

## ON-LINE COMPATIBLE ORIENTATION OF A MICRO-UAV BASED ON IMAGE TRIPLETS

M. Reich\*, J. Unger, F. Rottensteiner, and C. Heipke

Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany  
{reich, unger, rottensteiner, heipke}@ipi.uni-hannover.de

Commission III WG III/3

**KEY WORDS:** orientation, UAV, trifocal tensor, incremental bundle adjustment

### ABSTRACT:

In this paper we present a robust orientation approach for an imaging sensor flown on a micro-UAV based on image triplets. Our aim is to have the orientation available online, i.e. during image acquisition. The resulting point cloud and sensor orientations can then for instance be evaluated for navigation purposes of the UAV or to analyse the completeness of the point cloud. We use low quality imagery extracted from the downlink of an onboard PAL-camera. Trilinear constraints and cross-checked matches allow for a high robustness of the sensor orientation and the sparse 3D point cloud. In order to reach the goal of on-line processing given the large number of observations and unknowns, we make use of an incremental bundle adjustment. Estimated parameters are incrementally improved without explicitly considering previous observations.

Our approach combines linear projective geometry for obtaining initial values using the trifocal tensor with non-linear perspective geometry for the estimation of the unknowns. This combination allows for a high precision of the estimation, while eliminating the need for initial values. We evaluate the performance of our approach by means of imagery we acquired of the facade of the Welfenschloss in Hannover, collected with a Microdrones md4-200 micro-UAV. The results are the orientation parameters of the images and a sparse 3D point cloud representing the object. They are compared to those from a commercial bundle adjustment software and analysed in terms of geometric precision.

### 1 INTRODUCTION

Unmanned Aerial Vehicles (UAV) provide a flexible instrument for many tasks in photogrammetry. Especially vertical take-off and landing (VTOL) devices are able to capture images of complex three dimensional objects such as buildings. The use of the resulting 3D-models and their level of detail are diverse. Generalised models are interesting for planning and visualisation, whereas for cultural heritage and architectural purposes a more detailed model is usually required. Sometimes, the structure and the complexity of the object of interest are not known in advance. As an example, a building may be composed of courtyards or terraces, which may be unknown prior to data acquisition. If in such cases the initial flight path is not adequately refined on-the-fly once the additional detail becomes apparent during data acquisition, the result will be incomplete. In order not to miss relevant information about the object, an on-line adaptation of the flight path is required.

In this paper we present an approach that offers the possibility to evaluate the collected data on-line in terms of its usability for 3D modelling following the idea presented in (Hoppe et al., 2012). Based on low-resolution imagery which is transmitted from an on-board PAL-camera on the UAV, the sensor orientation, i.e. the camera position in space and its viewing direction, and a sparse point cloud are estimated by incremental bundle adjustment. Overlapping image triplets are used for robust keypoint matching and for determining initial values for the unknowns. The initial values of the unknown orientations for the first image triplet are obtained from the trifocal tensor, determined from triples of homologous points. Subsequent images are then appended to the oriented images via spatial resection.

The remainder of the paper is structured as follows. After summarising related work in section 2, we describe the hardware

and the way images are acquired in section 3. In section 4, our methodology is presented. Based on the implementation of the presented methods, we show results for a UAV-based dataset in section 5. Section 6 concludes this paper and gives a short outlook of our future work.

### 2 RELATED WORK

A comprehensive overview of projective geometry and the trifocal tensor is given in (Hartley and Zisserman, 2000). In (Ressl, 2000), the trifocal tensor is analysed more specifically in terms of its usability in photogrammetry. Strategies as well as fundamental theorems of bundle adjustment are extensively summarised in (Triggs et al., 2000). A detailed derivation of the incremental bundle adjustment is given in (Beder and Steffen, 2008). Another example of incremental orientation of monocular image sequences is given in (Meidow, 2012). The focus in that work lies on loop closure which has a significant positive effect on the estimation of the model parameters. The functional model in (Meidow, 2012) is based on homographies instead of collinearity equations and aims at stitching aerial images.

