A NEW APPROACH FOR AN INCREMENTAL ORIENTATION OF MICRO-UAV IMAGE SEQUENCES

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

KEY WORDS: orientation, UAV, convex optimisation, incremental bundle adjustment

ABSTRACT:

Civil applications for small size unmanned aerial vehicles (UAV) have become quite important in recent years and so have accurate orientation and navigation of these devices in unknown terrain. In this work we focus on on-line compatible positioning in facade observation based on monocular low resolution still images acquired by a camera mounted on a UAV. Also, a 3D point cloud of the facade is generated. This allows further processing steps, e. g. navigation assistance, collision avoidance or the evaluation of the point cloud density, verifying completeness of the data. To be able to deal with the increasing amount of observations and unknown parameters we implement an incremental bundle adjustment based on automatically determined tie points and sliding image triplets. The tripletwise orientation allows for an efficient double cross-check of the detected feature points and hence guarantees reliable initial values for the nonlinear bundle adjustment. The initial values are estimated within a convex formulation delivering a sound basis for the incremental adjustment. Our algorithm is evaluated by means of imagery we took of the facade of the Welfenschloss in Hannover, captured from a manually flown Microdrones md4-200 micro-UAV. We compare the orientation results of our approach with an approach in which initial values for the unknown object coordinates are computed algebraically.

1 INTRODUCTION

Comprehensive information about facades is important e. g. for different fields like architecture, cultural heritage or the construction of realistic 3D city models. UAVs allow for an efficient observation of facades because of their high flexibility and thus fill the gap between terrestrial and aerial photogrammetry which are often constrained by an inappropriate observation angle. Nowadays high-performance stabilising devices as well as light weight imaging sensors allow a manual navigation of a small UAV through narrow streets or close to buildings. Since the complexity of a building is not always known in advance, the completeness of the collected data cannot be guaranteed and a potential lack of data is not revealed until postprocessing. In order to avoid missing relevant information, the user must be able to adapt the flight path during the capture, which requires among other results, an on-line orientation of the UAV and a sparse point cloud of the environment.

In this work we present a new incremental orientation approach based on image triplets, which is able to deliver the position and orientation of the UAV and a sparse point cloud of the observed object in near real-time. We use image triplets to increase the robustness of image matching because triplets allow a double cross-check on the consistency of the keypoint matches between all image pairs that can be formed of an image triplet. We use a combination of projective and perspective geometry. The former is used to obtain initial values for the first image triplet. The latter forms the foundation of an incremental bundle adjustment. In this way we avoid effects of over-parametrisation and furthermore are able to integrate results of a precalibration of the camera in a straightforward way. A previous version of this approach is also described in (Reich et al., 2013), this paper contains two important extensions: Firstly, the estimation of initial values for the unknown object coordinates of the homologous feature points is formulated as a convex optimisation problem. Hence the risk of getting stuck into a local minimum leading to a distortion of the adjustment can be eliminated. Secondly, in the new version

we adaptively remove already oriented images from incremental adjustment based on the number of supporting object points measured in the related image to keep computation time constant.

The remainder of this paper is structured as follows. Section 2 gives an overview of work related to this topic. In section 3 the used hardware, namely the UAV and the camera are presented. In section 4 we describe the methodology of this approach with a focus on the convex formulation of three-image spatial intersection. Experiments and results are shown in section 5. Finally, section 6 summarises this work and gives an outlook into future prospects.

2 RELATED WORK

An extensive overview of UAV applications and experiments can be found in (Remondino et al., 2011). For our approach we need an on-line compatible implementation. The idea of on-line user assistance in UAV mapping is given in (Hoppe et al., 2012). Simultaneous georeferencing of images and 3D point cloud generation in real- or near real-time based on monocular imagery, which is also called simultaneous localisation and mapping (SLAM), has been in the focus of several publications: On-line orientation and dense-matching implementations based on projective geometry alone can be found in (Klein and Murray, 2009) and (Wendel et al., 2012), respectively. In these approaches a local and a global bundle adjustment optimise all unknown parameters in parallel to avoid drifting effects. In (Cesetti et al., 2011) a so called "visual global positioning system" is presented in which the position is estimated based on the matching of the acquired images with freely available georeferenced satellite imagery. A vision-aided navigation for UAVs based on a fusion of relatively oriented image pairs and data of an inertial measurement unit (IMU) is presented in (Wang et al., 2013). Steffen and Förstner (2008) describe the real-time orientation of a UAV for mapping purposes formulated in an unscented Kalman filter. The approach also deals with the drift problem in performing a loop closure based on points structured in a 3 dimensional octree.

