

## INTEGRATION OF A GENERALISED BUILDING MODEL INTO THE POSE ESTIMATION OF UAS IMAGES

J. Unger\*, F. Rottensteiner, C. Heipke

Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany  
unger@ipi.uni-hannover.de

**KEY WORDS:** pose estimation, hybrid bundle adjustment, building model, unmanned aerial system, UAS

### ABSTRACT:

A hybrid bundle adjustment is presented that allows for the integration of a generalised building model into the pose estimation of image sequences. These images are captured by an Unmanned Aerial System (UAS) equipped with a camera flying in between the buildings. The relation between the building model and the images is described by distances between the object coordinates of the tie points and building model planes. Relations are found by a simple 3D distance criterion and are modelled as fictitious observations in a Gauss-Markov adjustment. The coordinates of model vertices are part of the adjustment as directly observed unknowns which allows for changes in the model. Results of first experiments using a synthetic and a real image sequence demonstrate improvements of the image orientation in comparison to an adjustment without the building model, but also reveal limitations of the current state of the method.

### 1. INTRODUCTION

The civil market of Unmanned Aerial Systems (UAS) is growing as UAS are used in a wide range of applications, e.g. in 3D reconstruction for visualization and planning, monitoring, inspection, cultural heritage, security, search and rescue and logistics. UAS offer a flexible platform for imaging complex scenes. In most applications the knowledge of the pose (position and attitude, exterior orientation) of the sensors in a world coordinate system is of interest. The camera on a UAS can be seen as an instrument to derive pose relative to objects in its field of view. However, as scale cannot be inferred from images alone, a camera is not able to deliver poses in a world coordinate system without the aid of additional sensors or ground control information. In addition, even if robust methods are applied, image-based parameter estimation suffers from accumulating errors by uncertain image feature positions that lead to block deformation (called “drift” in the following). Also, the limited payload capability of UAS and cost considerations constrain the selection of positioning and attitude sensors such as GNSS (Global Navigation Satellite Systems) receivers and IMUs (Inertial Measurement Units). As a result, directly measured data for the image pose are typically not accurate enough for precise positioning.

We propose a method to incorporate an existing generalised building model into pose estimation from images taken with a camera on board of the UAS. Whereas both the geometric accuracy and the level of detail of such models may be limited, the integration of this information into bundle adjustment is helpful to compensate inaccurate camera positions measured by GNSS, e.g. in case of GNSS signal loss if the UAS flies through urban canyons, and drift effects of a purely image-based pose estimation. The integration of the building model into bundle adjustment is based on fictitious observations that require object points to be situated on building model planes.

This paper is structured as follows. The next section outlines related work in which a-priori knowledge about the objects

visible to the sensor is introduced into the process of pose estimation. Section 3 introduces our scenario and outlines the mathematical model that is used to describe related entities. Section 4 presents our hybrid bundle adjustment with a focus on fictitious observations, whereas Section 5 contains the overall workflow of sensor orientation. Experiments using synthetic and real data are presented in section 6, before we conclude and give an outlook on future work in section 7.

### 2. RELATED WORK

Reviews of UAS technology and applications in mapping and photogrammetry are given in (Colomina and Molina, 2014) and (Nex and Remondino, 2014). The integration of object knowledge in image pose estimation and 3D reconstruction processes beyond ground control points (GCP) has been dealt with in various applications and with different motivations. First, there is work on the integration of **generic knowledge** about the captured objects into bundle adjustment. McGlone et al. (1995) provide the generic mathematical framework for including geometric constraints into bundle adjustment. Based on this work, Rottensteiner (2006) reviews different approaches for that purpose, comparing two different strategies: in adjustment, one can use “hard constraints”, involving constraints between the unknowns that will be fulfilled exactly, or “soft constraints” related to observation equations which, thus, can be subject to robust estimation procedures for detecting outliers. Consequently, he uses soft constraints to estimate the parameters of building models from sensor data. Gerke (2011) makes use of horizontal and vertical lines to obtain additional fictitious observations as soft constraints in indirect sensor orientation including camera self-calibration.

**Digital Terrain Models (DTM)** provide knowledge of a scene that is useful in image orientation. Strunz (1993), Heipke et al. (2005) and Spiegel (2007) carry out hybrid bundle adjustment using image observations and a DTM to constrain the heights of object points for improving pose estimation. Geva et al. (2015)

---

\* Corresponding author

deal with the pose estimation of image sequences captured in nadir direction from an UAS flying at a height of 50m in non-urban areas. Assuming the pose of the first frame to be known, they also derive surface intersection constraints based on DTM heights. Avbelj et al. (2015) address the orientation of aerial hyperspectral images. In their work, matches of building outlines extracted from a Digital Surface Model (DSM) in an urban area and lines in the images are combined in a Gauss-Helmert adjustment process.

Methods for integrating linear features are found in the field of **texturing 3D models**. Frueh et al. (2004) detect lines in oblique aerial imagery and match them against outlines of a building model. The matches are used in image pose estimation by exhaustive search. Other authors make use of corner points (Ding et al., 2008) or plane features (Hoegner et al., 2007) for texture mapping of building models. Hoegner et al. (2007) outline two strategies for image-to-model matching: They search for horizontal and vertical edges in the image and use their intersections as façade corner points that are matched to corners of the model. Alternatively, if not enough such vertices are observed in the images, homographies based on interest points that lie in a plane are estimated to orient images relative to façades.

Kager (2004) deals with **airborne laser scanning (ALS) strip adjustment**. He identifies homologous planar patches as tie features in overlapping ALS strips and uses these planar features to derive fictitious observations for the homogenisation of ALS strips. Hebel et al. (2009) find planes in laser scans acquired by a helicopter and match them to a database of planar elements (also from ALS) for terrain-based navigation. Matches are used to formulate constraint equations requiring the two planes to be identical, which are used to estimate the pose parameters.