The idea of using image triplets in image sequences can be found in (Nistér, 2000). In contrast to our approach a hierarchy of triplet combinations represented by trifocal tensors over the whole sequence of images is used. Triplets for the orientation of large image datasets are also employed in (Bartelsen et al., 2012). However, instead of using the trifocal tensor, the relative orientation is represented by three combinations of image pairs. Also, the images of these datasets are unordered and come with a long baseline, which leads to a more complex orientation problem. Other implementations of on-line orientation, e.g. (Klein and Murray, 2009), and on-line dense reconstruction, e.g. (Wendel et al., 2012), are based on projective geometry alone and work with a combination of a local bundle adjustment with a global one. The local

\*Corresponding author

bundle adjustment is basically required to keep track of the orientation parameters between key frames for which the global adjustment is carried out to obtain a more stable overall solution. No incremental adjustment in the sense of (Beder and Steffen, 2008) is carried out. Incremental bundle adjustment taking into account points at infinity is carried out in (Schneider et al., 2013). They use it for the orientation of UAV-based fisheye cameras, using the software presented in (Kaess et al., 2008) for optimisation.

We combine linear projective geometry for obtaining initial values with a functional model based on the collinearity equations for incremental bundle adjustment. In doing so, we avoid the effects of over-parametrisation, and we are able to integrate information about the interior orientation of the camera in a straightforward way. We use image triplets in order to improve the robustness of matching by only accepting pairwise keypoint matches that are consistent within all pairs that can be formed from the images of a triplet.

### 3 HARDWARE AND DATA HANDLING

As a platform we use a *Microdrones md4-200* micro-UAV<sup>1</sup>, a VTOL-quadrocopter with a maximum payload of 300 grams. The potential flight duration is up to 25 minutes with one battery depending on the take-off-weight. The sensor we use for our investigations is a PAL camera with a resolution of 720 by 576 pixels. The UAV can transmit the analogue video signal on-line to a ground station unit in interlaced mode. In a frame grabber in the ground station the analogue signal is converted to digital images. The quality of the transmitted imagery is highly affected by the motion of the UAV and by disturbances during data transmission. The platform movement manifests itself in blurred images, which are automatically excluded from further computations based on a certain minimum number of image matches (see below). Images with disturbances caused by transmission errors are not yet excluded automatically.

### 4 METHOD

In this section we describe the mathematical approach to determine the orientation parameters and 3D structure information from the images. We assume the interior orientation parameters of the camera, including parameters related to lens distortion, to be known. The images are processed in the order in which they are acquired by the sensor. For the sake of robustness we use image triplets (Bartelsen et al., 2012), (Nistér, 2000) for matching and orientation.

We start the computation as soon as three images are available. SIFT features (Lowe, 2004) are extracted from each of these images, and consistent three-way correspondences between features from the three images are determined. For the first image triplet, we compute the trifocal tensor to obtain initial values for relative orientation (Hartley and Zisserman, 2000), (Ressl, 2000), which we use as an input for non-linear bundle adjustment to obtain the optimal estimate of the orientation parameters and the object coordinates of the tie points. The procedure applied for the orientation of the first image triplet is explained in section 4.1.

Starting from the fourth image in the sequence, we use a different procedure, because at this stage, object coordinates of tie points are already known from the images oriented previously. Whenever a new image is received, we extract SIFT features and form a triple consisting of the new image and the two images that were

most recently oriented. Three-way correspondences are again used to obtain consistent feature matches over three views. Features in the new image assigned to features for which object coordinates were already determined in the course of orienting the previous images in the sequence are used as quasi-control points to compute initial values of the exterior orientation parameters of the new image by spatial resection. After that, initial values for the object coordinates of tie points only available in the current image triple are determined by spatial intersection. Finally, an incremental bundle adjustment is carried out to improve the orientation parameters and the object coordinates. This procedure, including incremental bundle adjustment, is described in detail in section 4.2.

#### 4.1 Image Triplets and Trifocal Tensor

The first image pair is used to define the object coordinate system of the entire block. The origin of the object coordinate system is located in the projection centre of the first image. Its  $X$  and  $Y$  axes are parallel to the axes of the image coordinate system, whereas the  $Z$  axis is defined to coincide with the negative viewing direction. The scale of the coordinate system is fixed by the base length between the first and the second images, which is set equal to one.

