ON THE USE OF UAVS IN MINING AND ARCHAEOLOGY - GEO-ACCURATE 3D RECONSTRUCTIONS USING VARIOUS PLATFORMS AND TERRESTRIAL VIEWS

Alexander Tscharf[a,*], Markus Rumpler[b], Friedrich Fraundorfer[b], Gerhard Mayer[a] and Horst Bischof[b]

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

[b] Institute for Computer Graphics and Vision, Graz University of Technology, Austria
(`rumpler, fraundorfer, bischof`)`@icg.tugraz.at`

KEY WORDS: Photogrammetric Computer Vision, Unmanned Aerial Vehicles, Image-based 3D Reconstruction, Mapping, Georegistration, Accuracy Evaluation, Structure from Motion, Land Surveying, Mining, Archeology

ABSTRACT:

During the last decades photogrammetric computer vision systems have been well established in scientific and commercial applications. Especially the increasing affordability of unmanned aerial vehicles (UAVs) in conjunction with automated multi-view processing pipelines have resulted in an easy way of acquiring spatial data and creating realistic and accurate 3D models. With the use of multi-copter UAVs, it is possible to record highly overlapping images from almost terrestrial camera positions to oblique and nadir aerial images due to the ability to navigate slowly, hover and capture images at nearly any possible position. Multi-copter UAVs thus are bridging the gap between terrestrial and traditional aerial image acquisition and are therefore ideally suited to enable easy and safe data collection and inspection tasks in complex or hazardous environments. In this paper we present a fully automated processing pipeline for precise, metric and geo-accurate 3D reconstructions of complex geometries using various imaging platforms. Our workflow allows for georeferencing of UAV imagery based on GPS-measurements of camera stations from an on-board GPS receiver as well as tie and control point information. Ground control points (GCPs) are integrated directly in the bundle adjustment to refine the georegistration and correct for systematic distortions of the image block. We discuss our approach based on three different case studies for applications in mining and archaeology and present several accuracy related analyses investigating georegistration, camera network configuration and ground sampling distance. Our approach is furthermore suited for seamlessly matching and integrating images from different view points and cameras (aerial and terrestrial as well as inside views) into one single reconstruction. Together with aerial images from a UAV, we are able to enrich 3D models by combining terrestrial images as well inside views of an object by joint image processing to generate highly detailed, accurate and complete reconstructions.

1 INTRODUCTION

The use of unmanned flying platforms for aerial data acquisition has increased enormously in recent years. Besides a variety of commercially sold multirotor- and small fixed-wing aircraft models, there exists a committed community that designs and builds a wide variety of do-it-yourself aircraft that find use in various fields and applications. Typical fields of application reach from agriculture and environmental monitoring, surveying tasks for mining, archaeology or architecture as well as inspection and assessment of objects that are difficult and dangerous to reach for human operators.

Creating and visualizing realistic and accurate 3D models is becoming a central ambition of research in the field of geodetic data acquisition. Especially photogrammetric methods and image based measurement systems have been increasingly used in recent years. These systems have become very popular, mainly due to their inherent flexibility compared to traditional surveying equipment. Photogrammetric methods can be roughly classified based on the data acquisition strategy into aerial and terrestrial. Large-scale digital surface models are created from aerial photographs while terrestrial images are used for detailed object reconstructions of small and medium-sized close-range objects, e.g. for architectural and archaeological 3D documentation and preservation or mapping of quarry walls for blast design in open pit mining (Moser et al., 2006).

While manned aerial photogrammetry is only economical for large survey areas due to a high demand on resources, terrestrial photogrammetry is limited to ground based camera posi-

tions, thus visibility problems may arise depending on the scene geometry. Especially in mining or in cultural heritage reconstruction it is often not possible to access the object due to safety reasons. In this context photogrammetry with Unmanned Aerial Vehicles (UAVs) has recently emerged as a low-cost alternative to traditional manned surveying as well as to terrestrial photogrammetry. It helps to overcome geometrical constraints, closes the gap between aerial and terrestrial photogrammetry and combines the advantages of both. Multi-copter UAVs in particular, are able to record highly overlapping images from almost terrestrial camera positions to oblique and nadir aerial images due to the ability to navigate at very low airspeed, hover and capture images at nearly any possible position. Low-cost and low-weight UAV systems equipped with affordable high quality digital consumer-grade cameras present a considerable potential for close-range remote data acquisition in various fields of application (Rehak et al., 2013). Together with an automated multi-view processing pipeline, 3D reconstructions and dense point clouds from images can be generated in a more flexible, faster and cheaper way and can easily compete with point clouds from laser scans (Leberl et al., 2010).