Line matching between images and building models is also carried out with the direct goal of **orientation improvement**. Läbe and Ellenbeck (1996) use 3D-wireframe models of buildings as ground control, matching image edges to model edges and carrying out spatial resection for the orientation of aerial images. Li-Chee-Ming and Armenakis (2013) improve the trajectory of a UAS by matching image edges to edges of rendered images of an Level of Detail 3 (LoD3) building model and performing incremental triangulation.

In this paper, we incorporate object knowledge in the form of a generalised building model represented by planes and vertices. Instead of matching points, lines or planes directly, we use the object coordinates of tie points reconstructed from an image sequence and assign them to model planes based on a 3D distance criterion. In bundle adjustment, this assignment is considered by fictitious observations of the point distances to the model planes, using a mathematical model that can handle planes of any orientation. These fictitious observations act as soft constraints that improve the quality of pose determination beyond what can be achieved with low-cost GNSS receivers.

### 3. MATHEMATICAL MODEL

We address the scenario of a moving camera that observes objects in multi-view stereo configuration. Knowledge of the captured scene is given in form of a generalised building model. The building model is represented by its vertices and its faces. The topology is given by a list of the indices of vertices that belong to each model plane. Figure 1 depicts the relevant entities that we use to describe the building model and the cameras. In order to integrate the building model into bundle adjustment, we relate image coordinates to object points and assign these object

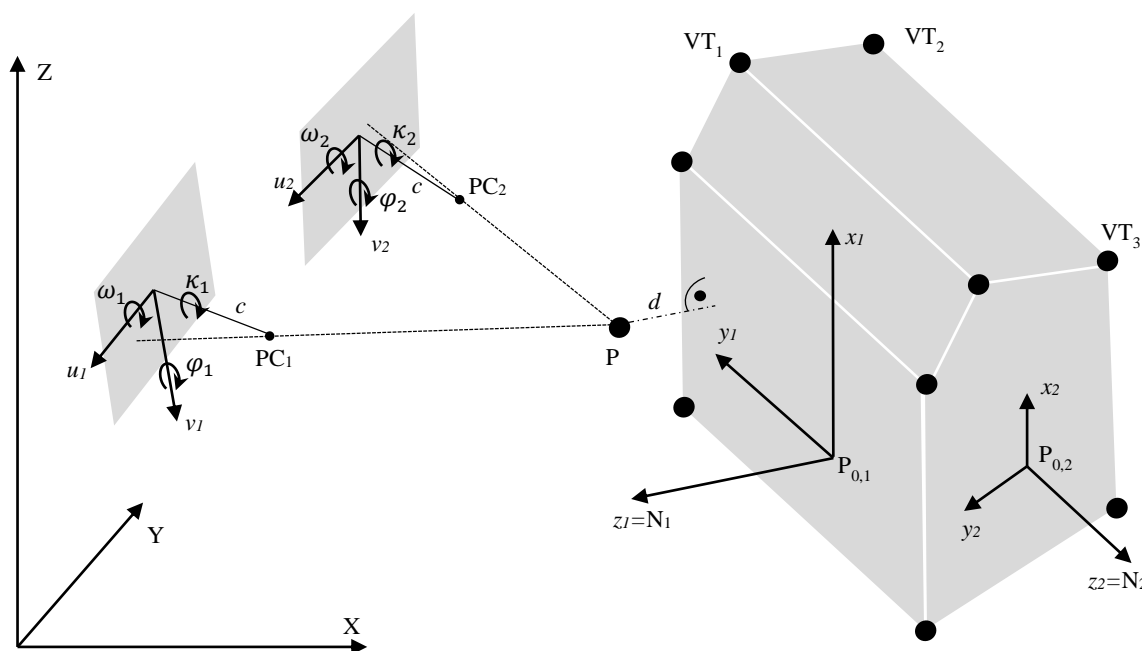


Figure 1: Relevant entities in our scenario. Two cameras  $i$  with identical camera constant  $c$ , image coordinate axes  $(u_i, v_i)$ , projection centres  $PC_i$  and three rotation angles  $(\omega_i, \phi_i, \kappa_i)$  with  $i \in \{1, 2\}$  represent the multi-view scenario where sensors capture an object point  $P$  in world coordinates  $X, Y, Z$ . The generalised building model is represented by corner points  $VT_k$  with  $i \in \{1, 2, \dots\}$  in world coordinates and by the planes they are situated in. Each plane  $j$  has a local coordinate system  $(x_j, y_j, z_j)$  where the local  $z_j$ -axis is the plane normal  $N_j$  and  $x_j, y_j$  are axes in the plane. The origin of the coordinate system of plane  $j$  is  $P_{0,j}$ , and each plane coordinate system is rotated relative to the world coordinate system by three angles  $(\omega_j, \phi_j, \kappa_j)$  that are not shown in the figure. The orthogonal distance of an object point  $P$  to a corresponding plane of the building model is denoted by  $d$ .

points to planes of the building model. Note, that there is no need to observe the vertices in the images, which would require to solve a complex image interpretation task.

The mathematical model that relates image coordinates  $u, v$  to the parameters of interior and exterior orientation and to the object coordinates  $X, Y, Z$  is given by the well-known collinearity equations (Eq. 1).

$$\begin{aligned} u &= u_0 - c \frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} \\ v &= v_0 - c \frac{r_{12}(X - X_0) + r_{22}(Y - Y_0) + r_{32}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} \end{aligned} \quad (1)$$

The exterior orientation (pose) of an image is given by the coordinates  $X_0, Y_0, Z_0$  of its projection centre PC and the elements  $r_{ij}$  of a rotation matrix which are functions of three rotation angles  $\omega, \varphi, \kappa$ . The coordinates of the principal point  $u_0$  and  $v_0$  (not shown in Figure 1) and the camera constant  $c$  are referred to as interior orientation parameters.

