfully decoded.

versions (∼50 cm diameter) of the markers printed on durable weather proof plastic foil to signal reference points in the scene used as GCPs. The markers are flexible, rolled up easy to carry, though robust and universally applicable even in harsh environments.

The markers are equally distributed in the target region and placed almost flat on the ground or attached to a facade. The 3D positions of the marker centers are then measured by theodolite, total station or differential GPS with improved location accuracy (DGPS), which is the only manual step in our reconstruction workflow besides image acquisition. All input images are then checked in the marker detection. After thresholding and edge detection, we extract contours from the input images and detect potential markers by ellipse fitting. The potential markers are then rectified to a canonical patch and verified, if circles are found using Hough transform. If the verification is positive, we sample the detected ellipse from the original gray scale image to unroll it and build a histogram (Figure 5). In the thresholded and binned histogram we extract the binary code of the marker if the code probability is high. The marker ID is obtained by checking the code in a precomputed lookup table.

The detected ellipse center describes the position of the image measurement of the respective marker (see Figure 6). By triangulating multiple image measurements of one and the same marker seen in several images, we calculate its 3D object point position in the local reference frame of the model. The markers can be directly matched and automatically associated with their corresponding ground control reference points as long as they share the same ID. Once, corresponding point pairs have been established between model and reference, they are used to calculate the transformation parameters. This results in a geo-registered model of the object.

2.3.2 Constrained Bundle Block Optimization with GCPs

In a purely image-based reconstruction it can be observed that the error at the border parts of the reconstruction slightly increases, as already shown in Figure 3. This leads to a deformation of the image block in the bundle adjustment. Two reasons can be

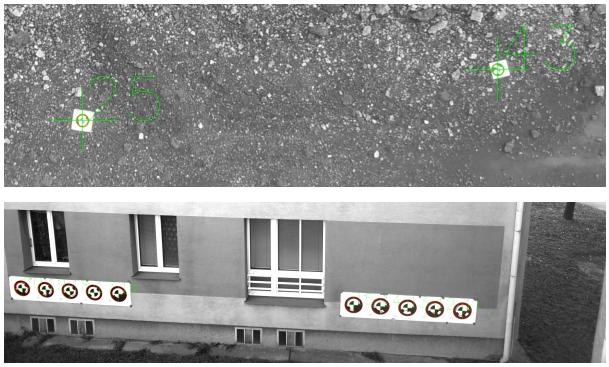

Figure 6: Automatically detected ground control points with plotted marker centers and corresponding marker IDs.

identified that cause this type of deformation. First cause is the quality of the calibration. An improper or inaccurate calibration leads to false projections of features and image measurements, e.g. due to a misalignment of the camera's principal point. A wrongly estimated focal length of the camera shifts the problem from the intrinsics away and may get compensated by the distortion parameters and vice versa. But, the original problem persists which is the reason for camera calibration to be carried out with adequate care. Even with a carefully calibrated camera, the parameters may change during image acquisition, e.g. due to harsh weather conditions and a temperature change between last calibration and the time of survey.

The second reason causing drift lies in the actual camera network. This can be explained that the scene is covered by fewer images towards the borders of the surveying area compared to the center of the object. Less image overlap leads to fewer image measurements per object point and thus cause the camera network and epipolar graph, as described in Section 2.2, to have fewer connections at the borders. This has the effect that the optimization of camera positions and 3D object points in the bundle adjustment is less constrained, thus the optimized positions can undergo larger changes.