Fully automated methods for image-based 3D reconstructions originate in the field of image processing (Hartley and Zisserman, 2004) and have now been integrated in many, partly freely available software packages (e.g. VisualSfM, Acute3D, Pix4D, Agisoft PhotoScan, PhotoModeler, etc.). These methods are able to calculate the intrinsic and extrinsic camera parameters as well as scene structure represented as a (sparse) 3D point cloud from an unordered set of images. Many of the mentioned 3D vision methods show increasing robustness and result in high quality
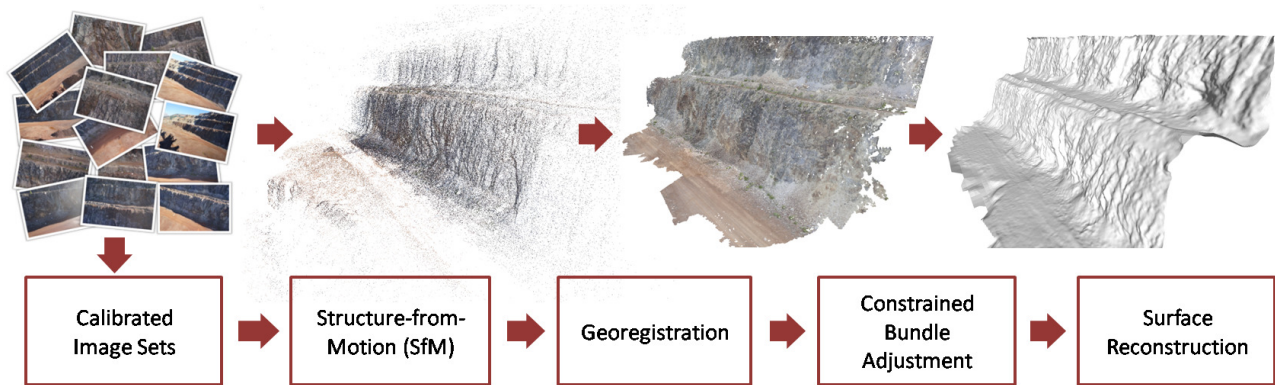
---

*Corresponding author.

Figure 1: Automated processing workflow for 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.

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. In this context, we present a user-friendly, fully automated processing pipeline, able to integrate images taken with different cameras in one single reconstruction and outputs an accurate georeferenced model with absolute geographic position and orientation and predictable reconstruction accuracy (Figure 1). Automated processes impose high demands on the quality and on the geometric configuration of the images. Especially complex object geometries require high overlap and a very dense image network to guarantee completeness, which cannot be ensured by using terrestrial or aerial nadir images exclusively. Only a combination of terrestrial and aerial viewpoints is able to guarantee completeness of the model.

We present in this paper a fully automated end-to-end workflow (Figure 1) to obtain precise and geo-accurate reconstructions especially for complex environments by the combined use of different camera platforms (aerial and terrestrial as well as inside views). The following sections outline the workflow of our fully automated multi-view reconstruction pipeline in more detail and evaluate how to obtain geo-accurate reconstructions for complex object geometries with high precision using UAVs in combination with terrestrial images. In three typical scenarios and datasets on surveying in open pit mining (Figure 4 and Figure 5) and archaeology (Figure 6) we show that highly accurate 3D reconstructions can be achieved.

2 RECONSTRUCTION PIPELINE

In this section, we describe our fully automated multi-view processing pipeline to reconstruct geo-accurate 3D models and camera positions with input images captured with different cameras at different scales and view points. The reconstruction pipeline takes pre-calibrated images, groups them according to their intrinsic parameters and processes them to generate textured polygonal surface models by performing the following steps:

- feature extraction and matching,
- Structure-from-Motion (SfM) / Aerial Triangulation (AT),
- geo-registration,
- meshing and texturing.

2.1 Structure-from-Motion

Calculation of the exterior camera orientations include feature extraction and feature matching, estimation of relative camera poses from known point correspondences and incrementally adding

new cameras by resection and computation of 3D object coordinates of the extracted feature points. Camera orientations and 3D coordinates of the object points are then optimized using bundle adjustment.

For our method we assume pre-calibrated images, i.e. images that have already been undistorted together with an initial guess of the focal length. We use the calibration routine of (Daftry et al., 2013). We group all input images into subsets sharing the same camera and focal length in a preprocessing step. The grouping and assignment to an initial calibration and focal length is performed according to meta information from specific tags provided with the image file (e.g. Exif information in JPEG or TIFF images), or given by the user.

The first processing step in our pipeline is feature extraction on every image in all subsets. A variety of methods exist for automated detection of feature points. The scale-invariant feature transform (SIFT) (Lowe, 2004) proved to be very robust against rotation, illumination changes and view point variations and scaling. It is therefore ideally suited to match images automatically from different view points, i.e. aerial images from a UAV and terrestrial images as well as inside views of an object taken with different cameras into one single reconstruction. The only prerequisite is that there is overlap between the images showing sufficient texture and salient features that can be matched across the views. Multi-copter UAVs are very flexible tools and therefore perfectly suited for the purpose of reconstructing complex objects, because they are able to navigate very slowly or hover at any possible position, thus recording images with high overlap from almost terrestrial camera positions, oblique aerial photos to traditional nadir looking images. The extracted features for all images are then stored and further processed.

Matching of the extracted features is performed between all images and all subsets. Exhaustive comparison of all extracted features in an unordered set of images between all possible pairs requires a lot of computation time and is the most time consuming step in every Structure-from-Motion pipeline. To speed up the correspondence analysis in large data sets, methods based on vocabulary trees are applied to achieve a rough pre-selection of similar image pairs (Nistér and Stewenius, 2006, Sivic and Zisserman, 2003). The computation time for feature extraction and matching can be additionally reduced through the extensive use of graphics processing hardware (GPUs).

