ASSIGNING TIE POINTS TO A GENERALISED BUILDING MODEL FOR UAS IMAGE ORIENTATION

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:

This paper addresses the integration of a building model into the pose estimation of image sequences. Images are captured by an Unmanned Aerial System (UAS) equipped with a camera flying in between buildings. Two approaches to assign tie points to a generalised building model in object space are presented. A direct approach is based on the distances between the object coordinates of tie points and planes of the building model. An indirect approach first finds planes within the tie point cloud that are subsequently matched to model planes; finally based on these matches, tie points are assigned to model planes. For both cases, the assignments are used in a hybrid bundle adjustment to refine the poses (image orientations). Experimental results for an image sequence demonstrate improvements in comparison to an adjustment without the building model. Differences and limitations of the two approaches for point-plane assignment are discussed - in the experiments they perform similar with respect to estimated standard deviations of tie points.

1. INTRODUCTION

This paper addresses the estimation of the position and attitude, also referred to as pose or exterior orientation, of an Unmanned Aerial System (UAS) considering limitations that occur when using this type of platform. The range of civil applications of UAS is still growing and includes, for example, 3D reconstruction for visualization and planning, monitoring, inspection, cultural heritage, security, search and rescue and logistics. Most applications make use of the UAS technology as a flexible platform and have a need to know where the UAS is situated in relation to objects and often also in a world coordinate system. Due to their low weight and small cost, cameras are often used as sensors to capture the surroundings and to derive the pose relative to objects. However, from camera observations only, the scale of the scene and the pose in a world coordinate system cannot be derived. In addition, purely camera based orientation procedures like structure-from-motion (SFM) are affected by unfavourable error propagation if no loops are closed for single flight strips. To derive poses in a world coordinate system, classical approaches use Ground Control Points (GCPs) for indirect georeferencing and/or perform direct georeferencing based on the observations of Global Navigation Satellite System (GNSS) receivers and Inertial Measurement Units (IMUs). In case of UAS, direct georeferencing capabilities are limited due to payload restrictions and cost considerations. In addition, GNSS signal loss will occur more likely than in classical airborne scenarios, e.g. if the UAS flies through urban canyons. On the other hand, indirect georeferencing using GCPs is often time consuming, and it may be infeasible in certain scenarios.

An alternative is to additionally use an existing generalised building model to improve the pose parameters of the images taken by 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 the bundle adjustment is helpful to compensate inaccurate camera positions measured by GNSS and drift effects of a purely image-based pose estimation. The integration of the building model does not only increase the quality of the pose estimation, but it can also be helpful for applications such as 3D reconstruction for monitoring buildings at different epochs, because the reconstructed tie points for each epoch are related to the same model. Other application examples include the refinement and rendering of generalised building models.

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. In this paper, we expand our previous method (Unger et al., 2016) with respect to the assignment of object points to model planes to overcome errors in the estimated model planes introduced due to wrong assignments. The assignments are found in object space. As an alternative to a simple distance criterion, we assess an indirect approach in two variants that aims at finding planes in the 3D tie point cloud and relate them to model planes based on certain criteria.

The 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 pose estimation. Section 3 introduces our hybrid bundle adjustment method. Section 4 focuses on the generation of fictitious observations with two strategies, whereas Section 5 contains the overall workflow of sensor orientation and sparse scene reconstruction. Experiments using real data to investigate the two strategies are presented in section 6, before we conclude and give an outlook on future work in section 7.

2. RELATED WORK

An overview of UAS applications, platforms and sensors in remote sensing is given by Pajares (2015). Several authors deal with the integration of object knowledge other than GCPs into pose estimation and 3D reconstruction using images and/or other sensors. In general, we distinguish between “soft constraints” and “hard constraints” that are used to model additional knowledge

* Corresponding author

in an adjustment procedure (Rottensteiner, 2006). "Hard constraints" are related to constraints between the unknowns that have to be fulfilled exactly. McGlone et al. (1995) use such constraints to integrate generic knowledge about the captured objects into bundle adjustment. "Soft constraints" are modelled as observation equations for so called fictitious observations with an a-priori standard deviation. This type is used e.g. by Gerke (2011) to constrain image orientation with knowledge about horizontal and vertical lines in the scene.