A way to avoid systematic errors arising from the deformation of the image block is to introduce known reference points as well as directly measured GPS coordinates of the camera positions in the bundle adjustment. On the one hand, this leads to smaller position error residuals, on the other hand a simultaneous transition into a reference coordinate system can be accomplished. The additional information introduced by the artificial fiducial markers in the scene can be seamlessly integrated into the reconstruction process. The global optimization of camera positions and 3D object point positions in the bundle adjustment is carried out based on Google's Ceres Solver for non-linear least squares problems (Agarwal et al., 2012). For this purpose, we perform an additional bundling step and extend the bundle adjustment besides the mass of object points from natural landmarks from SIFT points by a second class for ground control points and image measurements of the fiducial marker detections. In contrast to a naive approach constraining the 3D position errors of the GCPs, we avoid any metric scale problems and penalize the reprojection error of the GCPs in image space with a Huber error norm (Huber, 1964). The additional reference points are intended to constrain the optimization problem so that the solution sticks to the GCPs. Furthermore, we let the initial camera parameters for intrinsics and lens distortion be commonly refined and corrected for all cameras in the bundle adjustment step. Thus, GCP integration distributes the residual errors equally over all markers and allows 3D reconstructions with very low geometric distortions, even for elongated objects of large extent.

2.4 Densification and Meshing

Due to the comparably low number of triangulated feature points (about 500-1000 features per image, depending on the texture) and their non-uniform distribution on the surface compared to the dense number of pixels in one image (millions of pixels), the modeling of the surface is only an approximation of the real surface. To increase the number of 3D points, stereo (Hirschmueller, 2005) or multi-view methods (Irschara et al., 2012, Furukawa and Ponce, 2009) can be used for pixel-wise image matching. Stereo methods use two rectified images to determine a depth for each pixel in the image. In the case of unordered oblique images that were taken from a UAV, image rectification is often only possible with great loss in image resolution due to the geometric configuration of the cameras. Therefore, multi-view methods that are able to cope with arbitrary camera configurations such as the freely available PMVS2 (Furukawa and Ponce, 2009) are well suited for oblique image sets. For many tasks a closed surface model is necessary, such as visibility calculations where a point cloud is not sufficient. A well-known method for surface reconstruction from a set of densely sampled 3D points also used for laser scanning data is the Poisson surface reconstruction (Hoppe et al., 1992), which interpolates the densified points to a closed surface. Figure 1 shows a comparison between a sparse reconstruction, a densified point cloud and a reconstructed triangle surface mesh of a quarry wall consisting of about 10 million 3D points.

3 IMAGE ACQUISITION

To evaluate the presented automated workflow and the achieved accuracy respectively, several image flights were carried out to record typical datasets for architectural and mining applications.

3.1 Test Site and Flight Planning

One of our test sites is located at the "Styrian Erzberg", which is the biggest iron ore open pit mine in central Europe. To asses the achieved accuracy, a reference point network of ground truth measurements is needed. Therefore, one wall (which is about 24 m high and 100 m long) is equipped with 84 circular targets. This dense target network enables an extensive evaluation of reconstruction accuracy, and systematic deformations of the image block and reconstructed 3D geometry can be quantified. In addition and especially for automated georeferencing, all together 45 binary coded fiducial markers, as described in Section 2.1 and 2.3, are used as temporary GCPs on top and bottom of the wall and in the adjacent area around the object. The spatial distribution of the markers in the target region allows different selections of GCPs, and offers the opportunity to study the influence of the GCP network on the achievable accuracy of the 3D reconstruction. In addition to the well-textured quarry wall open pit mining dataset, we also recorded a facade dataset of a building with large homogeneous areas. All reference points were conventionally surveyed using a Trimble S6 total station with an average precision of 10 mm for 3D point surveying without prism. Figure 7 shows the spatial distribution of markers (green) and targets (red).

For image acquisition we used an AscTec Falcon 8 octocopter as a flying platform. The UAV is equipped with a Sony NEX-5N digital system camera. The APS-C CMOS sensor has an image resolution of 16.1 megapixels and is equipped with a fixed focus lens with a focal length of 16 mm (Table 1). The open pit mine dataset consists of 850 images, for the facade reconstruction 997 images were captured in three rows regarding the distance to the object.

Figure 7: The reference point network allows an extensive accuracy evaluation. Systematic deformations of the image block and reconstructed object geometry can be quantified. Markers (bottom right) indicating GCP positions are shown in green, circular targets (left) for quantitative evaluation are in red.