Similar to Kraus (1996), we use a local coordinate system attached to each plane in which we formulate the fictitious observation equations for points situated on that plane. Six parameters describe the pose of this local plane coordinate system  $x, y, z$ . These are three rotation angles (used to parameterise a 3D-rotation matrix  $R$ , not shown in Figure 1) and a 3D-shift  $P_0$  from the object coordinate system to the local one for each plane.  $P_0$  is initialised in the centre of gravity of the building model vertices of the plane. Initially, the  $x$ - $y$  plane of the local system corresponds to the model plane and the  $z$ -axis corresponds to the plane normal  $N$ .

To describe a plane within such a local system, it is parameterised by two angles  $\alpha, \beta$  defining the direction of the normal and a translation  $\delta$  along the (local)  $z$ -axis (see Figure 2). Using this parameterisation, the relation between a point and a plane is described by its orthogonal distance to that plane following Eq. 2.

$$d = N(\alpha, \beta)^T \cdot \bar{P}(X, Y, Z) + \delta \quad (2)$$

where  $N(\alpha, \beta) = \begin{bmatrix} \sin \alpha \\ -\sin \beta \cos \alpha \\ \cos \beta \cos \alpha \end{bmatrix}$  and  $\bar{P}(X, Y, Z) = R^T(P - P_0)$

is the object point expressed in the local coordinate system. Note that whenever the parameters  $\alpha, \beta$  and  $\delta$  are changed, we use these values to adapt  $R$  and  $P_0$ , so that after the parameter update, the adjusted plane again corresponds to the (slightly shifted and rotated)  $x$ - $y$  coordinate plane of the local system.

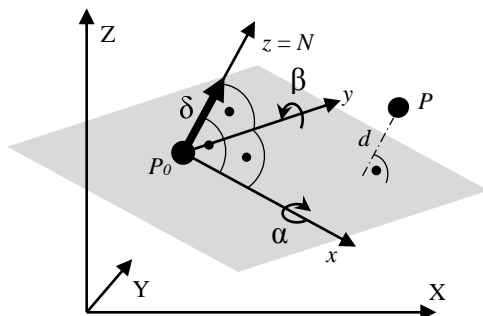


Figure 2: Local plane parameterisation with two angles  $\alpha, \beta$  and a shift  $\delta$  (bold arrow) along the local  $z$ -axis, which is the plane normal  $N$ .  $d$ : distance of a point  $P$  from the plane.

#### 4. HYBRID BUNDLE ADJUSTMENT

We use various types of observations in our adjustment problem:

- image coordinates of homologous points ( $u, v$ )
- direct observations  $(X, Y, Z)_{GNSS}$  for the projection centres of the cameras, obtained from low accuracy GNSS receivers
- direct observations for the vertices of the building model  $(X, Y, Z)_{VT}$
- fictitious observations relating object space coordinates of a tie point to the planes of the building model ( $d$ )
- fictitious observations relating object space coordinates of a vertex of the building model to the planes of the building model ( $d$ )

These observations are used as inputs into a Gauss-Markov model to estimate the following unknowns:

- the pose parameters for each image (three rotation angles  $\omega, \varphi, \kappa$  and projection centre coordinates  $X_0, Y_0, Z_0$ )
- the object space coordinates of the tie points  $(X, Y, Z)_{TP}$
- three parameters of each plane of the building model ( $\alpha, \beta, \delta$ )
- the object space coordinates of the vertices of the building model  $(X, Y, Z)_{VT}$

The latter two groups of unknowns reflect the fact that the building model is generalised. Due to the generalisation it is possible that the vertices of the building model do not correspond to real points at the object surface, so that they might not be observable in an image. The direct observations of vertex coordinates relate the estimated planes to the original building model.

##### 4.1 Functional Model

The following observation equations are formulated in our model:

- $v_{u,v} = f(\hat{X}, \hat{Y}, \hat{Z}, \hat{X}_0, \hat{Y}_0, \hat{Z}_0, \hat{\omega}, \hat{\varphi}, \hat{\kappa}) - (u, v)$   
according to the classical collinearity equations (Eq. 1). Note that we consider the interior orientation of the sensor as known and constant.
- $v_{(X,Y,Z)_{GNSS}} = (\hat{X}_0, \hat{Y}_0, \hat{Z}_0) - (X, Y, Z)_{GNSS}$   
for the direct observations of the projection centres. The offset of the GNSS receiver to the camera pose is neglected due to the expected relatively low accuracy of the receiver.
- $v_{(X,Y,Z)_{VT}} = (\hat{X}, \hat{Y}, \hat{Z})_{VT} - (X, Y, Z)_{VT}$   
for the direct observations of the building model vertices.
- $v_d = N(\alpha, \beta)^T \cdot \bar{P}(X, Y, Z) + \delta$   
for fictitious observations according to Eq. 2.

For each tie point and for each vertex of the building model one such fictitious observation according to Eq. 2 relates the object space coordinates to a plane of the building model. The distance  $d$  between the point and the plane is assumed to be zero, i.e. the point is assumed to lie in the plane. For the vertices of the building model it is exactly known which plane they are situated in. In contrast, relations between tie points and model planes must be established first (see section 5).

The three parameters  $\alpha, \beta$  and  $\delta$  per plane are unknowns in the iterative adjustment. However, the rotation  $R$  and the translation  $P_0$  are treated as constants during each iteration. As stated previously,  $R$  and  $P_0$  are updated after each iteration using the estimated local plane parameters  $\alpha, \beta$  and  $\delta$ .  $\alpha, \beta$  and  $\delta$  are initialised as zero and reset to zero after updating  $R^T$  and  $P_0$  for each iteration.