---
*Corresponding author

The incremental bundle adjustment is comprehensively derived in (Beder and Steffen, 2008). Using an incremental bundle adjustment with a functional model different from ours, Meidow (2012) presents a loop closure approach for the purpose of stitching subsequent images. The idea of using triplets of images for the orientation of image sequences can be found in (Nistér, 2000). A hierarchy of trifocal tensors is used for the whole sequence without incremental estimation. The nature of the trifocal tensor and its estimation is described in a general context in (Hartley and Zisserman, 2000). The formulation of optimisation problems via a convex objective function is of increasing interest in research in many different disciplines. The entire derivation of its principles can be found in (Boyd and Vandenberghe, 2004). One of the first works integrating this technique into geometric reconstruction problems is (Hartley and Schaffalitzky, 2004). In (Ke and Kanade, 2007), the geometric interpretation of reconstruction problems is pointed out. Furthermore they propose an L_1-norm based reprojection error integrated into the objective function. The works of Olsson and Kahl (2010) and Kahl and Hartley (2008) describe the spatial intersection problem from three images and show that there is only one minimum in optimisation with the L_∞-norm.

In our approach we combine projective and perspective geometry and implement an on-line compatible orientation within an incremental bundle adjustment. The initial values for the bundle adjustment are estimated using convex optimisation.

3 PRELIMINARIES

There are many types of UAVs. We use a *Microdrones md4-200* micro-UAV[1], a vertical take-off and landing quadrocopter. It has a maximum payload of 300 g, which is enough for a high quality compact camera. We use a *Canon PowerShot S110*[2] in combination with the *Canon Hack Development Kit*[3]. This software allows to manipulate many relevant parameters including exposure time, sensor sensitivity and focal length. Furthermore it enables automatic capturing based on a fixed time interval. The camera is calibrated before the flight based on a five-parameter-distortion model, three parameters for radial and two parameters for tangential distortion, respectively (Laganière, 2011).

During the flight images are acquired with a time interval of two seconds. In the following we tackle the transmission of the imagery as a problem that is solvable but whose solution is not in the focus of this work. We perform the estimation in postprocessing, although our implementation allows for a near real-time orientation and point cloud computation.

4 METHODOLOGY

The orientation of the UAV is based on imagery captured by a precalibrated camera, though our implementation is able to estimate the parameters of the interior orientation and lens distortion during bundle adjustment for refinement. For each incoming image SIFT features are extracted and matched (Lowe, 2004), which are the basis for the subsequent orientation work-flow. This workflow can be divided into two steps. The first one deals with the initial image triplet and is explained in section 4.1. The second part covers the spatial resection of subsequent images and convex spatial intersection of feature points for the derivation of initial values (section 4.2) as well as the incremental bundle adjustment, in which the unknown parameters are refined (section 4.3).

---

[1] http://www.microdrones.com/
[2] http://www.canon.com/
[3] http://chdk.wikia.com/wiki/CHDK

4.1 Initial Image Triplet

The computation starts as soon as features of the three first images are available. Matching of feature points in image triplets is more robust than in image pairs. Based on a pair of matching feature points $[x^I, x^{II}]$ in two images $[I^I, I^{II}]$ one can perform a double cross-check, because for a triple $[x^I, x^{II}, x^{III}]$ also $[x^I, x^{III}]$ and $[x^{II}, x^{III}]$ have to be a match. Nevertheless, erroneous matches generally still remain after double cross-check. Therefore, the estimation of the trifocal tensor, based on seven three-point correspondences, is carried out based on RANSAC (Fischler and Bolles, 1981). Inliers that support an estimation are found by a point transfer into the third image using the fundamental matrix of the first two images derived from the tensor. More details can be found in (Reich et al., 2013) and (Hartley and Zisserman, 2000, Chapter 15).

The first image triplet defines the local coordinate system of the whole block. A certain number of supporting matches has to be found to consider the estimation of the trifocal tensor as valid and to proceed with the following steps. Otherwise the first image is rejected and computation starts again with the three subsequent images.

When a valid initial image triplet is found, we derive the fundamental matrices of the first two images from the trifocal tensor and estimate the relative orientation of the two images based on the calibration data of the camera. The local coordinate system (XYZ) is defined as follows: The origin is located in the projection centre of the first image, the X and Y axes are parallel to the x and y axes in the image, respectively. The Z axis coincides with the negative viewing direction. The scale is given by the distance between the projection centres of the first and the second image. That base length is set to one.