Sensor dim.	Resolution	Focal len.	Pixel size
$23.4 \times 15.6\,mm$	4912×3264	$16\,mm$	$4.76\,\mu m$

Table 1: Camera and sensor specifications.

We define a desired minimum ground sampling distance of 1.5 cm per Pixel and a minimum overlap between images of at least 70%. Based on Equation 1 and 2 for nadir image acquisition in aerial photogrammetry,

$$PixelSize = \frac{SensorWidth\,[mm]}{ImageWidth\,[px]} \quad , \tag{1}$$

$$GSD = \frac{PixelSize\,[mm] * ElevationAboveGround\,[m]}{FocalLength\,[mm]} \quad , \tag{2}$$

we obtain a maximum flying height above ground and imaging distance to the object, respectively, of about 50 meters.

The angle of view calculates from Equation 3,

$$\alpha = 2 \cdot \arctan \frac{SensorWidth\,[mm]}{2 \cdot FocalLength\,[mm]} \quad , \tag{3}$$

to $\alpha = 72.35°$. The scene coverage for one image captured from height h above ground can be calculated from Equation 4,

$$c = 2 \cdot h \cdot \tan \frac{\alpha}{2} \approx ImageWidth\,[px] \cdot GSD \quad . \tag{4}$$

The required baseline b between views then calculates from the overlap ratio $o_r = \frac{o}{c}$ with o being the overlap $o = 2 \cdot h \cdot \tan \frac{\alpha}{2} - b$ to $b = (1 - o_r) \cdot c$, resulting in a maximum baseline between images of $b = 21.94\,m$ in 50 meters distance to the object and $b = 4.39\,m$ in close distance of 10 meters in front of the object. These geometric requirements together with the maximum resolution of the camera also constrains the size of the markers in the scene, since the robust decoding of the marker IDs requires an image of the marker with at least 25-30 pixels in diameter. The minimum marker size then yields a marker size of approximately 45-50 cm in diameter to be robustly decoded from a distance of 50 meters.

To enable analysis of which parameters influence the reconstruction accuracy with respect to triangulation angle, redundancy (i.e. overlap), distance to the object and camera network, we perform an oversampling of the scene and therefore record images with approximately 90% overlap in three different distances and heights from the object.

3.2 Online-Feedback for Image Acquisition

We support image acquisition by an online feedback system to assess the recorded images with respect to the parameters of image overlap, ground sampling distance and scene coverage defined in the previous section to ensure completeness and redundancy of the image block.

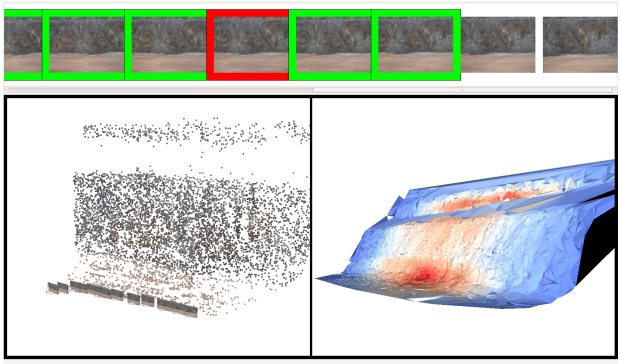

Figure 8: Visualization of ground resolution using an online Structure-from-Motion system to assess reconstruction accuracy and scene coverage during image acquisition.