## 4.2 Stochastic Model

We assume uncorrelated observations and a constant a-priori level of accuracy for each observation type. This leads to a diagonal variance-covariance matrix of the observations  $\Sigma_{ll}$ :

$$\Sigma_{ll} = \text{diag}(\sigma_{img}^2, \sigma_{GNSS}^2, \sigma_{VT}^2, \sigma_{dTP}^2, \sigma_{dVT}^2). \quad (3)$$

In Eq. 3, the variances of the measured image coordinates are denoted by  $\sigma_{img}^2$ . The variance of the GNSS receiver measurements is reflected by  $\sigma_{GNSS}^2$ . The variance  $\sigma_{VT}^2$  is related to the accuracy of the coordinates of the building model vertices. For the two groups of fictitious distance observations we introduce different variances, namely  $\sigma_{dTP}^2$  for tie points and  $\sigma_{dVT}^2$  for vertices. The vertices are known to lie exactly on their planes. Therefore, their fictitious distance observations conceptually must be zero (for numerical reasons we use a small variance  $\sigma_{dVT}^2$  resulting in high weights of these observations). On the other hand, the variance  $\sigma_{dTP}^2$  of the observed distance of tie points to their related planes mainly depends on the generalisation and the accuracy of the building model and has to be selected accordingly.

## 5. PROCESSING STEPS

Our processing workflow consists of the steps listed in Table 1. We first derive homologous points, estimate image poses and 3D object point coordinates based on a structure from motion (SFM) pipeline. Subsequently, we run a bundle adjustment including only images for which GNSS observations are available and without considering the building model (step 2). Images having GNSS coverage are assumed to be connected in the sequence (usually at the beginning or at the end of a flight).

Step 1	Image matching and SFM to derive tie points and image poses
Step 2	Bundle adjustment including only images, for which direct observations of projection centres are available.
Step 3	Establishment of relations between tie points and model planes
Step 4	Hybrid bundle adjustment including the planes (step 3 is carried out before each iteration)
Step 5	Hybrid bundle adjustment based on images and planes already used in step 4 and including new images, new tie points (step 3 is carried out before each iteration only considering planes already used in step 4)
Step 6	Hybrid bundle adjustment based on all images and planes considered in step 5 and new model planes for the points added in step 5 (step 3 considering all model planes is carried out before each iteration)

Table 1: Work flow of pose estimation.

In step 3 we assign tie points to the planes of the building model on the basis of their estimated 3D positions. Note that both, the observations of the image projection centres and the building model vertices, must be given in the same coordinate system, here the coordinate system of the GNSS observations. The assignment of a point to a plane is based on a distance criterion. In our current implementation tie points are assumed to be related to the closest plane provided that the Euclidean distance from the plane is below a given threshold. This threshold has to be selected in accordance with the accuracy and degree of generalisation of the building model.

Each tie point can add only one fictitious observation. We do not consider tie points to be related to more than one plane at the same time (e.g. points on plane intersections and corners). Only if the distance of a tie point to the nearest plane is below the threshold, the relation is considered to be correct and a fictitious observation is added to the adjustment. In contrast to the tie points, the relation of the vertices to the planes are known and each vertex can be related to more than one plane.

In step 4, hybrid bundle adjustment is carried out with the additional observations and parameters for the adjusted planes and tie points of step 2 as described in section 4. In each iteration the assignment of the tie points to the planes of the building models to set up the fictitious observations is recomputed based on the current parameter values (step 3). In contrast, the known relations of vertices to planes are not changed. Note that only planes containing more than a pre-defined minimum number of tie points are considered in adjustment.

Step 5 is a hybrid adjustment that additionally includes the images having no direct observations for the projection centre, which is carried out to transfer the remaining images into the object coordinate system using the ground control information of the part of the block already utilised in step 4. In step 5 additional model planes are not considered, in contrast to step 6, where the results of step 5 are used to find assignments of the new tie points to those additional model planes. Finally, a hybrid adjustment with all images including all planes that contain a sufficient number of tie points is carried out, which delivers the final results of our method.

## 6. EXPERIMENTS

In our experiments, we show results achieved both for simulated data and for real images captured by a micro UAS. Both scenarios use a 3D city model with Level of Detail 2 (LoD2) of a part of our campus as ground control information. For both sequences the viewing directions of the cameras are approximately horizontal and orthogonal to both, the flight direction and the facades. Both data sets have GNSS coverage for several images at the beginning of the image sequences. Both, GNSS observations and building vertices are given in WGS84/UTM Zone 32, which, after applying a fixed offset to reduce the number of digits, serves as our world coordinate system. The a-priori standard deviations of all observation types, used to define the stochastic model (cf. Eq. 3) are set as follows:

Image coordinates	$\sigma_{img}$	$\pm 1$ pix
GNSS obs. of projection centres	$\sigma_{GNSS}$	$\pm 2$ m
Building model vertices	$\sigma_{VT}$	$\pm 0.5$ m
Fictitious distance for tie points	$\sigma_{dTP}$	$\pm 0.3$ m
Fictitious distance for vertices	$\sigma_{dVT}$	$\pm 0.01$ m

$\sigma_{VT}$  reflects the accuracy and generalisation effects of the vertices of the building model.  $\sigma_{dTP}$  describes the deviation of the model planes due to the generalisation. In step 3 of the processing pipeline, we choose to take into account fictitious distances for points to planes only if the distance is smaller than 2 m. The threshold is chosen in accordance with the GNSS accuracy to obtain as many correct assignments as possible with few outliers only. Planes are adjusted only if at least 20 points are assigned to them. We noticed that planes having fewer points have a high probability to be reconstructed incorrectly.

## 6.1 Simulation

For the simulation, a trajectory of 41 images with a length of 190 m along the LoD2 model is simulated. Object points are distributed randomly in the planes of the building model with a density of 0.4 points per  $m^2$ . The points are re-projected into the images to generate image coordinate observations. Random Gaussian noise with a standard deviation of 1 pixel is added to these image coordinates. The positions of the first 10 images serve as simulated GNSS observations for the projection centres, they are contaminated by white noise of  $\sigma = 2$  m.