Initial values for the unknown object points are estimated by convex spatial intersection (see 4.2) and the relative orientation of the third image is computed by a RANSAC-based spatial resection (Kraus, 1997). Having initial values for all unknowns a robust bundle adjustment based on non-linear collinearity equations can be performed. Observations are iteratively re-weighted based on their residuals so that their contribution on the adjustment result can be controlled. In this way outliers in the observations can be detected and excluded from further estimation.

4.2 Convex Spatial Intersection

Subsequent images are added to the existing block based on spatial resection. First we perform a matching between the feature points of the new image and the ones of the previous images for which object coordinates have already been estimated. This allows us to compute the relative orientation of the new image analogously to the way described for the third image of the initial triplet in section 4.1.

In order to extend the existing 3D point cloud, we are interested in finding additional homologous points between the three images. The related 3D coordinates are to be estimated via incremental bundle adjustment, which needs appropriate initial values. These initial values are computed using convex spatial intersection.

Generally, the quality of a triangulated point is derived by its reprojection error in the images it is measured in:

$$f_{res,i}(\mathbf{X}) = \|\mathbf{x}_{proj,i} - \mathbf{x}_i\|_m \qquad (1)$$

with

$$x_{proj} = -c\frac{\mathbf{r}_{i,1}^T(\mathbf{X} - \mathbf{X}_0)}{\mathbf{r}_{i,3}^T(\mathbf{X} - \mathbf{X}_0)} + x_0,$$

$$y_{proj} = -c\frac{\mathbf{r}_{i,2}^T(\mathbf{X} - \mathbf{X}_0)}{\mathbf{r}_{i,3}^T(\mathbf{X} - \mathbf{X}_0)} + y_0. \tag{2}$$

$f_{res,i}(\mathbf{X})$ is the residual function for each image i based on the three-dimensional parameter vector $\mathbf{X}$, the object coordinates of the point to be intersected. It is the difference between the projected 3D point $\mathbf{x_{proj}}$ and the observation $\mathbf{x}_i$ in some norm m. c and (x_0, y_0) are the focal length and the principal point of the image, respectively, determined in camera calibration and considered to be constant. $\mathbf{X}_{0,i}$ is the position of the projection centre and $\mathbf{R}_i^T$ the transposed rotation matrix of image i with $\mathbf{r}_{i,j}^T$ being its jth row.

In the case of two images taking the midpoint of the shortest line that connects the two image rays which, in fact, is an algebraic minimisation with no obvious geometrical or statistical meaning (Hartley and Schaffalitzky, 2004), may not lead to suitable results. It is shown in (Kahl and Hartley, 2008) that algebraic minimisation in special cases may result in very high reprojection errors. Olsson and Kahl (2010) show that the reprojection error as a function of $\mathbf{X}$ (equation (1)) is quasiconvex. Since quasiconvexity is not preserved under summation (Boyd and Vandenberghe, 2004), L_2-minimisation (minimising the sum of squares) of the reprojection error does not represent a convex (or quasiconvex) optimisation. L_2-minimisation of the spatial intersection problem with three images mathematically means finding the minimum of a 47th order polynomial (Stewenius et al., 2005). Kahl and Hartley (2008) and Olsson and Kahl (2010) illustrate the chance to get stuck in a local minimum in three-image spatial intersection based on L_2-minimisation.

Therefore, we carry out an L_∞-minimisation as proposed in (Hartley and Schaffalitzky, 2004) and (Olsson and Kahl, 2010). This allows a quasiconvex formulation of the three-image spatial intersection problem:

$$\text{minimise} \qquad \max_{i=1}^{3} f_{res,i}(\mathbf{X}) \tag{3}$$

Taking the pointwise maximum of the three quasiconvex residual functions preserves quasiconvexity (Olsson and Kahl, 2010).

We reformulate (3) to derive a quasiconvex formulation for which an effective solver can be generated. By substituting equation (2) into equation (1) we obtain (note that while equation 3 represents the L_∞-norm, each reprojection error itself can be expressed in any norm m):

$$f_{res,i}(\mathbf{X}) = \left\| \begin{pmatrix} -c\frac{\mathbf{r}_{i,1}^T(\mathbf{X}-\mathbf{X}_0)}{\mathbf{r}_{i,3}^T(\mathbf{X}-\mathbf{X}_0)} + x_0 - x_i \\ -c\frac{\mathbf{r}_{i,2}^T(\mathbf{X}-\mathbf{X}_0)}{\mathbf{r}_{i,3}^T(\mathbf{X}-\mathbf{X}_0)} + y_0 - y_i \end{pmatrix} \right\|_m. \tag{4}$$