With respect to the type of a-priori knowledge of a scene, height models or 3D wireframe models e.g. 3D city models are employed frequently. Digital terrain models (DTM) are used to constrain the heights of object points in a hybrid bundle adjustment for satellite imagery by several authors (Strunz, 1993; Heipke et al., 2005; Spiegel, 2007). Lerner et al. (2006) present a method to use ray tracing based on initial pose parameters to directly constrain rays of homologous points of image pairs with a DTM. Hard constraints are used in a robust adjustment that derives image poses for image pairs. Experiments are only carried out using simulated data and images of a miniaturised scene. The idea to constrain the height of object points for nadir images captured by an UAS is found in (Geva et al., 2015). Assuming the pose of the first frame to be known, they also derive surface intersection constraints based on DTM heights that are used to form soft constraints for bundle adjustment. Avbelj et al. (2015) use a digital surface model (DSM) to refine the orientation of hyperspectral images in urban areas. They use building outlines extracted from the DSM to match them to lines in the images using statistical tests. Derived constraints are integrated into a Gauss-Helmert adjustment process.

Läbe and Ellenbeck (1996) use lines to match images to building outlines. Their approach is based on a 3D wireframe model and the aim is to improve the orientation of aerial images in bundle adjustment. Matching lines are found using pose clustering and robust spatial resection to filter outliers. Li-Chee-Ming and Armenakis (2013) match lines found in UAS images to the edges of a Level of Detail 3 (LoD3) building model based on approximate pose parameters. They also apply their approach for indoor datasets and propose a solution to compute the initial pose based on line matches (Li-Chee-Ming and Armenakis, 2017).

Eugster and Nebiker (2009) also use corresponding lines to refine the exterior orientation parameters. Lines are extracted from UAS images and from virtual views of a building model. They use approximate image poses from the measurements of low cost sensors (IMU, barometer, GPS) that result from a direct georeferencing strategy. For the lines they apply relational matching. Orientation parameters are iteratively refined by spatial resection. The procedure is reported to result in an improvement in accuracy of a factor 3 to 4 in comparison to the pure direct georeferencing.

Vysotska and Stachniss (2017) set up a constrained sliding window adjustment to refine laser scans of a ground platform moving in urban areas in 2D. Constraints are found by applying ICP to the scans. Approximate values for the starting pose are needed.

Lothe et al. (2010) address monocular simultaneous localisation and mapping (SLAM) with a camera fixed in a car looking in the driving direction. They present a two-step post-processing method to limit drift effects. The first step is an ICP on sub blocks of the image sequence with a generalised building model ("coarse correction"), the second one consists of a constrained bundle adjustment that refines the image pose parameters ("fine correction").

Our method does not rely on direct matches of points, lines or planes as features, but instead assigns tie points to model planes in object space. Such correspondences lead to soft constraints represented by fictitious observations that are integrated into an overall hybrid bundle adjustment. The adjustment simultaneously refines image poses, object points and the building model. In an iterative process assignments of points to model planes are updated. The expected degree of generalisation of the model is explicitly covered by the a-priori standard deviations of the fictitious observations.

3. ROBUST HYBRID BUNDLE ADJUSTMENT

In our scenario, a UAS is equipped with a low-cost GNSS receiver and a camera that takes images in a multi-view configuration flying in between buildings. In addition, we assume to have knowledge about the scene in the form of a generalised building model represented by its vertices and faces. The topology is given by an ordered list of vertex indices that describe the boundary polygon of each model face. We refer to a model face as *model plane*.