Figure 3a shows the resulting camera positions and tie points. The datum is defined by the GNSS observations of the first 10 images, which results in strong deviations of the block relative to the building model (highlighted by red ellipses in the figure). Figure 3b depicts the improvements for tie point and camera positions after including the model planes that are visible in the images of the first sub-block (step 4). After this step, the datum of the block is defined by both, the direct observations of the projection centres and the vertices of the building model. The adjusted points coincide very well with the buildings model planes. As expected, changes of the vertices of the building model are in a range of just a few centimetres due to the fact that the simulated points originally exactly coincided with the model planes. Figure 3c shows the tie points and camera positions of the last part of the simulated image sequence after adding images 11 to 41 to the hybrid adjustment in processing step 5; as ground control is only available for the first part of the sequence, there are considerable deviations of the resultant point cloud from the model. Figure 3d depicts the result after adding model planes to the hybrid adjustment in step 6. The hybrid adjustment is shown to be able to adjust the deviations that were present after step 5 by moving cameras and tie points towards the model. The black dots denote estimated tie points that can be seen to coincide with the walls in comparison to the magenta points estimated in step 5. The blue ellipses highlight areas where this improvement is most obvious.

Figure 4 compares the a-posteriori standard deviations of the estimated 3D tie points with and without considering the plane relations with the first 10 images and with all images. Between steps 2 and 4 as well as between steps 5 and 6, the tie point precision improves clearly when adjusting the points with the building model. Especially the Z-direction shows strong improvements. The relative differences in precision of points remain similar as they depend mainly on the number of images observing a point. Adding images without new planes in step 5 yields higher standard deviations for the new tie points (Note that point indices are not ordered and change from the top figures to the bottom ones). All points are clearly improved by considering additional model planes in step 6, with an estimated precision of the tie points in the order of  $\pm 0.2$  m.

## 6.2 Real Data

For the acquisition of a real image sequence we used a manually controlled DJI Matrice 100 quadcopter with gimbal stabilised Zenmuse X3 camera. We used the same area as for the simulations, but with a different trajectory. The camera has a fixed focus, 3.61mm focal length and a 1/2.3" CMOS sensor having 4000x3000 pixels and a pixel size of 1.5  $\mu m$ . Images were taken automatically every 2 seconds. The image sequence consists of 183 images with an average ground sampling distance of 6 mm/pixel. On average, there was a five-fold overlap, so that

on average, each tie point was observed in five images. The used GNSS device receives GPS, GLONASS and SBAS satellites signals. The flying height above ground was up to 20 m at the beginning to obtain good GNSS signals and about 2 m for the last part of the flight. The surrounding buildings are 4 to 30 m high. Even between the buildings, GNSS signals from at least 5 satellites were received at each camera position. To be able to also test our processing pipeline using images without GNSS coverage we only considered GNSS observations for the first 110 images of the sequence.

Image distortion was eliminated prior to processing based on available interior orientation parameters. The processing steps 1 and 2 were carried out using the commercial Software Agisoft PhotoScan Pro<sup>1</sup>. In the adjustment of step 2, the GNSS observations for the first 110 images were considered to define the datum. In the subsequent steps, image coordinates exported from PhotoScan were used as observations in our hybrid bundle adjustment (steps 3 and 4); similarly, exported orientation parameters and object point coordinates served as initial values for the unknowns. We only exploit tie points that are observed in at least three images and are considered to be inliers by PhotoScan. This is done to minimise the number of outliers, as at this stage our adjustment does not yet handle outliers in the observations. After eliminating points as described above, there remain 5400 object points in the block.

We show the result for the first part (110 images with GNSS) of the image sequence with considering images and model planes (after step 4) in Figure 5. A comparison of the initial point positions with the building model shows that the distances of most points from their corresponding planes are below 2 m (i.e., below the expected accuracy of GNSS). GNSS for all images yields proper georeferencing which allows for the initialisation of the fictitious observations for our hybrid adjustment despite the simple distance criterion used for assigning tie points to model planes.

Regarding corrections to planes, we often observe ground points being erroneously assigned to wall planes or planes only partly covered by tie points, which results in adjusted walls that are no longer vertical. Furthermore, tie points on building details not contained in the model due to generalisation or points on vegetation close to the building also introduce errors to the parameters of these planes. The profile shown in Figure 5 (right) shows a plane (blue ellipse) that is affected by complex structures not represented in the generalised building model.

We observe several limitations of the current state of the method: There are a few remaining outliers and quite a few points on structures not represented in the building model. The simple distance criterion leads to wrong assignments of points to planes that cannot adequately be handled by our method yet.

Figure 6 shows the results achieved after adding the remaining images in step 5 and including additional model planes in step 6; Figure 7 shows the improvement of the a-posteriori standard deviations of tie points. Whereas the improvement is relatively small between steps 2 and 3 due to the relative large number (110) of GNSS positions used in the adjustment, the tie points observed only in images without GNSS coverage profit most from the inclusion of the model planes: the corresponding object space coordinates have a standard deviation smaller by a factor of two for points in the range of point indices 3000 to 4000.