The automated offline pipeline described in Section 2.2 yields high accuracy, as we will show in Section 4. However, processing takes several hours, thus, results are only available hours later. If the reconstruction result does not meet the desired expectations, e.g. due to a lack of images in areas that would have been relevant to the survey, a repetition of the flight is necessary which cause additional costs and time delays. In order to be able to already judge the results on site whether the captured images are suited for fully automated processing, we apply a recently developed method for online Structure-from-Motion (Hoppe et al., 2012). The method calculates the exterior orientation of the cameras and a sparse point cloud reconstruction already during or right after the flight on site. The images may be streamed down from the UAV via an SD card in the camera with Wi-Fi capability to a laptop computer on the ground. The method is able to process high-resolution images on a notebook with quad-core CPU and powerful graphics card in real time. The method requires a stream of consecutively captured images and needs about two seconds to determine the outer orientation of a 10 megapixel image and to calculate new object points. Due to the restrictions in image matching and bundle adjustment to immediate neighboring cameras, the online reconstruction does not claim high accuracy. However, the method allows the estimation of achievable reconstruction quality and is very beneficial to determine completeness of the final reconstruction during image recording.

For the user, the quality of the reconstruction can be judged only poorly from the triangulated sparse feature points. Two main relevant parameters determine the accuracy: redundancy, which states how often a surface point is seen, and the ground resolution. To determine both parameters from the actual reconstruction, a surface is extracted from the sparse points using (Labatut et al., 2007). This surface is then used to visualize ground resolution and redundancy of the reconstruction using color coding. For the pilot it is then apparent, which parts of the scene are observed often and at which ground resolution they can be reconstructed. This assists the pilot in planning the next steps of the flight so that a uniform coverage of the scene with constant ground resolution can be achieved. Figure 8 shows the reconstruction result during image acquisition.

4 EVALUATION AND RESULTS

In this section we analyze the performance of our presented automated method for georeferenced 3D reconstructions. In literature, the reprojection error of object or feature points has often been used as an adequate measure for evaluating the accuracy of the exterior camera orientation. However, for photogrammetric applications the accuracy of the reconstructed object points is of prime interest. We thus perform a point-wise comparison of reconstructed object points to corresponding, clearly identifiable 3D reference point coordinates from circular targets. Since the reconstruction has been already geo-registered by a rigid transformation, we approximately know the location of the individual target points in the images, thus we perform a guided search for circular targets in the images for each of the reference points (see Figure 7). The object points for comparison with the known ground truth point positions are then triangulated from the center point measurements of the detected circular targets in the images.

It can be shown that highly geo-accurate reconstructions are obtained with our system. In Figure 9, we show the absolute point error for each evaluation target in the quarry wall after constrained bundle adjustment with GCPs. Using all 850 images of the open pit mining dataset and all available GCPs for the bundle block optimization a mean accuracy of less than 2.5 cm is reached. For the facade dataset we are even able to achieve an overall accuracy of 0.5 cm due to the closer distance to the object (4-10 m) and resulting a much higher GSD, respectively. To avoid

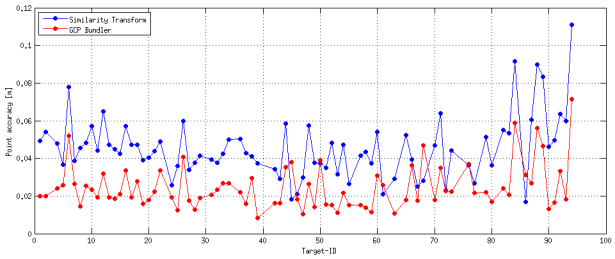

Figure 9: Using all 850 images and all available GCPs in the constrained bundle adjustment, a mean measurement uncertainty dropping from 4.54 cm to below 2.45 cm is reached.

systematic deformations of the image block, we use our fiducial markers as ground control points (GCPs) in a constrained bundle adjustment. Table 2 shows the improvement in accuracy by comparing the mean absolute point error for rigid similarity transform and optimization using GCP constrained bundle adjustment. The mean error which is already very good before GCP bundling then drops further from 3 times the GSD to a factor of 1.5 times the GSD. The decreasing standard deviation indicates an equalization of the error distribution and a less deformed image block after the optimization.