Established feature correspondences between images are then used to estimate the relative camera orientations between pairs of images. Geometric verification of the relative camera orientations is performed using the five-point algorithm (Nistér, 2003) within a RANSAC loop (Fischler and Bolles, 1981). Once an

initial image pair is found, new images are incrementally added to the reconstruction using the three-point algorithm (Haralick et al., 1991). The relative orientations between cameras can be represented in a graph structure, the so-called epipolar connectivity graph. Images in the graph are represented by the nodes and the relationships between them (based on common feature points and overlap) are represented by the edges of the graph that correspond to the relative orientations between cameras.

Camera orientations and triangulated 3D feature points are then simultaneously refined by minimizing the reprojection error between the projected 3D point and its corresponding 2D feature measurement in the image in a bundle adjustment step (Triggs et al., 2000). Optimization in the bundle adjustment step is carried out based on Google's Ceres Solver for non-linear least squares problems (Agarwal et al., 2012).

2.2 Automatic Georeferencing

Reconstructions created by purely image-based approaches like the method described here are initially not metric due to the lack of scale information in the images. A metric scale of the reconstruction can be accomplished easily for example by one known distance in the scene. This might be a distance measure between two distinct points that is also easily recognizable in the digitally reconstructed 3D model, or a known distance between two camera positions.

However, in surveying applications in general, the absolute position of object points is important. In addition, we want the created 3D model stored and displayed in position and orientation in its specific geographic context. This can be achieved by a rigid similarity transformation (also 3D Helmert transformation (Watson, 2006) or 7-parameter transform) of the model into a desired metric target coordinate system using at least 3 known non-collinear point correspondences between model points and points in the reference coordinate system (control points). A more robust transformation result can be obtained by a larger number of points and a robust estimation of the transformation parameters for rotation, translation and scaling. The method of least squares within a RANSAC loop (Fischler and Bolles, 1981) improves clearly the registration quality of the model in the presence of noise and outliers.

2.2.1 Georegistration and GPS Alignment

Flying platforms for aerial data acquisition are often equipped with a GPS receiver, that allows positioning of the aircraft in flight, stabilization and, depending on the application autonomous navigation between waypoints. Recording of GPS data during the flight enables to track and monitor positions and travelled distances of the UAV. It is then necessary to link the recorded images to the corresponding position data and use GPS information for georeferencing. This can be achieved by synchronized timestamps of the images with the GPS signal. Several professional products instead offer a direct interface between on-board GPS receiver and camera to instantly assign a GPS location to a captured image and store the information in the meta data of the image file. Recorded information from inertial sensors may also be available in the meta data, providing additional information about the orientation of the aircraft at the time of capturing the image, given by the rotation angles for roll, pitch and yaw.

Position data stored for each image is now used to metrically scale the previously calculated reconstruction and to transform the model into a desired reference system. However, the quality and accuracy of location data is not sufficient in most cases to allow an accurate three-dimensional reconstruction and reliable

measurements in the scene solely based on GPS positions and IMU data. Due to weight restrictions and a maximum payload depending on the used aircraft, usually very small GPS receivers are used that allow only limited accuracy, that is typically in the range of 1-2 meters (Pfeifer et al., 2012). But, the accuracy is sufficiently high for a rough positioning and metric scaling of the image-based reconstruction because transformation parameters can be estimated robustly when using a large number of images. The more images, and consequently also one GPS position for each of the images, the more robust the transformation gets. The accuracy of the absolute positioning of the reconstruction might be low, but the precision of the metric scaling is high enough, because relative position errors between GPS positions should get better distributed and compensated, the larger the number of position measurements, i.e. the number of images is.

2.2.2 Constrained Bundle Adjustment with GPS and Ground Control Points

Pure image-based approaches suffer from systematic errors. We experienced that especially for a few datasets showing long elongated, large-scale scenes our pipeline resulted in large errors up to a few meters due to a deformation of the whole image block introduced in the bundle adjustment. Depending on the control point locations from the georegistration step, the errors drift away from those fixed points and cause this "bending"-effect shown in Figure 2 and 3. Observed camera block deformations are very often caused by incorrectly estimated radial distortion parameters of the camera. As a consequence the reprojections of 3D points onto the image plane are not correct and thus cause wrong error measures in the bundle adjustment process. Furthermore, the reprojection error as the sole evaluated error measure has impact on many independent parameters (3D positions of the object points as well as intrinsic and extrinsic camera parameters). Errors can be passed back and forth during the optimization and camera positions may undergo large changes.
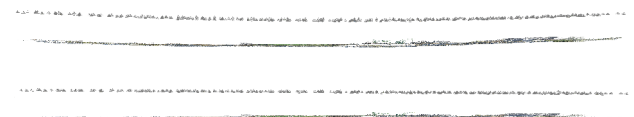

Figure 2: Illustration of the "bending"-effect (top). Camera positions and 3D points drift away from fixed control points due to systematic errors. The surveying area in this example has an extent of about 2.1×0.6 kilometers. Images were taken with a senseFly eBee fixed-wing drone at a constant flying height of 85 meters above ground level. Errors caused by the bending in this dataset resulted in positional shifts of 3D points and camera positions of up to 8 meters in altitude from their measured GPS position (bottom).