This can be extended to:

$$f_{res,i}(\mathbf{X}) = \left\| \begin{pmatrix} \frac{(-c\cdot\mathbf{r}_{i,1}^T+(x_0-x_i)\mathbf{r}_{i,3}^T)(\mathbf{X}-\mathbf{X}_0)}{\mathbf{r}_{i,3}^T(\mathbf{X}-\mathbf{X}_0)} \\ \frac{(-c\cdot\mathbf{r}_{i,2}^T+(y_0-y_i)\mathbf{r}_{i,3}^T)(\mathbf{X}-\mathbf{X}_0)}{\mathbf{r}_{i,3}^T(\mathbf{X}-\mathbf{X}_0)} \end{pmatrix} \right\|_m \tag{5}$$

$$= \left\| \begin{pmatrix} \frac{(-c\cdot\mathbf{r}_{i,1}^T+(x_0-x_i)\mathbf{r}_{i,3}^T)\mathbf{X}-(-c\cdot\mathbf{r}_{i,1}^T+(x_0-x_i)\mathbf{r}_{i,3}^T)\mathbf{X}_0}{\mathbf{r}_{i,3}^T\mathbf{X}-\mathbf{r}_{i,3}^T\mathbf{X}_0} \\ \frac{(-c\cdot\mathbf{r}_{i,2}^T+(y_0-y_i)\mathbf{r}_{i,3}^T)\mathbf{X}-(-c\cdot\mathbf{r}_{i,2}^T+(y_0-y_i)\mathbf{r}_{i,3}^T)\mathbf{X}_0}{\mathbf{r}_{i,3}^T\mathbf{X}-\mathbf{r}_{i,3}^T\mathbf{X}_0} \end{pmatrix} \right\|_m.$$

Now we can write equation (5) as

$$f_{res,i}(\mathbf{X}) = \frac{\|\mathbf{A}_i\mathbf{X} + \mathbf{b}_i\|_m}{\mathbf{c}_i^T\mathbf{X} + d_i}, \tag{6}$$

with

$$\underset{2\times 3}{\mathbf{A}_i} = \begin{pmatrix} -c\cdot\mathbf{r}_{i,1}^T + (x_0 - x_i)\mathbf{r}_{i,3}^T \\ -c\cdot\mathbf{r}_{i,2}^T + (y_0 - y_i)\mathbf{r}_{i,3}^T \end{pmatrix}, \qquad \underset{1\times 3}{\mathbf{c}_i^T} = \mathbf{r}_{i,3}^T,$$

$$\underset{2\times 1}{\mathbf{b}_i} = \begin{pmatrix} -\left(-c\cdot\mathbf{r}_{i,1}^T + (x_0 - x_i)\mathbf{r}_{i,3}^T\right)\mathbf{X}_0 \\ -\left(-c\cdot\mathbf{r}_{i,2}^T + (y_0 - y_i)\mathbf{r}_{i,3}^T\right)\mathbf{X}_0 \end{pmatrix}, \quad \underset{1\times 1}{d_i} = -\mathbf{r}_{i,3}^T\mathbf{X}_0.$$

Equation (6) is a quasiconvex function on the set $\mathbf{c}_i^T\mathbf{X} + d_i \geq 0$, implying that the point $\mathbf{X}$ is in front of the camera (Hartley and Schaffalitzky, 2004), (Kahl and Hartley, 2008). Equation (6) can geometrically be interpreted as a convex cone of norm m (Boyd and Vandenberghe, 2004) in front of the camera i with its apex in the projection centre. This cone intersects with the image plane defining some uncertainty measure (e. g. circle in case of the L_2-norm and square in case of the L_∞-norm).

Taking the pointwise maximum of equation (6) generally implies that the resulting function at some points is not differentiable anymore. Hence the minimum cannot be found by gradient based methods. Kahl and Hartley (2008) propose to write the minimisation of (6) as a convex feasibility problem

$$\begin{aligned} &\text{find} \qquad \mathbf{X} \\ &\text{subject to} \quad \|\mathbf{A}_i\mathbf{X} + \mathbf{b}_i\|_2 - \mu(\mathbf{c}_i^T\mathbf{X} + d_i) \leq 0, \quad (7) \\ &i = 1...3, \end{aligned}$$

with a fixed $\mu > 0$. Note that a L_2-norm is used for representing the reprojection error. The minimisation problem (7) can be geometrically thought of as finding the intersection of the three convex cones located at each of the projection centres of the three images. The size of the cones is controlled by the factor μ, which, in fact, is the radius of the intersecting circle in image plane. The problem (7) is feasible if μ is large enough so that the cones intersect (Ke and Kanade, 2007). Such a feasibility problem can be solved by a bisection algorithm (Olsson and Kahl, 2010), (Kahl and Hartley, 2008).