Method	Mean	Std.dev.	Median
Similarity Transform	4.54 cm	1.64 cm	4.40 cm
GCP bundler	2.45 cm	1.18 cm	2.16 cm

Table 2: Accuracy improvement by GCP constrained bundle adjustment

Next, we investigate relevant parameters influencing the reconstruction accuracy, which is important to understand better the aspects of block stability and the most influencing factors. A high oversampling of the quarry wall was done for this purpose and is represented in the open pit mining dataset as mentioned in section 3.1. The most prominent parameters that have a large impact on accuracy are, besides image overlap and triangulation

angle, foremost the ground sampling distance determined by image resolution and the imaging distance to the object. In order to identify and quantify these parameters and furthermore give guidelines for image acquisition, a systematic parameter analysis is carried out. We build different image subsets for both the open pit mine as well as for the facade dataset to study the effect of different camera configurations.

4.1 Number of Observations

As shown in (Rumpler et al., 2011) in a synthetic simulation on a regular aerial flight pattern, accuracy increases with a higher number of image measurements and with increasing triangulation angles. Plotting the mean object point error over the total number of images for different subsets, it can be shown in Figure 10 that the point error decreases with increasing total number of images. Figure 10 also shows, that there is a fast saturation and accuracy improvement within larger datasets.

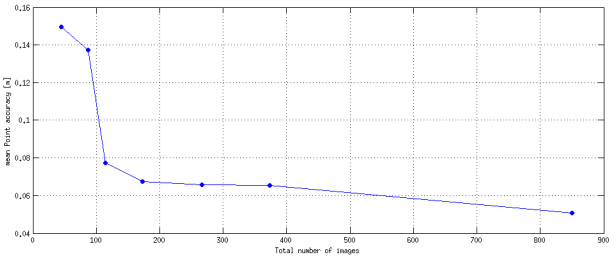

Figure 10: Error curve for different image subsets before GCP constrained bundle adjustment. With increasing total number of images per subset, the mean reconstruction accuracy increases.

It can be shown that a higher number of images in the dataset leads to an accuracy improvement. But, considering the number of image measurements per evaluation target does not necessarily reduce the error and does not directly lead to higher accuracy as shown in Figure 11. The error in the graph looks rather oscillating over the track length. Thus, it is not possible to exemplify the achievable accuracy alone on the number of used images or observations for unordered and oblique datasets as the camera configuration and its influence on feature matching, triangulation angle and ray intersection geometry may change drastically. From Figure 11 we argue that not every additional image

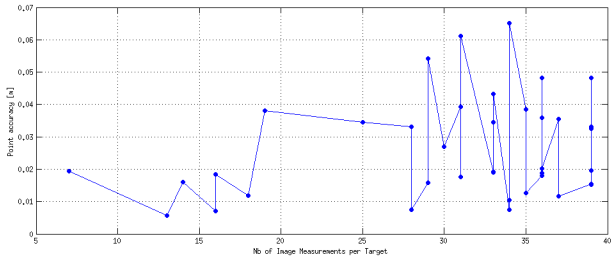

Figure 11: Mean evaluation target track length. A larger number of image measurements has not necessarily a positive effect on the achievable accuracy. The graph looks uniformly distributed or rather oscillating.

measurement necessarily leads to an improvement in accuracy. There are various influences which can not be compensated by high redundancy. Each image contributes to the resulting accuracy through different ways. Apart from redundancy especially camera network, triangulation angle as well as the quality of feature detection and localization of image measurements and target points due to image noise have large influence.

4.2 Camera Network and Resolution

As indicated above, the influence of the geometric configuration of the multi-view camera network on the resulting accuracy is higher than the influence of redundancy in image acquisition. To investigate the influence of the distance to the object and ground sampling distance, i.e. image resolution respectively, we build different image subsets. The direction of the camera's optical axis is kept constant in narrow bounds so that the cameras are looking almost perpendicular to the wall and the facade. The distance to the object varies from close to distant in three different rows (15 m, 35 m and 50 m). In each row we select images to ensure a constant overlap of 70% for all three rows between the views. First, each row is processed separately and subsequently all rows combined. The distance to the object has a large and systematic influence on the achievable accuracy. With increasing distance to the object the calculated error increases as well. Figure 12 shows the mean error for all targets with respect to the different subsets and distances to the wall. The best accuracy can be achieved using images from all three rows together.