These systematic errors causing a deformation of the image block can be avoided by either a more accurate initial camera calibration or by adding external constraints in the bundle adjustment. For photogrammetric applications, we therefore use (roughly) known GPS positions of the cameras determined by an on-board GPS receiver and fixed control points to allow for camera self-calibration within the optimization.

Georegistration of the reconstruction as described in the previous subsection alone does not solve this issue. The model deformations are still present due to the shape-preserving character of the transformation. Instead, after rough georegistration and GPS-alignment, we use known GPS locations of the images in the bundle adjustment to constrain the positions of the cameras and to reduce an initial distortion of the image block. We do
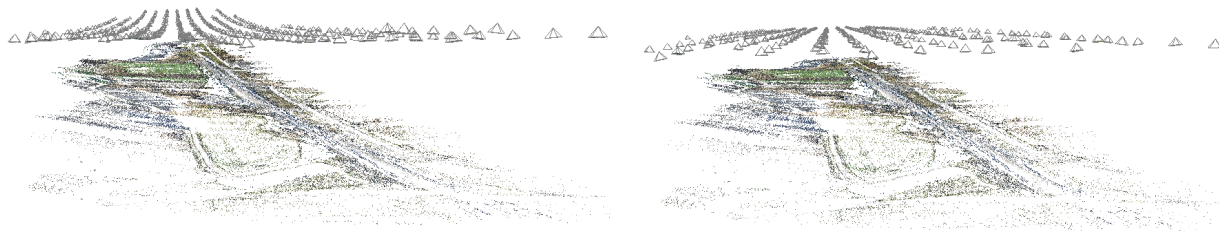
Figure 3: Results of the photogrammetric reconstruction without (left) and with additional GPS positions in the bundle adjustment (right). The direct comparison shows the reduction of the initially clearly visible distortion of image block and object points.

that by calculating the deviations of the calculated camera positions from the Structure-from-Motion result and penalize in the optimization step the deviation to their measured GPS positions. The influence of the deviation between the measured position is weighted by a Huber-error function (Huber, 1964). The camera positions can move only within a certain range around their measured positions and thus, are softly linked to their measured GPS positions. This leads to smaller residuals on the one hand, and on the other a direct transition from the model coordinate system into a desired geographic reference system can be accomplished simultaneously.

In addition, ground control points (GCPs) may also be used to correct distortions or a small geographic misalignment of the model and to tie the reconstruction to a certain geographic position. Besides camera positions and 3D points we therefore use the GCPs also for self-calibration in the bundle adjustment step and optimize common intrinsic camera and distortion parameters for each camera group.

GCPs signal points that are usually easily recognizable natural or artificial landmarks in the scene. Their position is known and for example determined by means of conventional survey methods or DGPS (Differential Global Positioning System) with high accuracy. For this purpose, the bundle adjustment is extended to the use of control points and their corresponding image measurements. The additional information can be seamlessly integrated into the reconstruction process. The reprojection error between the image measurements and projected control points is additionally weighted and penalized in the bundle adjustment in a similar way to the mass of natural features obtained by the SIFT keypoint detector. Important in this case is an appropriate weighting of the residual reprojection errors of the GCPs compared to the SIFT-generated points. Usually, low number of GCPs (around a couple of dozen) is confronted with a large number of natural feature points (hundreds of thousands or millions of points).

Integrating both mechanisms (using ground control points and GPS positions of the cameras) distributes the residual errors equally over all cameras and object points and allows for 3D reconstructions with very low geometric distortions. Furthermore, in the case of regular camera networks we experience that an additional cross flight and images taken at different distances to the object help to stabilize the intrinsic camera parameters. 2D image measurements, feature matches across overlapping images and triangulated 3D points are then better constrained. This leads to a more robust self-calibration result and furthermore to a more stable image block and increased point position accuracy even for very large, elongated surveying areas (Figure 2 and 3).

2.3 Surface Reconstruction and Texturing

The results of the previous steps so far are the external orientations of the cameras, optimized intrinsic camera parameters and a 3D point cloud from triangulated object feature points. Stereo (Hirschmueller, 2005) or multi-view methods (Irschara et al., 2012, Furukawa and Ponce, 2009) are common approaches to densify the initial point cloud and to increase the number of 3D points. For better visualization, we generate a closed surface model from the point cloud using a method described in (Labatut et al., 2007) based on 3D Delaunay triangulation and graph cuts. The method produces watertight triangle meshes from unstructured point clouds very robustly even in the presence of noise and gross outliers in the raw 3D sample points. The meshes can then be textured (Waechter et al., 2014) from the input images to generate a photorealistic representation of the scene.

3 DATA ACQUISITION

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

For our investigations we chose different test sites: One is located at the "Styrian Erzberg", another one is a small gravel pit situated in Upper Austria and we recorded an archaeological excavation in Turkey as well. The first two sites are equipped with a dense network of ground truth points to assess the quality of reconstruction by a point-wise comparison.