In our work we modify the feasibility problem (7) according to (Ke and Kanade, 2007) and reformulate it as a minimisation problem. μ is kept fixed and we try to minimise the distance ϵ to the cones defined by μ.

$$\begin{aligned} &\text{minimise} \qquad \epsilon \\ &\text{subject to} \qquad \|\mathbf{A}_i\mathbf{X} + \mathbf{b}_i\|_\infty + \mu(\mathbf{c}_i^T\mathbf{X} + d_i) \leq \epsilon, \quad (8) \\ &i = 1...3. \end{aligned}$$

In this way the intersection of the three cones defined by μ may be empty. If this is the case the minimal distance ϵ^* will be positive. If $\epsilon^* \leq 0$ an intersection of the cones exists and the optimal point X^* lies in that intersection. The implementation of the optimisation procedure is based on *CVXGEN*[4] (Mattingley and Boyd, 2012), which produces a *C*-based solver for convex optimisation problems. Note that in contrast to (Kahl and Hartley, 2008) we use a L_∞-based reprojection error, because *CVXGEN* is for quadratic programs only and using the L_2-norm would imply solving a second order cone program (Kahl and Hartley, 2008).

We are aware of the fact, that the L_∞-norm is sensitive to outliers. Hence, after the optimisation the reprojection error of each point (in each image) is required to lie below a certain maximum value.

4.3 Incremental Bundle Adjustment

As we want to provide an on-line compatible orientation procedure, unknown parameters are estimated incrementally. For each incoming image there are six parameters for the exterior orientation, namely the position and the orientation of the camera, and three object coordinates for each point measured in the current image triplet for the first time, enlarging the set of parameters of the whole block. Besides, already estimated parameters that are related to the current image triplet, can be improved by incremental bundle adjustment. For each triplet, two different types of unknowns occur:

1. unknowns which have already been estimated in previous triplets. These are

    (a) orientation parameters of the $N - 1$ already oriented images remaining in the incremental bundle adjustment

    (b) object coordinates of points that have been estimated in previous triplets and are used to tie the new image to the existing block (see beginning of section 4.2)

2. unknowns that are estimated for the first time, i. e.

    (a) orientation parameters of the new image

    (b) object coordinates of points that have been measured and triangulated in the current image triplet for the first time (see section 4.2).

The parameter vector $\mathbf{p}$ is split into two components, $\mathbf{p}_1$ and $\mathbf{p}_2$ for the first and second type of parameters, respectively. Similarly, we split the observation vector $\mathbf{l}$ into two components, $\mathbf{l}_1$ and $\mathbf{l}_2$. $\mathbf{l}_1$ contains all observations that were used in previous triplets, whereas $\mathbf{l}_2$ contains observations derived in the current image triplet. The incremental bundle adjustment can now be formulated as:

$$\begin{pmatrix} \mathbf{l}_1 + \mathbf{v}_1 \\ \mathbf{l}_2 + \mathbf{v}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} & 0 \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \end{pmatrix}. \qquad (9)$$

As one can see in equation (9) the new observations $\mathbf{l}_2$ are related to both types of parameters $\mathbf{p}_1$ and $\mathbf{p}_2$ via the design matrices $\mathbf{A}_{21}$ and $\mathbf{A}_{22}$, respectively. Therefore, not only the new parameters can be estimated ($\hat{\mathbf{p}}_2$) but also the previous parameters are improved ($\hat{\mathbf{p}}_{1,+}$):
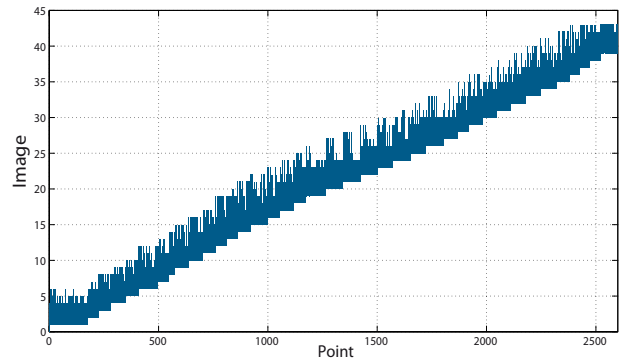
---
[4]http://cvxgen.com/