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

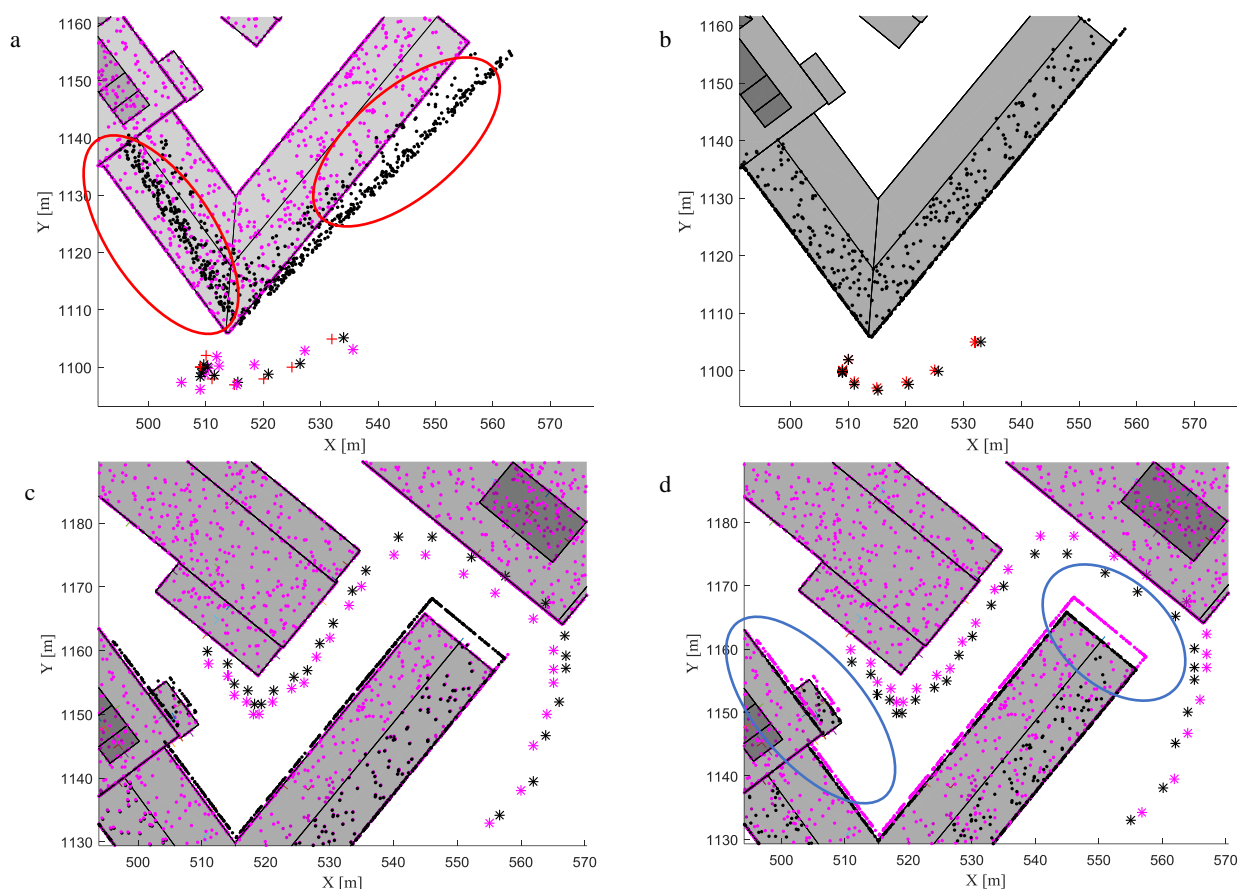


Figure 3: Results of adjustment after different processing steps. Black dots: estimated points; black asterisks: estimated camera positions; magenta: initial positions of points (dots) and camera positions (asterisks), typically the results of the previous step; red crosses: simulated noise-free camera positions. The building model is super-imposed to these results. a) Adjustment without planes (step 2); the initial values of the camera positions correspond to the GNSS positions. The red ellipses indicate deviations of the results relative to the building model. b) Adjustment with planes (step 4). c) Adjustment including new images (step 5). d) Adjustment including new images and new planes (step 6). The blue ellipses highlight areas where the results of step 5 differ from the model and the adjusted black points coincide with a wall.

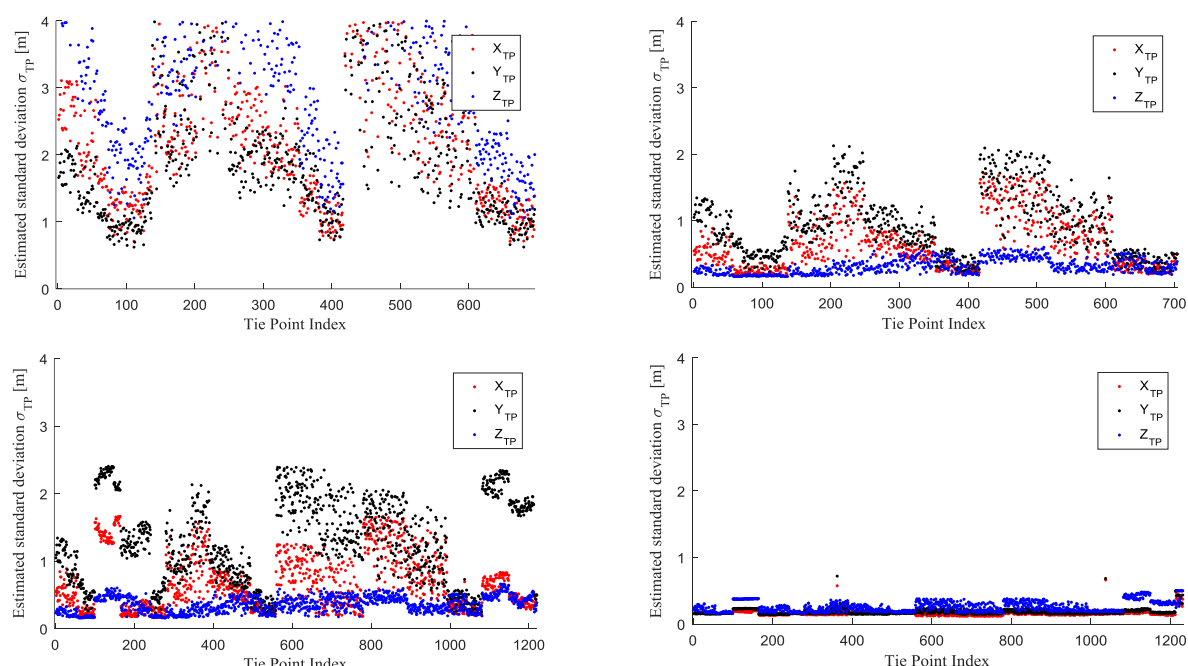


Figure 4: A-posteriori standard deviations of the estimated tie point coordinates in object space from the simulated data after steps 2 (top left), 4 (top right), 5 (bottom left) and 6 (bottom right).