Point signalling is mainly done with binary coded, individually identifiable fiducial markers (Rumpler et al., 2014) printed on durable weather proof plastic foil. In addition non coded, red circular targets are used to densify the reference network in certain parts. Different subsets of the points are then used as ground control points (GCPs) for automated georeferencing, and others are used as check points (CPs) to evaluate the achieved accuracy. 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.

We used different platforms and cameras to acquire each of the datasets. One is a Falcon 8 octocopter by AscTec, equipped with a Sony NEX-5N digital system camera. The second flying platform is a senseFly eBee, a small fixed-wing UAV with a Canon IXUS 127HS compact camera. The main advantages of multicopters are its flexibility, the ability to fly at very low airspeed to record datasets with high overlap, hover and observe objects from any possible position, even very close to an object to capture images at a very high level of detail. The fixed-wing UAV, however, is able to fly and survey large areas in short time with certain details not been detected due to the in general larger flying altitude and higher airspeed. In addition we use a Canon EOS 5D full-frame digital SLR and a consumer-grade Panasonic compact camera for terrestrial images in areas, where highly detailed results or views from the inside of the object are required or an airborne mission cannot be performed. A compiled summary of camera specifications for detailed information on the cameras and sensors used is given in Table 1.

To guarantee a certain accuracy, a desired image overlap and minimum ground sampling distance has to be defined beforehand.

Based on Equation 1 and 2 for nadir image acquisition in aerial photogrammetry,

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

$$GSD = \frac{PixelSize\left[\frac{mm}{px}\right] * ElevationAboveGround\,[m]}{FocalLength\,[mm]} \quad , \qquad (2)$$

we estimate a maximum flying height above ground and imaging distance to the object, respectively. To enable analysis of which parameters influence the reconstruction accuracy we oversample the scenes and record images with about 90% overlap in previously defined distances and heights from the object.

Apart from the imaging distance, the baseline between particular cameras has a strong influence on the triangulation geometry and ray intersection. Especially for the canonical stereo configuration with parallel optical axes, the distance to baseline ratio is a good parameter to quantify the quality of a camera network. Small baselines lead to small triangulation angles and to high depth uncertainty. But to enable feature matching, high image overlap and intersection angles below $30°$ are optimal (Zeisl et al., 2009).

3.1 Styrian Erzberg

The Styrian Erzberg is the biggest iron ore open pit mine in central Europe. Our test site represents one quarry wall, which is about 24 m high and 100 m long with the typical geometry of an open pit hard rock mine. It is equipped with 129 reference points with known ground truth positions. 45 are realized as fiducial markers on the top and bottom of the wall and on the adjacent benches and are used as temporary GCPs. Additionally, the wall is equipped with 84 circular targets, which are used to evaluate the reconstruction accuracy. This dense network (see Figure 4) enables an extensive evaluation of accuracy and allows us to quantify systematic deformations of the image block and reconstructed 3D geometry.
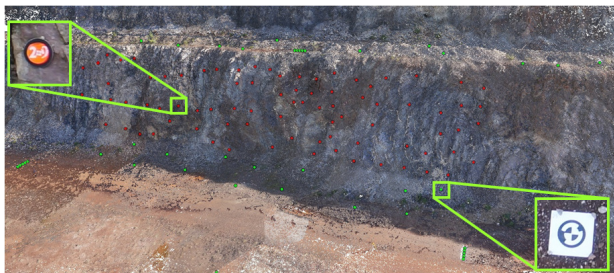

Figure 4: The reference point network allows an extensive accuracy evaluation. Markers (right) indicating GCP positions are shown in green, circular targets (left) for quantitative evaluation are in red.

Due to complex geometry and steep slopes at the test site, we used the AscTec Falcon 8 octocopter for image acquisition. Using the octocopter we were able to approach and hold any possible camera position, enabling the opportunity to acquire images under stable conditions for our further investigations. All together 850 images were recorded in different flying altitudes, view angles and distances to the object with a mean GSD of 1.5 cm.

3.2 Gravel Pit

Our second test site is a small gravel pit situated in Upper Austria. As shown in Figure 5 the scene includes the actual pit as well as the surroundings and covers an area of about $0.43\,km^2$. Reference points are temporarily signalled in the same manner as

described for the Erzberg dataset and are evenly distributed over the whole site. 27 control points are realized as fiducial markers and 19 as red circular targets. Additionally a small part of the pit was scanned at high level of detail (4 points per m^2) using the autonomous scan function of a Trimble S6 total station.

Figure 5: Colored model of a gravel pit with surroundings.

Images were recorded using a senseFly eBee fixed-wing UAV in different flying altitudes (75, 100 and 140 m). Due to camera specifications and higher elevation above ground the mean GSD is about 3.5 cm in this test scenario. The dataset consists of 533 images in total with an overlap within each altitude held constant at 70%. The resulting 3D model (Figure 5) includes more than 400 million points and represents the scene at a level of detail not achievable with manual surveying methods.