Figure 1: Indices of the images in the sequence in which each object point was visible.

$$\begin{pmatrix} \hat{\mathbf{p}}_{1,+} \\ \hat{\mathbf{p}}_2 \end{pmatrix} = \mathbf{N}_+^{-1} \begin{pmatrix} \mathbf{A}_{11}^T \mathbf{P}_{11} \mathbf{l}_1 + \mathbf{A}_{21}^T \mathbf{P}_{22} \mathbf{l}_2 \\ \mathbf{A}_{22}^T \mathbf{P}_{22} \mathbf{l}_2 \end{pmatrix}, \qquad (10)$$

with $\mathbf{N}_+^{-1}$ being the inverse normal equation matrix of the incremental adjustment and $\mathbf{P}_{11}$ and $\mathbf{P}_{22}$ the weight matrices derived from the inverse covariance matrices of the observations $\mathbf{l}_1$ and $\mathbf{l}_2$, respectively. The dimensions of the matrices involved in the estimation equations, thus the normal equation matrix of the already estimated parameters and $\mathbf{A}_{21}$, do increase over time. At some point older images (and their orientation parameters plus points that have been measured in that image for the first time) have to be removed from the orientation process to keep the size of the matrices small enough for near real-time processing. For each incoming image a minimum number of points found in the oldest image involved in the estimation process and re-measured in the current image has to be present to keep that image in the estimation process. Otherwise this old image is eliminated. For a complete derivation of the incremental bundle adjustment the reader is referred to (Beder and Steffen, 2008) or (Reich et al., 2013).

5 EXPERIMENTS

The evaluation of the presented approach is based on imagery showing the facade of the Welfenschloss in Hannover. The images were taken by the camera and micro-UAV presented in section 3 with a fixed focal length and with a time interval of two seconds between neighbouring images. The UAV was manually controlled in front of the facade in several distances and heights. We ended up with three separate flights each consisting about 250 images.

The analysis of these image sequences is focused on two specific aims. First, we want to show the general performance of our algorithm. Hence, we show if the computation time per image is kept constant and the effects of older images removed from the estimation if they are not relevant any more. Second, we demonstrate the improvement of the convex formulation of the estimation of the initial values with respect to an algebraic estimation that was presented in (Reich et al., 2013). The following analysis is based on one image sequence consisting of 233 images.

As mentioned in section 4.3 the criterion for keeping an image in the estimation process is a minimal number of points found in that image which are re-measured in the current image. In our experiments we set that number to 20 points. Figure 1 shows the distribution of points with respect to the images they are measured in for the first part (43 images) of the image sequence. The
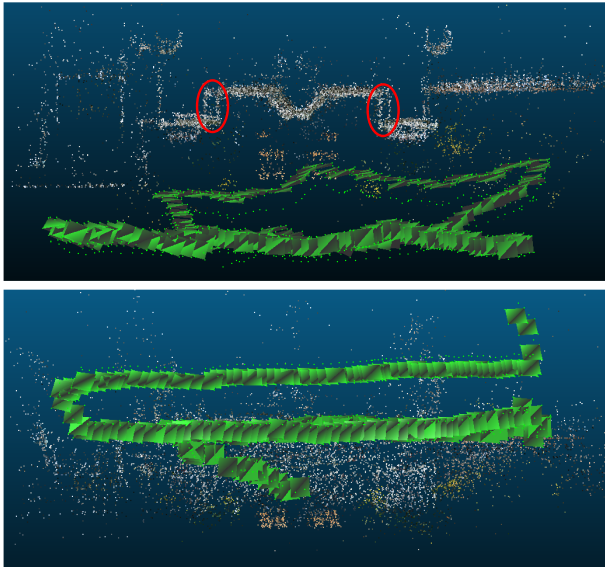
Figure 2: Point cloud and oriented images of an image sequence consisting of 233 images seen from above (first row) and from the front (second row). Regions that are highly affected by an offset are highlighted in red.

number of images involved in one iteration of the incremental bundle adjustment varies between five and seven. This interval is somehow restricted since we limit the number of new observations connected to already estimated points to 50 due to reasons of computation time. Furthermore, we select all observations depending on their location in image space to achieve a well distributed set of observations. A maximum number of 200 new observations is involved into each incremental bundle adjustment. The computation time per image is kept constant and amounts to about ten seconds on a standard desktop computer in a non-optimised implementation. This timing includes the computation of the whole covariance matrix of the parameters to be able to extract its precision. We estimate that an improvement by a factor of five is achievable, which delivers real-time results based on the time interval of two seconds between neighbouring images.