The mathematical representation of the scenario is the one described by Unger et al. (2016), see also Fig. 1: Image coordinates are related to object coordinates and pose parameters using the collinearity equations. Object points observed as homologous points in images are referred to as tie points in this paper. The vertices of the model are another type of object points. They are not explicitly observed in the images. Due to the generalisation of the model, it is possible that they even do not correspond to real object points. Both types of points, tie points and vertices, are related to the building model by assigning them to corresponding model planes. The model planes are parameterised in local plane coordinate systems (x, y, z) that are related to the object coordinate system by six parameters each. These are three rotation angles and a 3D-shift $P0$ from the origin of the object coordinate system to that of the local one for each plane. $P0$ 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. As we want the adjustment to not only affect the reconstructed tie points but also the building model, each plane is parameterised in such a local coordinate system by two angles α, β defining the direction of the normal and a translation δ along the (local) z-axis.

Using this parameterisation, the relation between an object point and a plane is described by its orthogonal distance d to that plane following Eq. 1.

$$d = N(\alpha,\beta)^T \cdot \bar{P}(X,Y,Z) + \delta \quad (1)$$

$\bar{P}(X,Y,Z)$ is the object point expressed in the plane's local coordinate system. $N(\alpha,\beta)^T$ represents the plane normal parameterised by the two angles α,β of the plane.

For the hybrid bundle adjustment, the functional and the stochastic models of Unger et al. (2016) are used. In the stochastic model, we assume a constant a-priori level of accuracy for each observation type and uncorrelated observations.

The following observation types are used: Next to the image coordinates of tie points, there are direct observations for the

projection centres of the cameras, obtained by the low cost GNSS receiver, and direct observations for the object space coordinates of the vertices of the building model. Two groups of fictitious observations relate object space coordinates of a point to planes of the building model using Eq. 1. The first group consists of the fictitious observations related to tie points and the second consists of the ones for the vertices of the building model. Both reflect the assumption that an object point that belongs to a plane should have a distance of zero ($d = 0$) to that plane which leads to the observation equation $v_d = N(\alpha,\beta)^T \cdot \bar{P}(X,Y,Z) + \delta$, with v_d being the residual of the fictitious distance observation. As the model planes are affected by generalisation effects, the real distances of tie points may typically deviate from 0. We allow for only one fictitious observation per tie point, which means we do not model tie points as corner points. In contrast, vertices are assigned to multiple model planes; in this way the topology of the building model is considered in the adjustment.

The aforementioned observations are used as inputs for a Gauss-Markov model to estimate the following unknowns:

- the pose parameters for each image (ω, φ, κ, X_0, Y_0, Z_0)
- the object space coordinates of the tie points $(X,Y,Z)_{TP}$
- three parameters of each model plane (α, β, δ)
- the object space coordinates of the vertices of the building model $(X,Y,Z)_{VT}$

By introducing the plane parameters α, β and δ and the coordinates of the vertices as unknowns into the adjustment, the building model planes can be corrected in case there are tie points that indicate such a change. Estimated changes for model planes are reflected by the parameters α, β and δ. The transformation parameters of the local plane coordinate systems R and P_0 are constants in the adjustment. As the planes are connected by the vertices, the model plane topology is not changed. Vertices will move according to the planes they are connected to. The size of this movement is limited by the a-priori standard deviations of vertex coordinates.

In contrast to our previous work, the hybrid adjustment is made robust w.r.t. gross errors in the observations. Outliers are detected by an iterative reweighting of the observations based on their normalised residuals $\tilde{v}_i = \frac{|v_i|}{\sigma_{l_i}}$, where σ_{l_i} is the a-priori standard deviation of the corresponding observation. It is the goal of reweighting to reduce the weights of observations that have large normalised residuals. We use the iteration scheme recommended by Förstner and Wrobel (2016) which consists of the Huber weight function for the first iterations, then iterates using the exponential weight function to reduce the weight of potential outliers and ends with an adjustment with the original weights after having eliminated the outliers. An observation is identified as an outlier and removed if its normalised residual $\tilde{v}_i$ is larger than three.

We assign tie points to planes of the building model based on their estimated 3D positions, computed in a structure-from-motion pipeline and a first robust adjustment without taking into account the building model. After this bundle adjustment, we make sure, that the two data sets (the 3D point cloud of tie points and the building model) refer to a common coordinate system (see section 5 for details).