3.3 Turkey

Our last test site is an archeological excavation in Turkey. The site shows complex geometry with arches, partly collapsed chambers and walls. We used an AscTec Falcon 8 for aerial image acquisition, together with terrestrial images in areas which could not be observed from the air (Figure 6). The terrestrial images were

Figure 6: Image acquisition with an AscTec Falcon 8 octocopter for archaeological site documentation and reconstruction.

recorded from the inside and outside of the object with a Canon EOS 5D DSLR with a 24 mm wide angle lens for high resolution terrestrial images and a small consumer-grade Panasonic DMC-TZ22 zoom camera.

We took 5.014 images within four days in total with all three cameras, giving 38.4 GB of raw image data. Aerial images were captures in a classic raster flight pattern with cross flights in two different heights (40 and 90 meters above ground with minimum overlap of about 80%) and in a hemisphere flight around the object with tilted camera to ensure enough overlap with terrestrial images for automated matching. We were able to align 4.722 images fully automatic into one single reconstruction of the site. Seven markers as ground control points were used to georeference the model. An overview image of the reconstruction together with detail views of the object are presented in Figure 7.

4 EVALUATION AND RESULTS

In this section we analyze the performance of the presented workflow. For photogrammetric applications the accuracy of reconstructed object points is of prime interest. Thus we perform

Platform	Camera	Sensor type	Sensor dimensions	Resolution	Focal length	Pixel size
AscTec Falcon 8	Sony NEX-5N	APS-C CMOS	$23.4 \times 15.6\,mm$	4912×3264	$24\,mm$	$4.76\,\mu m$
senseFly eBee	Canon IXUS 127HS	1/2.3" CMOS	$6.16 \times 4.62\,mm$	4608×3456	$24\,mm$	$1.35\,\mu m$
Terrestrial	Canon EOS 5D	full-frame CMOS	$36.0 \times 24.0\,mm$	4368×2912	$24\,mm$	$8.24\,\mu m$
Terrestrial	Panasonic DMC-TZ22	1/2.3" CMOS	$6.2 \times 4.6\,mm$	4320×3240	$24\,mm$	$1.44\,\mu m$

Table 1: Camera and sensor specifications.

Figure 7: Rendered views from an automatically reconstructed and textured 3D model of an archaeological excavation site in Turkey, obtained from 4.722 terrestrial and aerial images captured with 3 different cameras from the air and from the ground.

a point-wise comparison of reconstructed object points to corresponding, clearly identifiable 3D reference point coordinates as already described in section 3. We investigate what are the relevant parameters determining accuracy in general and try to answer the following questions: How does accuracy increase with the use of external information in the reconstruction process given by ground control points and, how many control points are necessary to achieve satisfactory results with respect to absolute position accuracy and how should they be distributed.

Figure 8 shows the absolute point error for each check point of the Erzberg dataset, where a mean accuracy of less than 2.5 cm is reached using all 850 images and GCP constrained bundle adjustment. For the gravel pit dataset an overall accuracy of 14 cm can be achieved, primarily due to a higher flying altitude, a different camera with lower resolution (see Table 1) and different camera network.
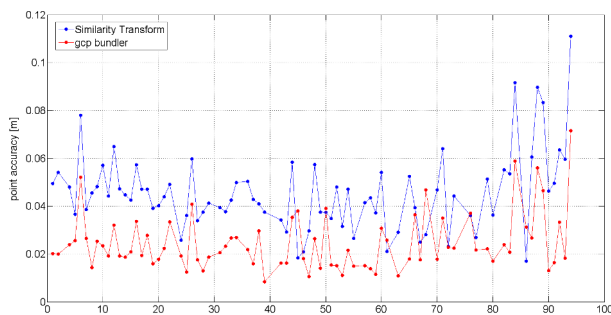

Figure 8: Using all 850 images and all available GCPs in the constrained bundle adjustment, a mean measurement uncertainty of 2.45 cm is reached (Rumpler et al., 2014).

For a better understanding of block stability and accuracy we investigate in the following relevant parameters influencing the reconstruction quality. For this purpose, a high oversampling of the scene was performed, as already described in section 3. Parameters with large impact on accuracy are, besides image overlap and triangulation angle, foremost the ground sampling distance determined by image resolution and imaging distance to the object and the distance to baseline ratio given by the camera network. In order to quantify the influence of these parameters and to give guidelines for image acquisition, a systematic parameter analysis is carried out based on different subsets of the previously described datasets.

4.1 Georegistration

One of the most important and critical steps with respect to the absolute position accuracy in the presented workflow is georegistration. Because of the fact that results of a Structure-from-Motion pipeline are initially in a local euclidian coordinate system, georegistration or at least scaling has to be done everytime, regardless of how images are recorded. As already mentioned, accurate georegistration is possible by integrating GCPs in the bundle adjustment. The number of points and their spatial distribution within the scene strongly affects the achievable accuracy. Figure 9 clearly shows that the error decreases with an increasing number of GCPs, but it is also apparent that even a small number of seven or eight GCPs is sufficient to get good results. In our three case studies adding more GCPs does not necessarily improve the result with respect to the overall accuracy.
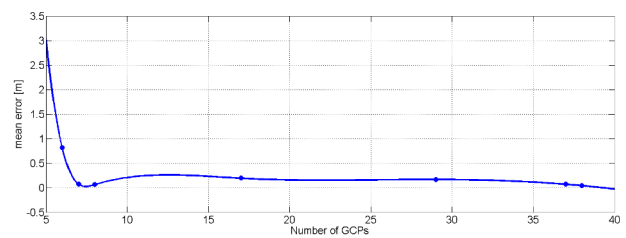

Figure 9: The best overall accuracy can be reached with 7 or 8 GCPs. A higher number is not necessarily needed for accuracy reasons.