Matching image triplets allows for an effective cross-check of the matching points. We first carry out a pair-wise matching of keypoints based on SIFT-descriptors and then check for consistency of the pair-wise matches over three views. If point  $p^I$  in image  $I_I$  matches point  $p^{II}$  in the second and point  $p^{III}$  in the third image,  $p^{II}$  and  $p^{III}$  must also be a pair-wise match for the triple  $[p^I, p^{II}, p^{III}]$  to fulfil the cross-check constraint. Especially for repetitive structures, which often arise in facade modeling, this constraint reduces the number of wrong matches. Then, we estimate the trifocal tensor based on three-point correspondences  $[p^I, p^{II}, p^{III}]$ , using the following constraints (Hartley and Zisserman, 2000, Chapter 15):

$$p^{I,i} \left( p^{II,j} \epsilon_{jpr} \right) \left( p^{III,k} \epsilon_{kqs} \right) T_i^{pq} = 0_{rs}. \quad (1)$$

In Eq. 1, the indices  $i, j$  and  $k$  refer to the homologous points triple in homogeneous coordinates, thus  $i, j, k = 1, \dots, 3$ , whereas  $p, q = 1, \dots, 3$  are the indices of the elements of the three matrices  $T_i$  of which the trifocal tensor is composed.  $\epsilon_{jpr}$  is a tensor whose entries are either 0, +1 or -1 and which is closely connected to the cross-product of two vectors; cf. (Hartley and Zisserman, 2000, Appendix 1.1) for details. The values for  $r$  and  $s$  are restricted to be 1 or 2, because for each three-point correspondence only four of the arising equations are linearly independent. The trifocal tensor can be uniquely determined based on the constraints in Eq. 1 if seven three-point correspondences are known. In order to cope with outliers in the three-point correspondences, we apply RANSAC (Fischler and Bolles, 1981) for estimating the trifocal tensor. For each RANSAC iteration, we randomly select seven point correspondences and use them to estimate the trifocal tensor. For the validation of each tensor estimate we perform a point transfer to the third image using the fundamental matrix of the first two images  $F_{21}$  retrieved from the tensor (Hartley and Zisserman, 2000, Chapter 15):

$$\hat{p}^{III,k} = p^{I,i} g_j^{II} T_i^{jk} \quad (2)$$

with  $g^{II} \perp g_e^{II} = F_{21} p^{II}$ .

In Eq. 2,  $g_j^{II}$  denotes the  $j^{th}$  element of line  $g^{II}$  passing through  $p^{II}$  and being perpendicular to the epipolar line  $g_e^{II}$ , and  $\hat{p}^{III}$  is

<sup>1</sup><http://www.microdrones.com/>

the point corresponding to  $p^I$  and  $p^{II}$  in the third view. Whether or not a triple  $[p^I, p^{II}, p^{III}]$  is an inlier given the current estimate for the trifocal tensor is decided on the basis of the Euclidean distance between the estimated and the measured points,  $\hat{p}^{III}$  and  $p^{III}$ . This distance must be below a certain threshold for the triple to be accepted as an inlier.

Once a valid trifocal tensor is found, we compute initial values for a relative orientation of the first image triplet on the basis of the fundamental matrices derived from the trifocal tensor (Hartley and Zisserman, 2000) and the camera calibration data. Then, we perform a robust bundle adjustment based on the nonlinear collinearity equations formulated in a Gauss-Markov model. Observations are weighted depending on their residuals, allowing outliers to be detected and excluded from the estimation (Kraus, 1997). The results of the initial bundle adjustment form the basis on which subsequent images are oriented in the incremental bundle adjustment.

#### 4.2 Incremental Bundle Adjustment

The trifocal tensor is only applied to the first image triplet to initialize the block. Subsequent images are appended to the existing block via resection and incremental bundle adjustment. In the remainder of this section we will refer to images whose orientation parameters have already been estimated as *oriented images*, and the tie points determined in this process will be called *existing points*, as opposed to the *new image* for which no orientation parameters are known and to *new points* for which we do not yet know object coordinates. Again, we form an image triplet, using the new image and the two oriented images added to the block most recently. We extract SIFT features from the new image. Then we search for consistent point triples  $[p^I, p^{II}, p^{III}]$  for which the correspondence  $[p^I, p^{II}]$  was already validated in a previous stage (i.e., part of a three-way correspondence using the previous image triple) and, thus, corresponds to an existing point for which object coordinates are already known. The existing points can be used as quasi-ground control for spatial resection in order to obtain initial values for the orientation parameters of the new image. Again, we apply RANSAC, this time based on the four-point algorithm for spatial resection described in (Kraus, 1997), which does not require any initial values.

