

3D reconstruction of buildings and landscapes – Evaluation of quality parameters and optimization of mission planning using RPAS

Stefan Schön, Patrick Ingwer, Arno Fischer, Martin Schafföner, Eberhard Hasche, Reiner Creutzburg
Brandenburg University of Applied Sciences, Department of Informatics and Media, P.O. Box 2132, D-14737 Brandenburg, Germany

Abstract

This paper provides an overview over different aspects of photogrammetry in combination with RPAS and mission planning. Video material captured with the help of a RPAS is used, to reconstruct 3D-models of buildings and landscapes. Overlap values are examined to find limits where the quality of the reconstruction can be preserved although the overlap is considerably smaller. The effects of different file formats for the reconstruction is investigated as well. Also georeferencing with the help of online map services is used to rectify the 3D-reconstruction.

Introduction

3D reconstruction is nowadays used in a lot of applications and for many different reasons, be it for archaeological excavations, marketing, or as asset for games. The essential part of data acquisition can be achieved with multiple approaches. In the simplest case, the object can be measured manually “on foot”. Also, options such as radar or electro-acoustic detection are applicable in different cases. The most common approach to acquire data from buildings or landscapes is to use LiDAR or Photogrammetry. Due to the technical development the formerly very complex process with media disruption, analogue-to-digital conversions, is today possible as a complete digital workflow. For data acquisition of large buildings or landscapes an aerial vehicle is also necessary to cover areas which cannot be reached from ground, or where a point of view which is further away from the object is simply advantageous. Since the cost of RPAS have fallen sharply with the advent of flying vehicles in the hobby sector and RPAS are used in a lot of applications [1], an adequate substitute for helicopters and airplanes is also available when it comes to aerial data collection. Even smaller RPAS can be used here, combined with a compact camera, to provide the data necessary for a photogrammetric approach. This can be a viable and affordable approach for close range photogrammetry and 3D-reconstruction. Therefore, this paper deals with the current state of the work process of photogrammetry in combination with a RPAS with the main focus of optimization potentials for data acquisition and the processing chain. On the basis of video footage, which is taken with the aid of a RPAS, the necessary quality parameters for a 3D-reconstruction of buildings of the University of Applied Sciences Brandenburg can be determined. A video is used in order to efficiently vary the amount of individual images which are generated. So instead of having to produce single image series for the quality assessment, where individual parameters might vary during different flights, numerous image series can be derived from a single recording. Recordings with highly redundant overlap are generated and examined. This redundancy is gradually reduced until a quality loss is visible in the 3D-reconstruction. A drop in

quality can be predicted as soon as the number of shared reference points significantly decreases. Simple georeferencing, using online map services, is applied to rectify the 3D-reconstruction and to allow positioning in the geo-coordinate system. Reconstruction effects of different image types and usage of a pre-calibrated camera are also considered. Finally, general rules for the optimal mission planning are derived from the overall result. In particular, how details and the overall quality can be retained despite reduced redundancy.

Evaluation of selected influences on photogrammetry

In addition to the standard procedure for reconstruction there are some additional possibilities which influence the result of the reconstruction. These could all be considered useful for a photogrammetric reconstruction. For this purpose several scenarios and their impact on the overall quality of the reconstruction are analyzed. Agisoft PhotoScan is used for all of the following applications. Unless stated, all calculations were performed with the parameters listed in table 1.

Table 1: Reconstruction settings in PhotoScan

Align Photos	
Accuracy:	high
Pair preselection:	disabled
Key point limit:	40000
Tie point limit:	1000
Constrain features by mask:	no
Build Dense Cloud	
Quality:	High
Depth filtering:	aggressive
Reuse depth maps:	no
Build Mesh	
Surface type:	Arbitrary
Source data:	Dense cloud
Face count:	Medium
Interpolation:	enabled
Point classes:	all
Build Texture	
Mapping mode:	generic
Texture from:	All cameras
Blending mode:	mosaic
Texture size:	4096
Texture count:	1
Color correction:	no

File format of the images

One factor, which should be considered in the reconstruction with PhotoScan is the file format of the images. Depending on the file format and the compression, two images may vary greatly in their size. This is particularly of interest when the project is quite large, the reconstruction is done on a single computer and a single chunk could already exceed the RAM limit. In addition the programs used to create image sequences from the video file need a varying amount of time, depending on the file format. For this reason, the differences and the impact of various file formats are analyzed below.