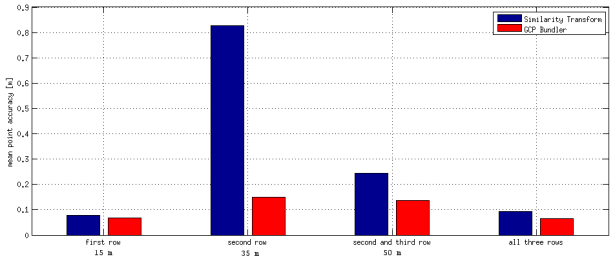

Figure 12: Considering the mean error over all targets, it can be observed, that the mean error increases with larger distances to the object, and with decreasing ground sampling distance, i.e. resolution respectively.

Flying at different altitudes is a common approach in airborne photogrammetry to enhance the accuracy of the IO parameters, thus a combination of the different flying heights/distances delivers better results. This indicates that the images of the first row influence the accuracy positively due to their higher GSD, but the first row is more affected by drift and distortions, which is not obvious directly from the shown mean error. Images that are further away cause larger errors. On the one hand, this is caused by a lower ground sampling distance, and thus, lower level of detail in the images. On the other hand, the influence of image noise on the reconstruction uncertainty increases with point depth. Image noise and small localization errors of the image measurements are approximately constant for all images, however, the resulting positional error increases with larger distances due to a larger angular error in the triangulation. Nevertheless, the combination of different image resolutions effects the achievable accuracy positively, because images taken further away mitigate the error propagation within each row and help connecting the camera network over longer tracks. Additionally, the first row consists of 173 images, whereby the second and third row are containing only half the number, namely 87 images. According to Figure 10 the higher number of images in the first row leads also to a higher accuracy.

In aerial photogrammetry, the typical depth error for a triangulated 3D object point is in the range of 1.5-2 times the GSD. Based on Equation 2 and a GSD of 1.5 cm per Pixel in images taken from 50 m distance, the expected point uncertainty would then be in the range of 2.25 cm. Overall, we achieve a mean position error of the object points of 2.45 cm (Table 2) which is perfectly in consent with the expected measurement uncertainty.

5 CONCLUSIONS

We have presented a system for fully automated generation of precise and georeferenced 3D reconstructions based on fiducial markers. Firstly, we advocated the use of planar printed paper based fiducial markers as a target pattern to obtain accurate and reliable camera calibration. Secondly, we integrated an online feedback to guide the user during data acquisition regarding ground sampling resolution and image overlap so that automated photogrammetric processing is possible, the final reconstruction meets predefined accuracy requirements, and results in a complete reconstruction of the object. Lastly, we utilize known ground control points signalled by fiducial markers in the scene and integrate them into our image-based reconstruction pipeline. The produced 3D models are accurately reconstructed and transformed into a geographic reference coordinate frame by seamlessly integrating the GCPs given by the markers and additional optimization of camera calibration parameters in the bundle adjustment.

We showed that our approach is able to produce very good results in two typical scenarios and datasets in open pit mining and an architectural facade reconstruction. We achieve an equally distributed measurement uncertainty of about 1.5 times the ground sampling distance. The most prominent parameter with large impact on accuracy is, besides image overlap and triangulation angle given by the camera network, foremost the ground sampling distance determined by image resolution and imaging distance to the object.

In the case of nadir aerial imaging mainly the camera network geometry is crucial for determining reconstruction accuracy, but that cannot be inferred to unordered image sets of oblique views straight forward. Combining images taken at different distances leads to better block stability of the camera network and points and helps to enhance the accuracy of the IO parameters. But there are various influences which cannot be compensated by high resolution, redundancy or larger triangulation angles. Apart from those parameters, the quality of feature detection and localization of image measurements has a large influence due to heavily changing view points, illumination changes or image noise. This will be the subject of future research.