Having obtained initial values for the orientation parameters of the new image in the way just described, we search for *new points* to stabilise the estimation of the orientation parameters of the new image. In order to do so, we again perform three-way matching in the way described in section 4.1, but discarding any SIFT features in any of the three images that already has been validated to correspond to an existing point. In this case, because initial values for the orientation parameters are known for all images of the triple, the matches are not verified on the basis of the trifocal tensor, but by pairwise reprojection. First, initial values for the object coordinates of each new point are estimated for each pair of images  $[I, II]$ ,  $[I, III]$  and  $[II, III]$  via spatial intersection. In a second step the mean of these three estimates is projected into each image. The match is classified as an inlier if the Euclidean distance between the reprojection of the mean object points and the observed points is below a user-defined threshold in all three images.

If the number of triple correspondences is lower than a user-defined threshold, the new image is rejected and the orientation procedure moves on to the next image in the sequence. In this way images that are affected by extreme movement of the UAV or are deficient in image quality, are excluded from the processing chain automatically.

If a sufficient number of correspondences is found, we carry out an incremental bundle adjustment. For that purpose, the vector of unknowns is split into two components. The first component is a vector  $x_1$  containing the unknown exterior orientation parameters of the images that have already been included into bundle adjustment and the object coordinates of the existing points, whereas the second component is a vector  $x_2$  that contains the orientation parameters of the new image and the object coordinates of the new points, the latter ones expanding the point cloud determined by bundle adjustment. For the first group of unknowns we do have the results of the previous bundle adjustment which can be used as initial values, whereas for the second group of unknowns, the initial values have been derived in the way just described. Similarly, we split the vector of observations into a component  $l_1$  containing all the observations that were already used in previous adjustments and a component  $l_2$  that contains the new observations that were derived in the matching procedure for the current image triple. There are three different groups of new observations in  $l_2$ :

1. image points observed in the oriented images that correspond to new object points (in  $x_2$ ) but are also related to the orientation parameters in  $x_1$
2. image points observed in the new image that correspond to existing object points (in  $x_1$ ), but are related to the orientation parameters of the new image in  $x_2$
3. image points observed in the new image that correspond to new object points (in  $x_2$ ) that are also related to the orientation parameters of the new image in  $x_2$ .

After linearisation, we formulate an extended Gauss-Markov model:

$$\begin{pmatrix} l_1 + v_1 \\ l_2 + v_2 \end{pmatrix} = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \quad (3)$$

The first line of Eq. 3 represents the standard Gauss-Markov model for an adjustment solely based on  $l_1$  for determining  $x_1$ . As pointed out above, the new observations  $l_2$  depend on the unknown parameters both in  $x_1$  and in  $x_2$ . The relation between  $l_2$  and the parameters  $x_1$  and  $x_2$  is described by the second line in Eq. 3. The design matrix  $A_{21}$  contains the derivatives of the collinearity equations with respect to the parameters  $x_1$ , whereas  $A_{22}$  contains the derivatives of the collinearity equations with respect to the new parameters  $x_2$ . It is the goal of incremental bundle adjustment to solve Eq. 3 for the unknown parameters  $x_2$  using the inverted normal equation matrix of the previous iteration. Following (Beder and Steffen, 2008), we want to use the new observations to improve the old parameters  $x_1$  and update them to a new parameter vector  $\hat{x}_{1,+}$ . The estimation of both groups of parameters reads as:

$$\begin{pmatrix} \hat{x}_{1,+} \\ \hat{x}_2 \end{pmatrix} = N_+^{-1} \begin{pmatrix} A_{11}^T P_{11} l_1 + A_{21}^T P_{22} l_2 \\ A_{22}^T P_{22} l_2 \end{pmatrix}, \quad (4)$$

where  $P_{11}$  and  $P_{22}$  are the weight matrices derived from the inverse covariance matrices of  $l_1$  and  $l_2$ , respectively. The normal equation matrix  $N_+$  for the incremental bundle adjustment can be decomposed into three submatrices  $B$ ,  $C$  and  $D$ :

$$N_+ = \begin{pmatrix} B & C \\ C^T & D \end{pmatrix} \quad (5)$$

with

$$B = A_{11}^T P_{11} A_{11} + A_{21}^T P_{22} A_{21} \quad (6)$$

$$C = A_{21}^T P_{22} A_{22} \quad (7)$$

$$D = A_{22}^T P_{22} A_{22}. \quad (8)$$

The inversion of  $N_+$  can be performed as

$$N_+^{-1} = \begin{pmatrix} B & C \\ C^T & D \end{pmatrix}^{-1} = \begin{pmatrix} K & L \\ L^T & M \end{pmatrix} \quad (9)$$