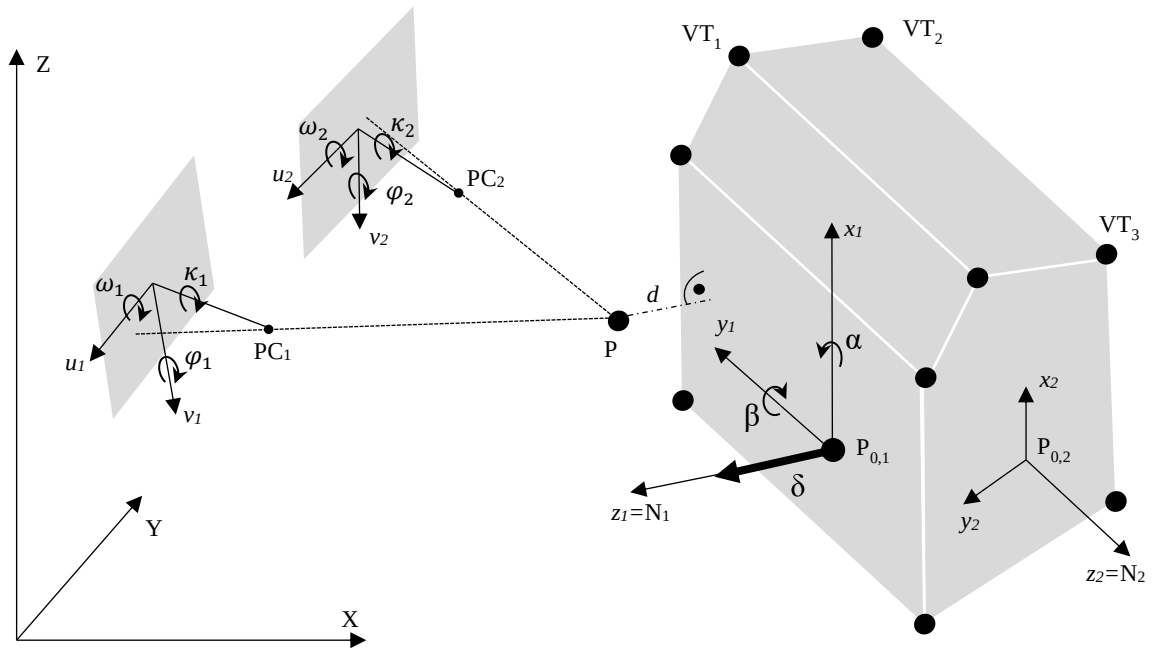

Figure 1: Relevant entities in our scenario (adapted from Unger et al. (2016)): Two poses *i* of the same camera with image coordinate axes (u_i, v_i), projection centres PC_i and three rotation angles (ω_i, φ_i, κ_i) with $i \in \{1,2\}$ represent the multi-view scenario where sensors capture an object point P with world coordinates X, Y, Z. The generalised building model is represented by its vertices VT_k with $k \in \{1,2,\dots\}$ 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 initially 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 (ω_j, φ_j, κ_j) that are not shown in the figure. Two angles α, β and a shift δ (bold arrow) along the local *z*-axis represent the parameterisation of the local plane. The orthogonal distance of an object point *P* to a corresponding plane of the building model is denoted by *d*.

4. ASSIGNMENT OF TIE POINTS TO MODEL PLANES

The focus of this work lies in finding the assignments of tie points to model planes, given the inaccurate image poses and the generalised building model, to generate the first group of fictitious observations.

In this section, two strategies for the assignment of tie points to model planes are discussed. The first one uses a simple distance criterion; the second strategy is based on finding planes in the tie point cloud and matching these planes to the planes of the building model.

The second strategy can be executed in two different variants: the first variant extracts planes from the whole point cloud independent of the building model. Planes thus detected are matched to those of the model. The second variant uses the model planes as input and considers only points in the vicinity of a model plane in order to find a corresponding extracted plane; explicit plane matching is thus avoided.