In figure 2 the point cloud and the orientation results of the whole image sequence are illustrated. One can see the uppermost image strip is captured in a rather oblique orientation so that also surface elements and trees are detected. Although the point cloud appears noisy it has to be said that the facade is quite rough with many decorative elements, window sills and balconies. Nevertheless, offset effects are present. The regions that are obviously affected either in scale or in translation are highlighted in red. The facade at that part of the building consists of a planar wall.

In a second experiment we compared the results of our approach with results for which the initial values of the object coordinates have been estimated using the shortest distance between the related rays (called 'algebraic approach' below). After each incremental bundle adjustment we extracted the standard deviations from the covariance matrix of the newly estimated parameters $\Sigma_{\hat{\mathbf{P}}_2 \hat{\mathbf{P}}_2}$. We computed the trace of that part of the matrix concerning the position of the projection centre and its three rotation angles, respectively, representing the sum of the standard deviations of the projection centre and that of the three rotation angles. Furthermore, we extracted the mean precision of all new object points after each incremental bundle adjustment iteration by averaging the trace of $\Sigma_{\hat{\mathbf{P}}_2 \hat{\mathbf{P}}_2}$ concerning the object points. Figure 3 illustrates the mean standard deviations of the estimated posi-
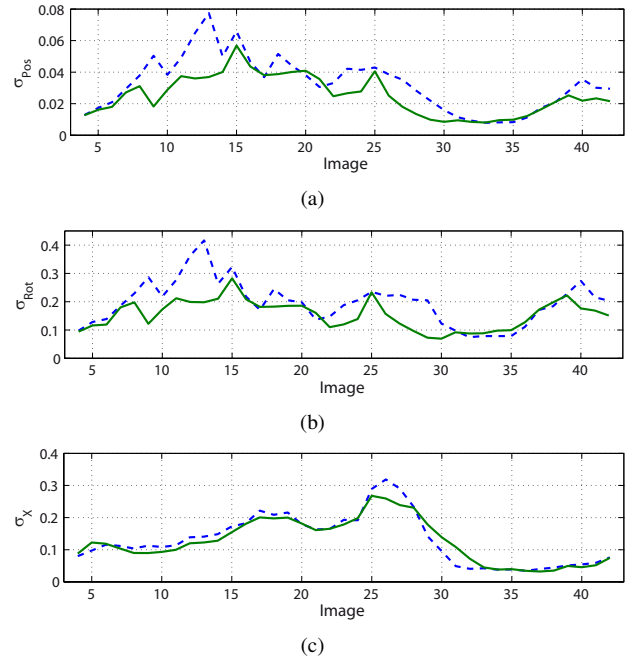

Figure 3: Mean standard deviations of the position $\mathbf{X}_0$ (a) and the rotation parameters $[\omega, \phi, \kappa]$ of the projection centre (b) and the object points (c) in $[baselengths]$ and $[gon]$ after each incremental bundle adjustment iteration for our approach (green/solid) and the algebraic approach (blue/dotted).

tions, rotations and object point coordinates of the first part of the sequence after each incremental bundle adjustment iteration. The color and type of the curve decode the type of initial value computation. The solid green curve shows the standard deviations for our approach, whereas the dotted blue curve shows the standard deviations for the algebraic approach. It can be seen that the precision of the orientation parameters in the case of convexly estimated initial values is nearly always better than for the algebraic approach, in particular for the images around image 13. The precision of the object points is rather equal for both approaches except for a few images near image 26. The fact, that differences in precision between the two approaches are visible for different images in the case of figure 3(a) and 3(b) compared to figure 3(c), may be an indication that the initial values itself play a secondary role only for the solution of the incremental bundle adjustment. Nevertheless, the way of initial value computation causes another factor influencing the precision of the bundle adjustment results as there is the number and distribution in space of points used. In fact, the distribution of object points observed in image 13 is better in our approach (std. deviation with respect to its centre of gravity: $3.2 \, [baselengths]$) than for the algebraic approach ($2.8 \, [baselengths]$).