ACKNOWLEDGEMENTS

This work has been supported by the Austrian Research Promotion Agency (FFG) BRIDGE program under grant 3770439.

REFERENCES

Agarwal, S., Mierle, K. and Others, 2012. Ceres Solver. `https://code.google.com/p/ceres-solver/`.

Bradski, G., 2000. The OpenCV Library.

Daftry, S., Maurer, M., Wendel, A. and Bischof, H., 2013. Flexible and User-Centric Camera Calibration using Planar Fiducial Markers. In: British Machine Vision Conference (BMVC).

Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. Communication Association and Computing Machine 24(6), pp. 381–395.

Furukawa, Y. and Ponce, J., 2009. Accurate, Dense, and Robust Multi-View Stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI).

Haralick, R. M., Lee, C., Ottenberg, K. and Nölle, M., 1991. Analysis and Solutions of the Three Point Perspective Pose Estimation Problem. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 592–598.

Hartley, R. and Zisserman, A., 2004. Multiple View Geometry in Computer Vision. Second edn, Cambridge University Press.

Hirschmueller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Hoppe, C., Klopschitz, M., Rumpler, M., Wendel, A., Kluckner, S., Bischof, H. and Reitmayr, G., 2012. Online Feedback for Structure-from-Motion Image Acquisition. In: British Machine Vision Conference (BMVC).

Hoppe, H., DeRose, T., Duchamp, T. and J. McDonald, W. S., 1992. Surface reconstruction from unorganized points. In: SIGGRAPH, pp. 71–78.

Huber, P. J., 1964. Robust Estimation of a Location Parameter. The Annals of Mathematical Statistics 35(1), pp. 73–101.

Irschara, A., Rumpler, M., Meixner, P., Pock, T. and Bischof, H., 2012. Efficient and Globally Optimal Multi View Dense Matching for Aerial Images. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences.

Irschara, A., Zach, C. and Bischof, H., 2007. Towards wiki-based dense city modeling. In: IEEE International Conference on Computer Vision (ICCV).

Labatut, P., Pons, J. P. and Keriven, R., 2007. Efficient Multi-View Reconstruction of Large-Scale Scenes using Interest Points, Delaunay Triangulation and Graph Cuts. In: IEEE International Conference on Computer Vision (ICCV).

Leberl, F., Irschara, A., Pock, T., Meixner, P., Gruber, M., Scholz, S. and Wiechert, A., 2010. Point Clouds: Lidar versus 3D Vision. Photogrammetric Engineering and Remote Sensing.

Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision (IJCV) 60, pp. 91–110.

Nilosek, D. and Salvaggio, C., 2012. Geo-Accurate Dense Point Cloud Generation. `http://dirsapps.cis.rit.edu/3d-workflow/index.html`.

Nistér, D., 2003. An efficient solution to the five-point relative pose problem. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 195–202.

Pfeifer, N., Glira, P. and Briese, C., 2012. Direct georeferencing with on board navigation components of light weight UAV platforms. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.

Rumpler, M., Irschara, A. and Bischof, H., 2011. Multi-View Stereo: Redundancy Benefits for 3D Reconstruction. In: 35th Workshop of the Austrian Association for Pattern Recognition.

Stretcha, C., Von Hansen, W., Van Gool, L., Fua, P. and Thoennessen, U., 2008. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8.

Triggs, B., McLauchlan, P., Hartley, R. and Fitzgibbon, A., 2000. Bundle Adjustment - A Modern Synthesis. In: Vision Algorithms: Theory and Practice, pp. 298–375.

Watson, G. A., 2006. Computing Helmert transformations. In: Journal of Computational and Applied Mathematics, Vol. 197, pp. 387–395.

Zhang, Z., 2000. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 22(11), pp. 1330–1334.