with

$$M = \left( D - C^T B^{-1} C \right)^{-1} \quad (10)$$

$$L^T = -M C^T B^{-1} \quad (11)$$

$$K = \left( I + B^{-1} C M C^T \right) B^{-1}. \quad (12)$$

As one part of  $B$  is the normal equation matrix  $N$  of the previous bundle adjustment (Eq. 6), which has been inverted before the new image was added, the inversion of  $B$  can be computed as

$$B^{-1} = F N^{-1} \quad (13)$$

$$F = I - N^{-1} A_{21}^T G A_{21} \quad (14)$$

$$G = \left( P_{22}^{-1} + A_{21} N^{-1} A_{21}^T \right)^{-1}. \quad (15)$$

Substituting Eq. 9 into Eq. 4 and applying the matrix-vector products yields:

$$\hat{x}_{1,+} = K \left( A_{11}^T P_{11} l_1 + A_{21}^T P_{22} l_2 \right) + L A_{22}^T P_{22} l_2 \quad (16)$$

$$\hat{x}_2 = L^T \left( A_{11}^T P_{11} l_1 + A_{21}^T P_{22} l_2 \right) + M A_{22}^T P_{22} l_2. \quad (17)$$

Substituting  $K$  from Eq. 12 and  $L^T$  from Eq. 11 into Eqs. 16 and 17 and considering the fact that the parameters estimated in the previous adjustment are

$$\hat{x}_1 = N^{-1} A_{11}^T P_{11} l_1 \quad (18)$$

results in an update equation for the old parameters ( $\hat{x}_{1,+}$ ) and an estimate for the new parameters ( $\hat{x}_2$ ) that no longer depends on the old observation vector  $l_1$ , but only on the previously estimated parameters  $\hat{x}_1$  and the new observations  $l_2$ :

$$\hat{x}_{1,+} = \left( I + B^{-1} C M C^T \right) F \left( \hat{x}_1 + N^{-1} A_{21}^T P_{22} l_2 \right) \quad (19)$$

$$- B^{-1} C M A_{22}^T P_{22} l_2$$

$$\hat{x}_2 = - M C^T F \left( \hat{x}_1 - N^{-1} A_{21}^T P_{22} l_2 \right) + M A_{22}^T P_{22} l_2. \quad (20)$$

The estimation of the parameters (Eqs. 5-20) is iterated until convergence. To assure robustness of the adjustment the observations are weighted as described in section 4.1. The main advantage of incremental bundle adjustment is a reduction of computation time. There is no need to invert the normal equation matrix  $N_+$  in Eq. 4, which becomes larger whenever a new image is added to the block. Instead, only the matrices  $M$  and  $G$  have to be inverted (Eqs. 10 and 15), whose size is equal to the number of the new unknowns  $x_2$  ( $u_2 \times u_2$ ) and the number of new observations  $l_2$  ( $n_2 \times n_2$ ), respectively. Assuming the number of new points added with each new image and, consequently,  $u_2$  and  $n_2$  to be approximately constant, so are the sizes of  $M$  and  $G$ . Furthermore, neither the design matrix  $A_{11}$  nor the observations  $l_1$

and the corresponding weight matrices  $P_{11}$  of the previous iteration are needed for the estimation. Once the incremental bundle adjustment is terminated for the current image triplet, the inverse  $N_+^{-1}$  becomes  $N^{-1}$  when the next image is added to the sequence, whereas  $\hat{x}_{1,+}$  and  $\hat{x}_2$  are combined to parameter vector  $\hat{x}_1$ .

Although the size of the matrices  $M$  and  $G$ , which have to be inverted, remains constant,  $N^{-1}$  and  $\hat{x}_1$  become larger with every newly added image, which still slows down the adjustment. After a certain number of oriented image triplets, older ones are excluded from the incremental adjustment in favour of new incoming images.

## 5 RESULTS

In this section we present the results of our algorithm obtained for the orientation of an image sequence showing the facade of the Welfenschloss in Hannover. As described in section 3, the quality of the imagery in terms of resolution and sharpness is relatively poor. Before the flight we stored images of a calibration pattern for the estimation of an accurate interior orientation. The camera was calibrated based on the calibration tool of the *OpenCV* library<sup>2</sup> using a distortion model with three parameters for the radial and two parameters for the tangential distortion (Laganière, 2011). We assumed the geometry of the camera and, thus, the calibrated values of the interior orientation and the distortion parameters, to be constant during the flight and thus assigned these parameter values to all the collected images. Altogether, we collected about 100 images of the facade. Most of them were rejected because they were affected by disturbances due to transmission errors. Finally, we applied our algorithm to a sequence consisting of 20 images. The minimum number of matches for an image to be accepted for incremental bundle adjustment was set to 20.