The first factor to be considered is the time factor. Photoshop CC 2015 is used to create image sequences from the video file. The same video sequence is exported as Jpeg, Tiff (uncompressed) and Tiff (LZW). To guarantee that the write limit of the hard disk is not reached during file export, a SSD, which is connected via PCIe, is used. The video was taken with a GoPro Hero 3 Black Edition with the settings: 3840x2160 px; 12.5 fps; wide. The results are listed in the following table.

Table 2: Computation time depending on the file format

	Jpeg (quality 12)	Tiff	Tiff (LZW)
Images: 91			
Computation Time:	~ 1 min	~ 0,5 min	~ 1 min
Images per second:	1,5	3	1,5
Images: 1307			
Computation Time:	12 min	3 min	12
Images per second:	1,8	7,2	1,81
Images: 2018			
Computation Time:	30 min	6 min	18 min
Images per second:	1,1	5,6	1,87
Images: 4040			
Computation Time:	84 min	12,5 min	35 min
Images per second:	0,8	5,4	1,9
Images: 9101			
Computation Time:	217 min	27 min	90 min
Images per second:	0,7	5,6	1,7

Regardless of the recorded scene, it can be determined that the export of the video file as uncompressed Tiff image is the fastest. Also, the Tiff images with LZW compression are exported faster than the Jpeg images. The CPU usage during all export processes is around 10 % to 15 % on a hexa-core CPU, which means that the export itself is not limited by the computing power. The declining processing speed in the Jpeg image series can be explained by the more complex picture scenes, as these require more computing time per image. The processing speed for Tiff images with LZW compression is relatively constant. Larger differences can be examined in the file size. The Jpeg images are at

average 6 to 7 MB, Tiff images with LZW around 11 MB and Tiff images without compression around 26 MB. With 7.2 images per second an average write speed of 191 MB per second is reached. Hard disks or slower SSDs could reach their maximum write speed, which means this factor is effectively limited to the storage drive and not CPU or RAM. Similar differences arise for the entire series. Thus, the sequence with 2018 images is around 14 GB, uncompressed Tiff images need 54 GB of storage space. Since the file size and the RAM limit the maximum size for a chunk and the render settings, using uncompressed Tiff images should be limited so smaller projects. To examine the reconstruction differences between the various file types, a new project is created for every file type. All images from one file type are processed in a single chunk. The flight mission is a circular flight around the auditorium on the campus of the University of Applied Sciences Brandenburg. Two circles were flown, one with a radius of 30 m and one with a radius of 50 m. Both times the flight height was 30 m. The video was recorded with the camera of the Phantom 3 Professional in the format: 4096x2160 px, 24fps. The reconstruction parameters were set to the values listed in table 5. The quality for the reconstruction of the dense point cloud was set to “ultra high”, so images would not be scaled down during processing. The following figures show the results of the reconstruction.



Figure 1. Reconstruction using Jpeg images (quality 12)



Figure 2. Reconstruction using Tiff images (uncompressed)

Visually, there are no differences between the reconstructions. Neither the dense point cloud, yet the mesh have differences. The only differences can be recognized at the edge of the bounding boxes, which are not important as the bounding box is larger than the area that should be reconstructed. The reconstructions based on the Tiff images have a higher noise in the point clouds, as more points were detected. In addition to the visual comparison, the exact reconstruction values can also be compared.

The data from the reconstruction values confirm the visual assessment. The reconstruction on the basis of the Tiff images has a negligible higher quality. Due to a larger number of points in the dense point cloud and therefore a higher amount of polygons, the computing times for these parts are slightly longer. Overall, the differences of the file format have a minimal impact on the reconstruction result. This can be explained by the very high effective overlap and the low reprojection error in all projects.