Regarding the spatial distribution, the GCPs should be evenly distributed over the whole scene, especially concerning the height-component. Height tie points are at least as important as control points of position. If for example all GCPs are along one row systematic deformations can be observed, because the reconstruction can tilt around that axis. Moreover, in contrast to traditional bundle adjustment approaches (Kraus, 1994), control points should be situated not entirely at the boundaries of the scene, because of less image coverage and a weak triangulation network. To guarantee a desired accuracy the used ground control point should be robustly detected in at least 10 images.

Our investigations also show that georeferencing using GPS information of the aircraft exlusively without any additional position constraints is not sufficient for surveying tasks with respect to the absolute pose of the reconstruction. Indeed, integrating a large number of camera positions in the reconstruction process mitigates systematic deformations of the image block and might result in highly precise metric scaling, but it is not possible to achieve absolute position accuracies below the meter range due to the high uncertainty of the small on-board GPS sensors on UAVs.

4.2 Number of Observations

(Rumpler et al., 2011) shows in a synthetic simulation on a regular aerial flight pattern that accuracy increases with a higher number of image measurements and with increasing triangulation angles. Figure 10 derived from the Erzberg dataset including oblique views shows as well, that the mean object point error decreases with increasing total number of images used for the reconstruction. But it is also obvious, that there is a fast saturation in accuracy improvement within larger datasets.
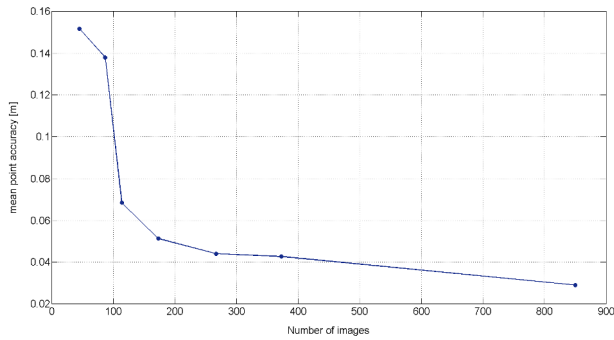

Figure 10: Error curve for different image subsets. With increasing total number of images used for the reconstruction, the mean point error decreases.

Thus, a higher number of images in the dataset leads to an accuracy improvement, but considering the number of image measurements per reference point does not necessarily reduce the error, as already shown in (Rumpler et al., 2014). In contradiction to synthetic results of (Rumpler et al., 2011), it is not possible to exemplify the achievable accuracy alone on the number of used images or observations for unordered and oblique datasets. The changing camera configuration influences feature matching, triangulation angle and ray intersection geometry, and from this we argue, opposing to (Fraser, 1996), that not every additional image measurement necessarily leads to an improvement in accuracy in practice with real world image data.

4.3 Camera Network and Resolution

We have shown that the influence of geometric configuration of the multi-view camera network on the resulting accuracy is higher than the influence of redundancy in image acquisition. In this section we present further investigations on the influence of camera network and resolution and compare a terrestrial dataset with different aerial networks for the Erzberg scene.

The ground sampling distance, or resolution respectively, has a strong influence on the achievable accuracy. The uncertainty of a point in 3D increases with increasing distance to the camera, thus images that are further away introduce larger errors. First, this is because of lower ground sampling distance, and thus, lower level of detail in the images. Secondly, the influence of localization errors on the reconstruction uncertainty increases with point depth. Image noise is approximately constant for all images, however, the resulting positional error increases with larger distances due to a larger angular error due to smaller triangulation angles (cp. Equation 3 with b being the baseline, f the focal length, d the disparity and z the the point depth.).

$$\epsilon_z \approx \frac{bf}{d} - \frac{bf}{d + \epsilon_d} \approx \frac{z^2}{bf}, \qquad (3)$$

Figure 11 shows the mean error for all targets of the Erzberg datset with respect to the different subsets. It clearly shows that the

viewing angle has to be carefully adapted to the object geometry. Using exclusively vertical images, the steep wall is shadowed and the mean error increases to 17.1 cm. The smallest error is achieved using a combination of different views (vertical, horizontal and oblique), which is only possible by using a multi-copter UAV. Because of the adjustable camera angle and low airspeed, images can be always optimally adapted with respect to the surface geometry and a high overlap and level of detail can be achieved easily.
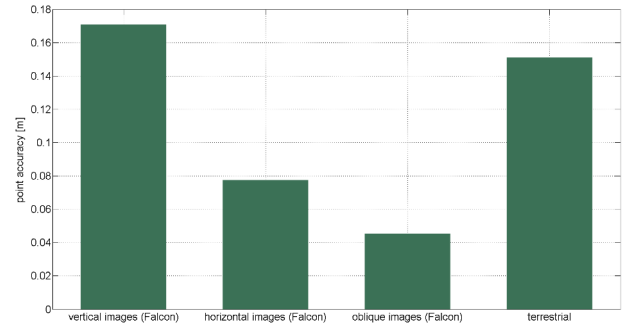

Figure 11: Adjustable camera angle, low airspeed and high image overlap using a multi-rotor UAV for image acquisition enables optimal results.