For key point detection as well as for their descriptors we used SIFT features with three octave layers (Lowe, 2004). The detected key points were matched in the way described in section 4.1. Performing the estimation of the trifocal tensor with a point transfer threshold of five pixels resulted in a nearly blunder-free matching result (cf. figure 1, where the accepted matches are shown in green, the ones rejected by the validation of the tensor are shown in red). The facade has a rather repetitive structure, and though the majority of the matches seems to be correct, which in the figure is indicated by more or less parallel lines, there exists a considerable amount of blunders. The accepted matches provide an excellent basis for the estimation of initial values for the orientation of the triplet and the object points.

To give a visual quality analysis of our results we also determined the orientation parameters and the object coordinates of the tie points by a bundle adjustment using the software *PhotoModeler*<sup>3</sup>. We used identical observations and the same definition of the object coordinate system. However, *PhotoModeler* excluded some of the observations, either because they were assumed to be blunders or because they did not fulfil specific constraints, e.g. a minimum intersection angle in object space. The *PhotoModeler* result is considered as the reference solution in this paper. Figures 2(a) and 2(b) show the resulting point clouds generated by our algorithm after orienting the first 10 and all of the 20 images of our sequence, respectively. The point colours indicate their distances to the model points computed by *PhotoModeler*. The distance is given in a metric scale which we determined by evaluating the

<sup>2</sup><http://opencv.willowgarage.com/wiki/>

<sup>3</sup><http://www.photomodeler.com/>

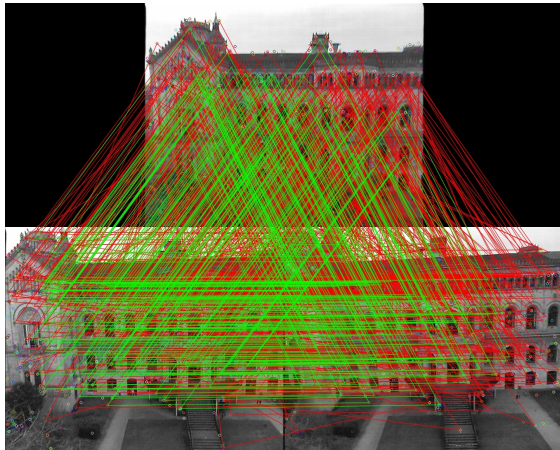
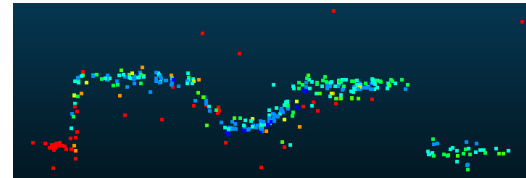


Figure 1: Triplet matching results after refinement by the trifocal tensor. Accepted matches are shown in green, rejected matches are shown in red colour.

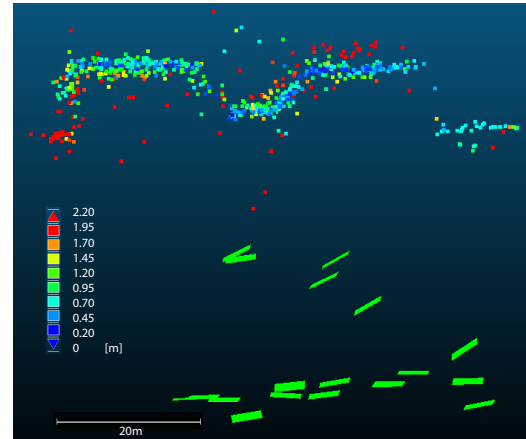
GPS-positions of the UAV during image acquisition. In figure 2(a) one can see that after ten images the structure of the facade is reconstructed rather well by the point cloud generated by our method. The front of the turret on the left side as well as some points on the ground in front of the building were rejected by *PhotoModeler*, which is the reason why for these points the distance to the nearest point in the model generated by *PhotoModeler* is rather large. After 20 oriented images our result becomes noisier (figure 2(b)). Especially on the right side of the central turret the facade plane is rotated probably due to a drift of the orientation.