As a result, Jpeg images can be used if the conversion time from video to image sequence is not taken into account, or if the project has to be processed in a single chunk and could run into the RAM limit with other file formats. However, should the "centimeters per pixel"-resolution be higher, the usage of Jpeg images is an issue, due to compression and artifacts. Uncompressed Tiff images can be created very fast, but the amount of storage space required make them unsuitable for large projects. Accordingly, using Tiff images with LZW compression is the best starting point for most reconstructions.

Camera calibration

In classical photogrammetry, special pre-calibrated cameras are used, where their distortion coefficients of the camera lens and other parameters are known. This advantageous for an exact reconstruction. PhotoScan also offers the possibility to use pre-calibrated cameras. The camera data can be directly inserted for every chunk in a project. For this purpose, it is examined whether it is useful to calibrate a simple digital camera, to increase the quality of the reconstruction or to reduce the computing time.

To calibrate the camera lens a number of software products are available. The developer of PhotoScan, the company Agisoft, also provides a free application to calibrate the camera lens. The data calculated during this process can therefore immediately used in PhotoScan. Camera calibration is done by showing a checkered pattern on a computer monitor. This pattern has to be photographed with the camera. Based on these images Agisoft Lens calculates the focal length, the coordinates of the principal point and the radial distortion [2]. The radial distortion is based on Browns distortion model [3]. After calibration, the following values are obtained for the GoPro Hero 3 Black Edition used in this experiment.

Table 3: Camera calibration results for the GoPro Hero 3 Black Edition

Focal length	
f_x	1227,55
f_y	1230,02
Coordinates of the principal points	
c_x	2013,88
c_y	979,651
radial distortion coefficients	
K_1	0,169326
K_2	0,0101711
K_3	0,0434207
K_4	-0,0126214
P_1	0,00987558
P_2	-0,00691739

This data can be entered in PhotoScan under „Tools“ → „Camera Calibration“ for the currently active chunk. PhotoScan uses this data as initial values in the reconstruction.

In order to determine the differences between calibrated and uncalibrated camera a project with two chunks is created. The same pre-selected photos are assigned to both chunks. For one chunk the data acquired from the camera calibration is provide, in the other chunk the camera calibration data is left empty. The flight mission is a circular flight around the auditorium on the campus of the University of Applied Sciences Brandenburg with a radius of 60 m and a height of 30 m. The source for this reconstruction is a video which was taken with the GoPro Hero 3 Black Edition in the format: 3840x2160 px; 12.5 fps; wide. The reconstruction parameters were set to the values listed in table 1. The following figures show the results of the reconstruction.



Figure 3. Dense Point Cloud, calibrated camera



Figure 4. Dense Point Cloud, camera not calibrated

In both chunks all of the 45 images, each taken from a different angle, were aligned. Either the dense point cloud, as well as the textured mesh show minimal deviations between both chunks. Critical areas, such as the entrance which is made out of glass or the house edges show no differences. The overall quality can therefore be considered equal. The only difference can be found in the size and positioning of the bounding box, which is due to the fact that the project was calculated as a batch job. However, the building and the area around it are in the middle of the bounding box in both chunks. This fact can therefore be considered unimportant. In addition to the visual subjective differences, the exact reconstruction results can be compared.

Slight deviations can be seen between both reconstructions. Less tie points were found, although the camera was calibrated. This leads to a minimal higher reprojection error, which is, however, marginal. The effective overlap is slightly higher, which suggests a better alignment of the cameras in the chunk where the camera calibration data was provided. The computing time required for the photo alignment is practically identical in both reconstructions, which indicates that the camera calibration data, as a startup parameter for the bundle adjustment, is only of minor importance. Due to better overlap, more points were generated in the dense point cloud. This has therefore also an impact on the computation time which is needed for generating the dense point cloud. The number of polygons is higher as well, because these are directly tied to the amount of points in the dense point cloud. Since PhotoScan creates objects with far too many polygons this deviation can be neglected. Due to the higher number of polygons, the computation time for the UV-mapping is also slightly higher.