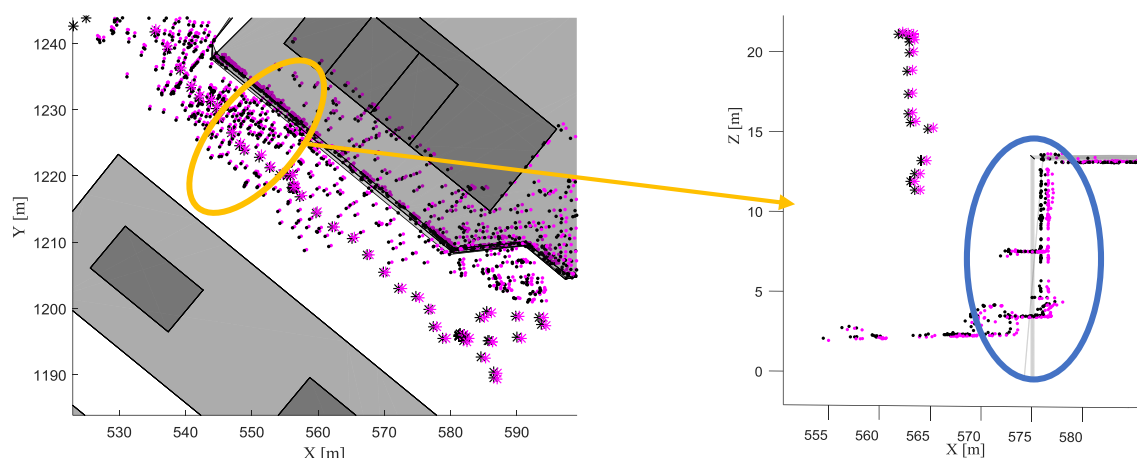


Figure 5: Detail of the results of hybrid adjustment with real data after including images with GNSS and planes (step 4). Black dots: estimated tie points; black asterisks: estimated camera positions; magenta dots/asterisks: initial tie point and camera positions from step 2, respectively. The right part of the figure shows a cross-section through the area highlighted by an orange ellipse in the left part of the figure (see main text for a discussion).

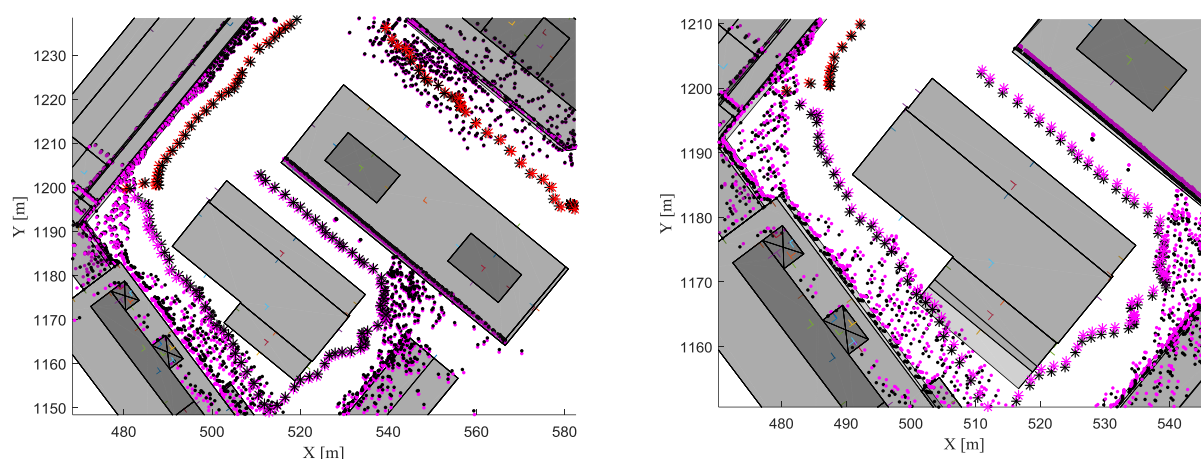


Figure 6: Results of two variants of hybrid adjustment. Black dots: estimated tie points; black asterisks: estimated camera positions; magenta dots/asterisks: initial positions of tie points/camera positions, i.e., results of the previous processing steps. Red asterisks: GNSS observations of camera positions. Left: Results of the hybrid adjustment with real data after step 5 including new images but no new planes. Right: Results after step 6, including all planes..

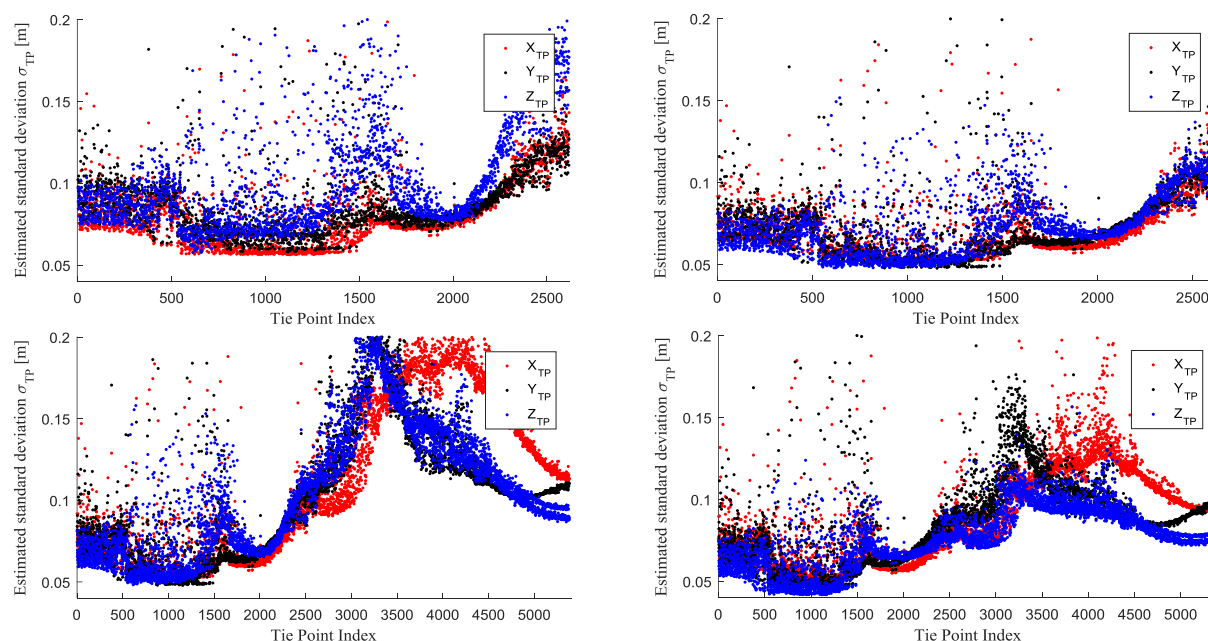


Figure 7: A-posteriori standard deviations of the tie point coordinates in object space after steps 2 (top left), 4 (top right), 5 (bottom left) and 6 (bottom right).