Differences can also be seen in the point cloud. Figure 4 shows the resulting point clouds of the image sequence used above. The white point cloud results from our approach. A second point cloud estimated using the algebraic approach is depicted in colour where different colours depict different distances between the two point clouds. One can see that the observations and hence the estimated object points differ between the two approaches. The wall on the left side only exists in the result of our approach. As can be seen, the differences between the point clouds increase with each iteration (UAV flew from right to left). This can be explained by a rotation between the two solutions.
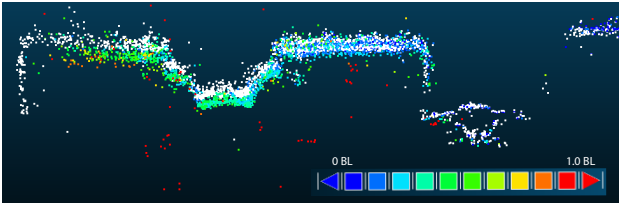
Figure 4: Comparison of the two resulting point clouds using our approach (white) and algebraically estimated initial values (coloured, depending on the distance to the other point cloud).

6 CONCLUSION

In this work we have presented a new incremental orientation approach of a micro-UAV based on triplets of images acquired at a certain time interval. We use convex estimation of the object coordinates as initial values for the incremental bundle adjustment. Furthermore, older images that become irrelevant for the orientation process are automatically removed from computation. We were able to show that our implementation estimates the orientation and a sparse point cloud of a whole flight consisting of hundreds of images with a constant computation time per image. In addition, we investigated the influence of different initial values for the unknown object point coordinates stemming from an algebraic computation and from our new approach.

However, there are several problems which need further investigation. Firstly, the comparison of the two approaches computing initial values for the bundle adjustment revealed differences in the estimation results. Whether the use of convexly optimised initial values itself or a deduced effect like the distribution of the object points in space is a significant reason for the enhanced quality of the results of our approach could not finally be verified. This will be part of our future work. Secondly, to be able to analyse the reliability and geometrical quality of our results we need a reference model with predefined and precisely measured points. Finally, the implementation is far from being applicable in real-time. But, irrespective of computation time, our approach is real-time compatible in terms of its incremental implementation.

References

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.

Boyd, S. P. and Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press.

Cesetti, A., Frontoni, E., Mancini, A., Ascani, A., Zingaretti, P. and Longhi, S., 2011. A visual global positioning system for unmanned aerial vehicles used in photogrammetric applications. Journal of Intelligent & Robotic Systems 61(1-4), pp. 157–168.

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 Schaffalitzky, F., 2004. L_∞ minimization in geometric reconstruction problems. In: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, Vol. 1, pp. I–504–I–509 Vol.1.

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.

Kahl, F. and Hartley, R., 2008. Multiple-view geometry under the L_∞-norm. Pattern Analysis and Machine Intelligence, IEEE Transactions on 30(9), pp. 1603–1617.

Ke, Q. and Kanade, T., 2007. Quasiconvex optimization for robust geometric reconstruction. Pattern Analysis and Machine Intelligence, IEEE Transactions on 29(10), pp. 1834–1847.

Klein, G. and Murray, D., 2009. Parallel tracking and mapping on a camera phone. In: Proc. 8th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'09), Orlando.

Kraus, K., 1997. Photogrammetry. Vol. 2, 3rd 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.

Mattingley, J. and Boyd, S., 2012. CVXGEN: a code generator for embedded convex optimization. Optimization and Engineering 13(1), pp. 1–27.

Meidow, J., 2012. Effcient 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.

Olsson, C. and Kahl, F., 2010. Generalized convexity in multiple view geometry. Journal of Mathematical Imaging and Vision 38(1), pp. 35–51.

Reich, M., Unger, J., Rottensteiner, F. and Heipke, C., 2013. Online compatible orientation of a micro-UAV based on image triplets. International Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W2, pp. 37–42.

Remondino, F., Barazzetti, L., Nex, F., Scaioni, M. and Sarazzi, D., 2011. Uav photogrammetry for mapping and 3d modeling–current status and future perspectives. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII-1/C22, pp. 25–31.

Steffen, R. and Förstner, W., 2008. On visual real time mapping for unmanned aerial vehicles. In: International Archives of the Photogrammetry and Remote Sensing, Vol. XXXVII-B3a, Beijing, pp. 57–62.

Stewenius, H., Schaffalitzky, F. and Nister, D., 2005. How hard is 3-view triangulation really? In: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, Vol. 1, IEEE, pp. 686–693.

Wang, T., Wang, C., Liang, J., Chen, Y. and Zhang, Y., 2013. Vision-aided inertial navigation for small unmanned aerial vehicles in GPS-denied environments. International Journal of Advanced Robotic Systems 10(276), pp. 1–12.

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.