As a result can therefore be found, that the camera calibration is not relevant in PhotoScan in this scenario. However, if the same camera is used for several reconstructions, it may still be useful to perform a calibration, as the time required for camera calibration is quite low and the data could be beneficial in larger and more complex projects.

EXIF-data

It is advantageous for many of the calculations in photogrammetry, if certain camera data are known. If pictures are taken with a digital camera, various EXIF-data is added to the image which is created. This data can be read during image import in PhotoScan and can be used for the reconstruction process. However, by recording a video and the subsequent conversion into an image sequence, the images do not have any EXIF-data. Therefore, data like the type of the camera, the focal length, or the sensor size are undefined in the project. Since these values are needed for the calculation, PhotoScan uses default values if they are not provided. For the focal length, PhotoScan assumes a value of 50 mm [4]. Some cameras, like the GoPro Hero 3 Black Edition have a small focal length of 15 mm (35 mm equivalent), which is a large deviation from the default value. Because of this, the computation time and the quality of the reconstruction could be affected adversely. Since the GoPro Hero 3 can also take individual pictures, where EXIF-data is written, parts of this data could be transferred to the image sequence which was created from a video.

In order to read EXIF-data, the “ExifTool”-software is used. This is a console application, where parameters are specified during the launch. To display the exact names of the EXIF-data in an image, the application needs to be started with the following command (Windows):

```
Exiftool.exe -s <image>
```

The data is output in the “EXIF name : value” format. Because PhotoScan does not define which EXIF-data is used, all data that appears reasonable should be selected. The following EXIF-data is transferred from the single image to the image sequence for this test:

Table 4: EXIF-data used from the image

EXIF description:	Value
Make	GoPro
Model	Hero3-Black Edition
FNumber	2.8
ApertureValue	2.8
FocalLength	2.8 mm
FocalLengthIn35mmFormat	15 mm
FOV	100.4 deg
FocalLength35efl	2.8 mm (35 mm equivalent 15.0 mm)

In order to transfer the EXIF-data from the single image to the image sequence, ExifTool is used again. The following command can be used to transfer the data:

```
Exiftool.exe -tagsFromFile <Source> <Parameter> <Target>
```

The image with the EXIF-data is used as the source. The target can either be a single image, as well as a folder containing images. After the EXIF-data has been transferred to the image sequence, the images can be imported into PhotoScan. If the data has been transferred correctly, values such as focal length, camera manufacturer, or the type of camera will be displayed under "Tools" → "Camera Calibration".

In order to determine the differences between images with EXIF-data and images without, a new project with two chunks is created in PhotoScan. The mission for this test is a circular flight around the auditorium on the campus of the University of Applied Sciences Brandenburg at a height of 30 m and a radius of 60 m. Both chunks are assigned the same images. In one chunk images without EXIF-data are assigned. For the other chunk images, where the EXIF-data was transferred from a single image to the image sequence, were used. The source for the image sequence is a video taken with a GoPro Hero 3 Black Edition in the format: 3840x2160 px; 12.5 fps; wide. The reconstruction parameters are set to the values listed in Table 1. The following figures show the results of this reconstruction.



Figure 5. Reconstruction of the auditorium, textured mesh, without EXIF-data



Figure 6. Reconstruction of the auditorium, textured mesh, with EXIF-data

Out of 113 images, taken from a different angle, all 113 were aligned. In the dense point cloud, as well as the textured mesh, minimal deviations can be seen. Complicated areas, like the glass entrance or the house edges, have been reconstructed with the same quality in both reconstructions. The only difference can be found in the size and positioning of the bounding box, as in the camera calibration test. Since this is negligible, the overall visual quality of both reconstructions can be described as equal. In addition to the visual differences, the exact reconstruction results can be compared.

The reconstruction with EXIF-data shows a slightly lower quality, than the reconstruction without EXIF-data. More Tie points could be found, but the reprojection error is not lower. The effective overlap is minimal lower, which can be ignored as the difference is only 0.02 at 18.3 effective overlap. The computing time for the photo alignment is equal. However, a significantly lower amount of points in the dense point cloud were generated. Since the visual analysis showed no difference in the building and the surrounding landscape, the deviation in the number points is due to the fact that the bounding box differs in both reconstructions. Accordingly, the number of polygons is lower in the reconstruction where EXIF-data was not provided. Differences in the computation time for the UV-map are negligible, as this value is always only a few minutes, even for very large projects.