Figure 3 indicates the images each object point is visible in. One can see that there are points that are measured in up to eleven images. The higher the number of views per point, the more stable the solution is. The fact that points are observed in more than three images should increase whenever a new image with observations for the respective point is included. We analysed the covariance matrix  $\Sigma_{\hat{x}\hat{x}}$  of the estimated parameters. As we use the variance factor  $\sigma_0 = 1$ , the covariance matrix of the estimated unknowns  $\Sigma_{\hat{x}\hat{x}}$  equals  $N_+^{-1}$  (Eq. 9). The blue curve in figure 4 shows the trace of the part of  $\Sigma_{\hat{x}\hat{x}}$  corresponding to the first 161 estimated object points. In green one can see the trace of the part of  $\Sigma_{\hat{x}\hat{x}}$  corresponding to the exterior orientation parameters of the first two images. The more images are introduced into the bundle adjustment the higher is the overall precision of the estimated parameters. Figure 4 also illustrates the improvement of the already estimated parameters in subsequent iterations (eq. 19). The mean point precision is approximately 6 – 7 cm in  $X$ - and  $Y$ - direction (approximately main plane of the facade) and about 25 – 28 cm in  $Z$ - direction (orthogonal to the facade). Given the block configuration (focal length  $c \approx 600$  pixel, distance between facade and projection centres about 50 m, base-height ratio between the extreme images of a triple 1:10), this corresponds to a precision of about 1 pixel in planimetry and 1/3 pixel in depth. We also cross-checked the relative distances between the projection centres of the first triplet derived by combined bundle adjustment using *PhotoModeler* with the distances estimated by our method. The variations are below 5 % of the estimated base length.

The computation time highly depends on the number of newly found object points. During the orientation of the 20 images shown in figure 2(b) the mean computation time for one triplet was about four seconds on a current desktop computer in a non-optimised implementation. In our experience this is an adequate time for the orientation of images acquired by a VTOL-quadro-



(a)



(b)

Figure 2: Vertical view of the resulting point cloud after orienting 10 images (a) and after orienting 20 images (b) using our algorithm. The positions of the image planes are represented by the green parallelograms in (b). Variations in size are caused by different viewing angles. Note that some camera stations only differ in height, so that the respective parallelograms are superimposed in (b). The point colours encode their point-to-point distances to the model computed using *PhotoModeler*.

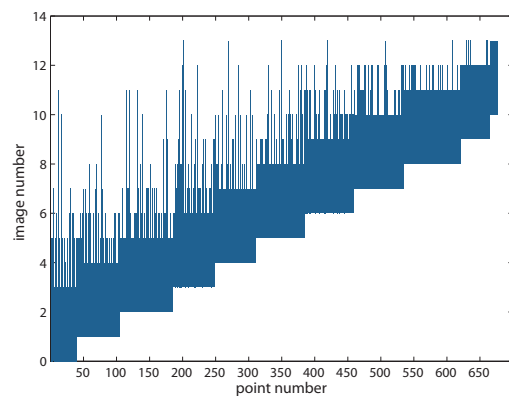


Figure 3: Indices of the images in the sequence in which each object point was visible.

copter to provide a visual on-line check of the completeness of the data acquisition as long as the velocity is low (in case of observing buildings normally the UAV is navigated manually with a velocity of about one meter per second).

## 6 CONCLUSIONS AND FUTURE WORK

In this paper we have presented an approach for an on-line orientation of a micro-UAV based on low resolution imagery. We were able to show that a low quality image sequence transmitted by a PAL-camera mounted on the UAV can be oriented using our algorithm, obtaining exterior orientations of the images and

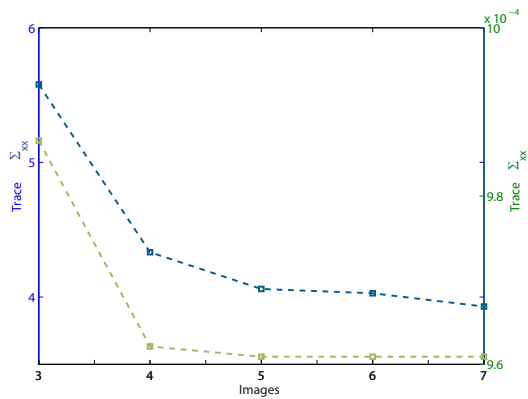


Figure 4: Trace criterion of the covariance matrix  $\Sigma_{\hat{x}\hat{x}}$  of the first 161 estimated object coordinates (blue) and of the exterior orientation parameters of the first two images (green) over five subsequent epochs of incremental bundle adjustment.

a sparse point cloud. Furthermore, the computation time is kept nearly constant and short using an incremental bundle adjustment allowing for a determination of the orientation parameters and the point cloud in near-real-time.