It is apparent that pure terrestrial photogrammetric systems are not that flexible compared to data acquisition with UAVs. Due to imaging positions bound to ground level it is mostly not possible to observe the object completely or from a certain distance or view point, due to geometric or safety reasons, especially in hazardous environments. The combination of different distances and image resolutions also affects the achievable accuracy positively. Images taken further away mitigate the error propagation within the first row, they help connecting the camera network over longer tracks and the image block is stabilized. In general, flying at different altitudes is a common approach in airborne photogrammetry, to optimize the intrinsic camera parameters, which furthermore also results in better reconstruction accuracy. We advocate the joint processing of views from different view points and especially to combine various platforms, aerial as well as terrestrial, which can considerably enhance the quality of the reconstruction and completeness of the model. We recommend the usage of multi-rotor UAVs for a flexible way of creating highly accurate reconstructions of complex geometries.

5 CONCLUSION

We presented an automated image-based reconstruction workflow to generate detailed and geo-accurate 3D models of complex scenes from unordered multi-view datasets, captured with different aerial imaging platforms and multiple terrestrial cameras. In two typical scenarios in open pit mining and one dataset from an archeological excavation we demonstrated that our system produces highly accurate and complete models of the scene, integrating aerial and terrestrial views into one single reconstruction. Low and equally distributed mean point position errors are achieved when integrating additional external constraints (ground control points and measured GPS positions of image locations) in the bundle adjustment to avoid systematic deformations and bending of the reconstruction due to an initially inaccurate camera calibration.

We showed that one of the most prominent parameters strongly impacting accuracy is, besides image overlap given by the camera network, foremost the ground sampling distance determined by image resolution and imaging distance to the object. Images

taken further away cause larger errors, but when using only images taken from a very close view point to the object, the reconstruction is more affected by drift and distortions. Combining images taken at different distances, view points and viewing angles stabilizes the image block and mitigates the error propagation. We suggest a combined use of different imaging platforms, especially for complex geometries and advocate the joint processing of aerial and terrestrial views to enhance the quality and completeness of the reconstruction.

Although many investigations and concepts discussed in this paper including bundle block adjustment approaches, camera self-calibration or optimal distribution of control points are well known in photogrammetric literature for decades, we presented a best practice example for different use cases, engineered to state-of-the-art performance.

We are able to directly and seamlessly integrate various cameras and view points into one single photogrammetric reconstruction process, as long as sufficient overlap between the different image subsets can be ensured. Especially the use of multi-copter UAVs enables the recording of highly overlapping image datasets due to the ability to navigate at very low airspeed, hover and capture images at nearly any possible position. These aircrafts close the gap between terrestrial close range and aerial photogrammetry and are therefore flexible enough and ideally suited to enable easy and safe data acquisition in complex or hazardous environments.

ACKNOWLEDGEMENTS

REFERENCES

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

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.

Fraser, C., 1996. Network Design. In: Atkinson, Close-range Photogrammetry and Machine Vision, Whittles Publishing UK, pp. 256–282.

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).

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.

Kraus, K., 1994. Photogrammetrie, Band 1 Grundlagen und Standardverfahren. Fifth edn, Ferd. Duemmlers Verlag, Bonn.

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.

Moser, P., Gaich, A., Grasedieck, A. and Zechmann, E., 2006. The SMX Blast Metrix - A new tool to determine the geometrical parameters of a blast based on 3D imaging. In: ISEE 2006, Texas, USA, pp. 80–84.

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.

Nistér, D. and Stewenius, H., 2006. Scalable Recognition with a Vocabulary Tree. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2161–2168.

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.

Rehak, M., Mabillard, R. and Skaloud, J., 2013. A mocro UAV with the capibility of direct Georeferencing. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-1/W2, pp. 317 – 323.

Rumpler, M., Daftry, S., Tscharf, A., Prettenthaler, R., Hoppe, C., Mayer, G. and Bischof, H., 2014. Automated End-to-End Workflow for Precise and Geo-accurate Reconstructions using Fiducial Markers. In: Photogrammetric Computer Vision - PCV 2014, pp. 135–142.

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.

Sivic, J. and Zisserman, A., 2003. Video google: A text retrieval approach to object matching in videos. In: IEEE International Conference on Computer Vision (ICCV), pp. 1470–1477.

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

Waechter, M., Moehrle, N. and Goesele, M., 2014. Let there be color! Large-scale texturing of 3D reconstructions. In: European Conference on Computer Vision (ECCV), pp. 836–850.

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

Zeisl, B., Georgel, P. F., Schweiger, F., Steinbach, E. and Navab, N., 2009. Estimation of Location Uncertainty for Scale Invariant Feature Points. In: British Machine Vision Conference (BMVC).