As a result, the EXIF-data does not seem to matter, if only a single camera is used. This means, that EXIF-data in images is not necessary for most projects and the usage of a video as source material does not impact the quality of the reconstruction or the computation time negatively. For larger projects, where different cameras are used, EXIF-data is still useful as it provides information to separate these cameras, e.g. by a different focal length.

Threshold values for overlap

The number of images used for a reconstruction has a direct impact on the required computation time. Therefore, it makes sense to limit the number of images. At the same time, the amount of images must be high enough so that the reconstruction can still be carried out successfully. For this reason, the following part will examine the thresholds of images which are needed for reconstruction in a specific scenario.

To determine the threshold values for overlap, a project with multiple chunks is created. For each chunk, a different number of images from the same flight is assigned. The mission is a circular flight around the auditorium on the campus of the University of Applied Sciences Brandenburg at a height of 30 m and a radius of 35 m. The source for the images is a video which is captured with

the camera of the Phantom 3 Professional in the format: 4096x2160 px, 24fps. The flying speed is set to 4.3 m/s. Accordingly, approximately 5.6 frames per meter are recorded. As a comparison model, a chunk where every 24th image was extracted is used. Therefore, 32 images are used for a new reconstruction. The reconstruction parameters are set to the values listed in table 1. For the reconstruction with 32 images the results can be seen in the subsequent pictures.



Figure 7. Reconstruction using 32 images (Dense Point Cloud)

As expected, the images were aligned correctly. The object was reconstructed completely, even large parts of the surrounding terrain were detected and reconstructed. The qualitatively worst part is located on a side wall, which is partially covered by trees. For this reason, this part is considered important for the following tests. Since the entrance of the building is made of glass, it has also been reconstructed poorly. Which is a general problem with photogrammetry and will therefore not be examined further in this test. However, the overall quality of the reconstruction is still satisfactory, so the number of images can be reduced further. For this purpose, every 48th image is extracted from the video, leaving only 16 images for the reconstruction. The results can be seen in the picture below.



Figure 8. Reconstruction using 16 images (Dense Point Cloud)

All images were aligned in this reconstruction as well. The overall quality is noticeable lower. The lawn area and the parking lots show large gaps and distortions in the mesh. The auditorium itself is still reconstructed with relatively good results. The important side wall has significant gaps in the dense point cloud, due to the lower number of images. This translates accordingly to the polygon model. The trees in front of the auditorium were no longer recognized. Overall, plenty of areas could be reconstructed, so the whole building could be restored in an external 3D modeling

program. Because of this, the number of images can be reduced further, to determine at which point the reconstruction quality is too bad to be restored with a 3D modeling program. For this purpose, every 60th image is extracted from the video, leaving only 13 images for the reconstruction. The results can be seen in the picture below.



Figure 9. Reconstruction using 13 images (Dense Point Cloud)

The photo alignment on this reconstruction shows significant shortcomings. The circular shape of the flight is hardly given. Accordingly, significant less areas could be reconstructed. This is also confirmed on the side wall of the auditorium. The dense point cloud is almost completely missing in this area. As more than half of the building is affected, this reconstruction can be considered unusable. The reconstruction with every 24th image can therefore be considered as the lower threshold, if further processing of the model shall not be necessary. However, the reconstruction with every 24th image still has some slight issues in the dense point cloud. To verify the maximum threshold, beyond which no qualitative improvement is noticeable, further tests can be done. For the new reconstruction, every 12th image is extracted from the video, which means that 65 images are now available.