## 7. CONCLUSION AND FUTURE WORK

The method presented in this paper allows for the integration of a generalised building model into the pose estimation of image sequences captured by an UAS. The building model is integrated by fictitious observations of the distances between tie points and model planes. Points are assigned to model planes on the basis of a simple distance criterion.

Our experiments based on simulated data show that the inclusion of a building model results in a considerable improvement of the precision of the resultant 3D points and in a better alignment of the estimated object points with the model. On the other hand, the experiments based on real data show remaining challenges. The main problem is to find correct matches between tie points and model planes; our simplistic technique based on a distance criterion proves not to be sufficient. Nevertheless, the adjustment procedure did result in an improvement of the estimated precisions of the tie point coordinates.

In our future work we will address the problem of ambiguous assignments between points and model planes. A next step will be the implementation of robust estimation to detect outliers. The matching process between tie points and planes can be improved by considering the estimated precision of tie point coordinates to adapt the distance threshold for assigning points to planes, replacing the decision by a hypothesis test. Further, to examine the influence of the LoD of the model on the results of our method, experiments with models of different degrees of generalisation will be carried out. Further developments will consist in a proper handling of occlusions to reduce the number of plane candidates for each tie point and the integration of a point cloud segmentation to detect planes that are not part of the model.

## REFERENCES

- Avbelj, J., Iwaszczuk, D., Müller, R., Reinartz, P., Stilla, U., 2015. Coregistration refinement of hyperspectral images and DSM. An object-based approach using spectral information. *ISPRS Journal of Photogrammetry and Remote Sensing* 100, pp. 23–34.
- Colomina, I., Molina, P., 2014. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 92, pp. 79–97.
- Ding, M., Lyngbaek, K., Zakhor, A., 2008. Automatic registration of aerial imagery with untextured 3D LiDAR models. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8.
- Früh, C., Sammon, R., Zakhor, A., 2004. Automated texture mapping of 3D city models with oblique aerial imagery. *Proceedings. 2nd International Symposium on 3D Data Processing, Visualization and Transmission. 3DPVT 2004*, pp. 396–403.
- Gerke, M., 2011. Using horizontal and vertical building structure to constrain indirect sensor orientation. *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (3), pp. 307–316.
- Geva, A., Briskin, G., Rivlin, E., Rotstein, H., 2015. Estimating camera pose using Bundle Adjustment and Digital Terrain Model constraints. *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4000–4005.
- Hebel, M., Arens, M., Stilla, U., 2009. Utilization of 3D city models and airborne laser scanning for terrain-based navigation of helicopters and UAVs. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII*, p. Part 3/W4, pp. 187–192.
- Heipke, C., Albertz, J., Attwenger, M., Buchroithner, M., Dorninger, P., Dorrer, E., Gehrke, S., Gwinner, K., Lehmann, H., Mayer, H., Ostrovskiy, A., Pacher, G., Rentsch, M., Schmidt, R., Scholten, F., Spiegel, M., Stilla, U., Neukum, G., and the HRSC Co-Investigator Team, 2005. HRSC auf Mars-Express-Photogrammetrische und kartographische Auswertungen. *Zeitschrift für Vermessungswesen* (6), pp. 379–386.
- Hoegner, L., Kumke, H., Meng, L., Stilla, U., 2007. Automatic extraction of textures from infrared image sequences and database integration for 3D building models. *Photogrammetrie - Fernerkundung - Geoinformation* (7), pp. 459–468.
- Kager, H., 2004. Discrepancies between overlapping laser scanner strips—simultaneous fitting of aerial laser scanner strips. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (35), pp. 555–560.
- Kraus, K., 1996. *Photogrammetrie. Verfeinerte Methoden und Anwendungen*, 3rd ed. Dümmler, Bonn.
- Läbe, T., Ellenbeck, K.H., 1996. 3D-wireframe models as ground control points for the automatic exterior orientation. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 218–223.
- Li-Chee-Ming, J., Armenakis, C., 2013. Determination of UAS trajectory in a known environment from FPV Video. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1/W2*, pp. 247–252.
- McGlone, J.C., McKeown, Jr., David M., Dowman, I.J., 1995. Bundle adjustment with object space geometric constraints for site modeling. In: *SPIE Symposium on OE/Aerospace Sensing*, pp. 25–36.
- Nex, F., Remondino, F., 2014. UAV for 3D mapping applications: a review. *Applied Geomatics* 6 (1), pp. 1–15.
- Rottensteiner, F., 2006. Consistent estimation of building parameters considering geometric regularities by soft constraints. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 13–18.
- Spiegel, M., 2007. Kombinierte Ausgleichung der Mars Express HRSC Zeilenbilddaten und des Mars Global Surveyor MOLA DGM. PhD Thesis. German Geodetic Commission, Series C, Volume DGK-C-610, Verl. der Bayer. Akad. d. Wiss., München.
- Strunz, G., 1993. Bildorientierung und Objektrekonstruktion mit Punkten, Linien und Flächen. PhD Thesis. German Geodetic Commission, Series C, Volume DGK-C-408, Verl. der Bayer. Akad. d. Wiss., München.