A possible use for the on-line orientation is to visually discover gaps in the sparse point cloud during the flight to avoid missing data. In figure 2(a) and 2(b) one can see that the vertical wall of the right turret is not represented sufficiently in the model. The orientations of the images shown in figure 2(b) reveal that this part of the facade is not directly observed by an adequate number of the oriented images. Hence for a complete model of the whole facade the flight plan would have to be refined, which could be done on-the-fly if the point cloud were inspected on-line by the operator.

However, our implementation also reveals several limitations. Firstly, the more images are iteratively added, the more unstable the solution becomes. We will approach this effect in future work by loop closure, following a strategy similar to (Meidow, 2012). Secondly, an absolute orientation of the resulting model is currently only possible in post-processing. To overcome this deficiency, we want to integrate telemetry data (containing GPS and IMU information) provided on-line by the downlink of the UAV in the bundle adjustment. Thirdly, the quality of the imagery and hence of the result of our algorithm is dependent on the transmission quality. On the one hand we will investigate ways to exclude defective imagery automatically, e.g. based on an analysis of the variance of the grey values. On the other hand, as the transmission defects are to a large degree a consequence of our hardware setup, we will try to avoid them altogether by using other transmission methods e.g. wireless local area network. Finally, once the incremental orientation has lost track, for instance due to severe wind, it may become necessary to navigate the UAV back to the last position successfully processed in order to generate new overlapping imagery. We want to automatically detect such situations in order to then be able to use autonomous way-point navigation to bring the UAV back on track.

For a further analysis of our results we plan to create a reference model of a distinct object in which the coordinate system is based on terrestrially measured ground control points. With this reference model we hope to be able to validate our approach also in terms of reliability and geometrical quality beyond a comparison of our results with those achieved by a commercial software.

## REFERENCES

- Bartelsen, J., Mayer, H., Hirschmüller, H., Kuhn, A. and Michelini, M., 2012. Orientation and dense reconstruction from unordered wide baseline image sets. *Photogrammetrie, Fernerkundung, Geoinformation (PFG) 2012(4)*, pp. 421–432.
- Beder, C. and Steffen, R., 2008. Incremental estimation without specifying a-priori covariance matrices for the novel parameters. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08.*, Anchorage, USA, pp. 1–6.
- Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM 24(6)*, pp. 381–395.
- Hartley, R. and Zisserman, A., 2000. *Multiple view geometry*. Vol. 642, Cambridge University Press.
- 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*, pp. 1–12.
- Kaess, M., Ranganathan, A. and Dellaert, F., 2008. iSAM: Incremental smoothing and mapping. *IEEE Transactions on Robotics 24(6)*, pp. 1365–1378.
- Klein, G. and Murray, D., 2009. Parallel tracking and mapping on a camera phone. In: *Proc. 8<sup>th</sup> IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'09)*, Orlando.
- Kraus, K., 1997. *Photogrammetry*. Vol. 2, 3 edn, Dümmler Verlag, Bonn.
- Laganière, R., 2011. *OpenCV 2 Computer Vision*. Packt-Publishing, 304 pages.
- Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60(2)*, pp. 91–110.
- Meidow, J., 2012. Efficient multiple loop adjustment for computer vision tasks. *Photogrammetrie-Fernerkundung-Geoinformation 2012(5)*, pp. 501–510.
- Nistér, D., 2000. Reconstruction from uncalibrated sequences with a hierarchy of trifocal tensors. In: *European Conference on Computer Vision, LNCS, Vol. 1842*, Springer Berlin Heidelberg, pp. 649–663.
- Ressl, C., 2000. An introduction to the relative orientation using the trifocal tensor. In: *International Archives of the Photogrammetry and Remote Sensing, Vol. XXXIII-B3*, pp. 769–776.
- Schneider, J., Läbe, T. and Förstner, W., 2013. Incremental real-time bundle adjustment for multi-camera systems with points at infinity. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XL-1/W2*, pp. 335–360.
- Triggs, B., McLauchlan, P. F., Hartley, R. I. and Fitzgibbon, A. W., 2000. Bundle adjustment - a modern synthesis. In: *Vision Algorithms: Theory and Practice, LNCS, Vol. 1883*, Springer Verlag, pp. 298–375.
- Wendel, A., Maurer, M., Graber, G., Pock, T. and Bischof, H., 2012. Dense reconstruction on-the-fly. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1450–1457.