Figure 10. Reconstruction using 65 images, overview



Figure 11. Reconstruction using 65 images (Dense Point Cloud)

In this reconstruction, the camera positions were detected very precisely. This is due to the significantly higher overlap of the individual images. Accordingly, the edge areas of the model, such as the parking lot and the Department of Computer Sciences and Media, were reconstructed far better. Also in the side wall of the auditorium, no missing areas are visible in the dense point cloud or the mesh. Compared to the reconstructions with less images, even the treetops were detected. The lateral roofs show a strong noise, independently from the number of images. This is partly due to the angle from which these building areas were captured. In previously well reconstructed areas, no improvement can be detected. Both the dense point cloud, and mesh are at the same level. This means, that areas where no parts are concealed and where the capture angle is good, the number of images can be reduced significantly. In addition to the visual differences, the exact reconstruction results can be compared.

The data from the various reconstructions show expected values. The number of images simultaneously increases the number of tie points. However, since the maximum number of tie points is limited, this value only rises by a small amount between the reconstruction with every 24th and 12th image. A slightly higher reprojection error in the reconstruction with every 12th image can be explained with the overall higher number of points. These points are mostly parts of trees and shrubs and are only detected because of the higher number of images. The matching time increases with every new image, as does the computation time, as expected. Because this value increases exponentially, the number of images should be kept low. The alignment time only needs marginal more computation time, even with hundreds of images. The points in the dense point cloud only rise by limited amount, due to the same reasons the number of points in the sparse point cloud is limited. Border areas and trees are reconstructed in greater detail the more images are used, the overall quality of these objects is still quite low and more often than not these objects are not needed at all. Since the number of polygons is directly linked to the number of points in the dense point cloud, the differences between the reconstructions are expected as well. The time needed for mesh generation and UV-mapping only rises with $O(\log n)$ up to a limit of several minutes and can therefore be ignored, especially for larger projects.

As a result, it is not useful to increase the number of images just to get a better reconstruction. Only for areas where important parts are covered, the number of images should be increased. If the object is captured from a good angle and the complexity is not too high, a good reconstruction can be achieved with a very low number of images.

Georeferencing

An important factor for a reconstruction is the transformation of the computed model into a fixed coordinate system. In this way, surfaces and areas can be equalized or realigned, thus rectified. The definition of coordinates in PhotoScan is possible in various ways. In the simplest case, GPS coordinates are directly embedded in the images. These coordinates can be read during photo alignment and reduce the computing time significantly. Alternatively, images can be georeferenced using ground control points (GCPs). For each GCP entry, longitude, latitude and altitude can be set. If there are no GPS coordinates, blueprints or other data available, internet map services can be used to georeference a model. Provided that the object can be detected on these maps. This method can easily be applied and is used in this case, because the video which is used as source for the image sequence has no GPS data. When choosing map material, several sources with differing qualities are available. The Google maps for the campus of the University of Applied Sciences was last updated around 2009. The maximum zoom level is quite low. Other map services, like Apple maps and Bing Maps, provide more recent map material in a higher quality.

Because the map material varies greatly between these services, an examination with regard to the implications for georeferencing in PhotoScan seems reasonable. For this purpose, the existing reconstruction from "Threshold values for overlap" with every 48th image is used. A number of GCPs is assigned to this reconstruction. They are placed directly onto the textured 3D-model. This way PhotoScan projects the GCP into every image where this point is visible. A manual adjustment of the GCP positions in the images might be necessary if the overall quality of the reconstruction is too low. After that, the model is duplicated for every map service accordingly. Due to the selected area, only points in the vicinity of the auditorium can be used. The following GCPs were selected for this test.



Figure 12. Selected Ground Control Points (GCPs)

This selection is not optimal, as the points are not distributed uniformly over the entire area. However, no other distinguishable points can be found on the maps. The geo-coordinates for Google Maps and Bing Maps are listed in the table below.

Table 5: Geo-coordinates from Bing Maps

Bing Point	Latitude	Longitude
------------	----------	-----------

1	52,410935	12,538924
2	52,411072	12,538887
3	52,411183	12,539120
4	52,411241	12,539297
5	52,410906	12,539777
6	52,410750	12,539676
7	52,410656	12,539466

Table 6: Geo-coordinates from Google Maps

Google Point	Latitude	Longitude
1	52,410934	12,538919
2	52,411069	12,538875
3	52,411188	12,539123
4	52,411242	12,539291
5	52,410907	12,539771
6	52,410755	12,539664
7	52,410669	12,539460

Although the positions of the GCPs were selected as accurately as possible, the values nevertheless differ. Point 7 is specified with 52.410656° in Bing Maps, but 52.410669° in Google Maps. This is a deviation of around 1.4 m. In addition to the latitude and longitude, the height is also an important factor. Because the height value is not listed in Google Maps or Bing Maps, the Google Elevation API can be used to get this data. A request can be send to the API, which will respond with JSON data containing the elevation. The structure of the HTTP request has the following format [5]:

<https://maps.googleapis.com/maps/api/elevation/json?&locations=lat1,long1|lat2,long2|...>

If the status in the JSON response is "OK" the data can be used in the project. To handle larger amounts of data a simple script is used. It reads GCPs from a file, forms an Elevation API request and adds the elevation data to the file. PhotoScan is able to import CSV-files and auto-assign values to GCPs, if the label is identical in the project and the file.

During import, the coordinate system has to be set to WGS84 (World Geodetic System 1984). Also, the columns of the table have to be in the right order. Google Maps and Bing Maps show data in the latitude/longitude format, whereas PhotoScan uses a longitude/latitude order in their tables. The accuracy for the data should be set to 1 m, as the data from the two map services already shows a high deviation. After the data has been imported, the estimated camera positions and the deviation of the GCPs can be retraced.

Several GCPs with Google Maps positions show a deviation of around 1 m. The reconstruction where Bing Maps data was used has a far smaller deviation of less than 0.4 m. This has to be expected, as the zoom level in Google Maps is lower and the exact positions of the GCPs can hardly be pinpointed. Since the positions of the images are now known, the images can be realigned using the "optimize camera alignment"-option. The generation of the dense point cloud, the mesh and the texturing have to be carried out again. This will rectify the reconstruction, in case distortions were present.

Because the reconstruction now includes real coordinates, it can be used to import the model in software like Google Earth.

This is useful, as the accuracy of the GCPs can be checked. If the reconstruction is quite large, an export as orthophoto can be an alternative. The project can be exported as “.kmz”-file, as this file format can be directly imported into Google Earth. Such an orthophoto can be seen in the picture below.



Figure 13. Orthophoto in Google Earth

The positioning of orthophotos is automatically done, based on the geo-data contained in the images. It is clearly visible that the image fits very well into the underlying map. All streets and buildings are precisely aligned. Even the areas with poor quality blend well into the underlying map. As the import and the positioning of the orthophotos was successful, the 3D-model can be exported.

Because Google Earth can only handle about 20,000 polygons, the mesh of the reconstructions has to be reduced by a fair amount. Normally, this should be done in an external 3D-program manually to preserve all details in the model. However, since remodeling of objects is not part of this paper the polygon reduction is done automatically in PhotoScan. The reconstructions are exported using the “Collada”-format, due to the fact that this file format can be imported directly in Google Earth. During export, the coordinate system has to be set to “local coordinates”, to preserve the geo-coordinates. To import the files correctly, the height has to be set to “absolute, 0 m” in the import window. This way, the model is placed directly onto the surface. The imported reconstructions can be seen in the pictures below.

Both reconstructions are easily imported and positioned. Although both models have slightly different GCPs, they were placed at almost identical positions. The reconstruction based on Bing Maps shows a better positioning. In contrast to the orthophoto, the roads are not aligned perfectly, which is due to the mesh reduction. Also, some areas of the lawn and the parking lot are below the Google Earth map. This is due to the fact that the GCPs were only placed directly around the auditorium. Both models have therefore still some weaknesses, particularly in relation to the height data. However, since the reconstruction of these outlying areas was not the goal and because they have a generally poor quality and would accordingly be discarded, this is not a disadvantage.

It can therefore be stated that georeferencing, based on Google Maps or Bing Maps, is quite possible and can lead to good results. Distortions in the model, which originate from the camera lens or the focal length can also be rectified. It also shows, that centimeter-accurate positioning is not possible due to the general

uncertainty of the maps. Moreover GCPs can only be set at points which are visible in the model and in the map service. This limits the range of possible points and can make georeferencing impossible, depending on the location which should be reconstructed and the relevance of the maps.



Figure 14. PhotoScan model edited for Google Earth

Lessons learned

As is shown in this work, RPAS represent an easy and quick way to take pictures which can be used to generate point clouds and 3D-models. In particular, the fact that RPAS can fly closer to the target than ordinary manned aviation systems represents a significant advantage. However, to obtain a high quality reconstruction an exact mission planning before the actual flight is required. Safety and security are also an important factor, especially due to the significantly increasing amount of RPAS in the civil sector. In addition to the safety of the hardware, the reliability of the flight missions must also be ensured. The latter can be done by an increased automation and checklists, as is known from aviation. The security of RPAS systems in the civil sector still continues to pose a problem [6]. The selection of the system is an important factor as well. Cameras with a high field of view, like the GoPro, reduce the number of images that are required for a successful reconstruction. But the reconstruction is also more vulnerable to distortions, especially if the pictures are taken from a poor angle.

The usage of videos as source for the image sequences, provides some advantages compared to the capture of images with a fixed overlap. Since the number of images can be greatly reduced, if the object which should be captured is not covered, the required computing time can be greatly reduced. This fact can be utilized especially well when videos are used, as the overlap can be adjusted depending on the area. An overlap of 50 to 60 % can be enough to reconstruct an area with high quality. Even the vertical overlap can be completely left out, at least for buildings which are smaller than 20 m. Areas which are obscured or of high complexity need at least 80 % overlap to be reconstructed with good quality. However, depending on the area which is covered, even an overlap of 90 % will not provide good results. In this case the trajectory of the RPAS has to be varied, especially the angle from which this area is captured.

The lack of camera data is not relevant, at least in the scenario examined in this paper. It is beneficial if different cameras are used, as this will provide the necessary information to differentiate the cameras. Because the addition of camera data can be easily achieved with existing software, the usage of camera data should still be considered. Especially for large projects, where the photo

alignment takes a considerable amount of time, existing camera data can provide an advantage. Also, because the transfer or creation of camera data can be combined with the image pre-selection which has to be done anyway.

The effects of camera calibration are, in the case of the GoPro Hero 3 Black Edition and in the tested scenario, extremely low and therefore not necessary. If the same camera is used for several projects it should be considered, as the calibration has to be done only once and it might provide some improvements in other scenarios.

Simple georeferencing, using online map services, is possible and can provide decent data for the rectification of the reconstructions. However, the accuracy of the positioning is limited and the georeferencing can be impossible, if no distinctive points can be found or if the map data is heavily deprecated. If georeferencing can be done, the accuracy is sufficient for Google Earth or similar programs. For archaeological excavations or other scenarios which require a higher accuracy, the use of other methods, such as differential GPS, is still necessary.

References

- [1] “Applications of multimedia technology on autonomous flying robots for university technology transfer projects” 2013. Available: https://www.researchgate.net/publication/258332971_Applications_of_multimedia_technology_on_autonomous_flying_robots_for_university_technology_transfer_projects
- [2] “Agisoft Lens User Manual” 2011. Available: <http://downloads.agisoft.ru/lens/doc/en/lens.pdf>
- [3] „Photogrammetric Engineering, Decentering Distortion of Lenses,“ Mai 1966. Available: https://eserv.asprs.org/PERS/1966journal/may/1966_may_444-462.pdf.
- [4] Agisoft, „Lack of EXIF data,“ in *Agisoft PhotoScan User Manual*, 2015, pp. 6-7.
- [5] “The Google Maps Elevation API” 2015. Available: <https://developers.google.com/maps/documentation/elevation/intro>
- [6] “Hacking and securing the AR.Drone 2.0 quadcopter - Investigations for improving the security of a toy” 2014. Available: https://www.researchgate.net/publication/260420467_Hacking_and_securing_the_ARDrone_20_quadcopter_-_Investigations_for_improving_the_security_of_a_toy