4.1 Point-to-model matching: The direct approach

In our previous work, the assignment of a point to a plane was based on a simple distance criterion. Tie points were assumed to be related to the closest plane provided that the Euclidean distance from the plane was below a threshold. This threshold was selected in accordance with the accuracy and degree of generalisation of the building model and potential datum problems of the point cloud. It was chosen to be relatively large. As a result, many wrong assignments of tie points to model planes were found that led to wrong results.

We therefore decided to refine this strategy by separating the initial datum effects from those mainly stemming from the generalisation, and solving for the datum in a first processing step (see section 5). Thus now we can use a significantly reduced threshold for the Euclidean distance of tie points to the model planes; it is set in accordance with the size of the generalisation effects we expect to occur. We use the same value as for the a-priori standard deviation of the fictitious observations that relate tie points to model planes.

4.2 Plane-to-model matching: The indirect approach

This is an alternative strategy to first search for planes, referred to as *extracted planes*, in the tie point cloud instead of directly relating tie points to model planes. The aim is to find model planes that correspond to the extracted planes. Based on these plane-to-model correspondences, tie points are assigned to model planes: If a corresponding plane is found, a point belonging to the extracted plane leads to a fictitious observation that relates this tie point to the corresponding model plane.

For both variants mentioned above, planes are detected in the tie point cloud using Maximum Likelihood Estimation SAmple Consensus (MLESAC) (Torr and Zisserman, 2000). MLESAC is a variant of RANSAC that does not just maximise the number of inliers but rather maximises their likelihood. The algorithm requires the maximum allowable distance of points to a plane as a parameter. The proportion of outliers is automatically estimated by the algorithm.

First variant: In the first variant, MLESAC is used on the whole point cloud. We sequentially extract a plane, exclude the inliers with respect to connected components from the given point cloud of 3D tie points and then search for the next plane in the remaining point cloud. This is done until a given number of planes has been found, no more planes are found or the number of points found per plane is repeatedly smaller than a threshold.

We allow for multiple extracted planes per model plane. This reflects the fact that only parts of a larger generalised model plane might be found in the tie point cloud. In order to do so, for each plane thus detected, the related points are projected onto that plane and connected components are found using alpha shapes (Edelsbrunner et al., 1983) with a given radius ϑ_α that defines the maximum distance of points to the shape. Points which are part of the largest connected component are kept as inliers and are then used to determine the plane parameters. The points of the boundary of the connected component in the refined extracted plane define its boundary polygon.

In the next step, correspondences between the extracted planes and the model planes are found. We check each combination of an extracted plane and a model plane for correspondence. For each such pair, we compute the angle between the normal vectors and the orthogonal distance of the centre of gravity (COG) of the boundary points of the extracted plane from the model plane. Both the angle and the distance must be smaller than pre-defined thresholds for the pair to become a candidate for a correspondence. Another criterion to be fulfilled is that the polygon and the model plane must overlap. To check this requirement the COG of the boundary points of the extracted plane is projected to the model plane. If the projection of the COG is outside the boundary polygon of the model plane the candidate is eliminated.

In a last step, tie points of extracted planes are eliminated if their orthogonal projection onto the matching model plane falls outside the model plane's boundary polygon. Tie points inside the boundary of the model plane finally lead to fictitious observations that relate those points to the corresponding model plane.

Second variant: In the second variant, no separate matching step is required, because for each plane a corresponding extracted plane is searched for in the points in its vicinity only. For each model plane, we extract all points having a distance to that plane smaller than a given threshold and a projection onto the model plane that is inside the model plane boundary. The parameters of an extracted plane are then computed using MLESAC. To further reduce the search space, MLESAC is configured to only find planes with normal vectors that do not exceed a given angular distance to the normal of the model plane. Similar to the first variant, only points of the largest connected component and inside the planes boundary polygon are kept and are assumed to correspond to the model plane. These points lead to the fictitious observations relating tie points to model planes.

5. PROCESSING STEPS

Our overall workflow is listed in table 1. The first step is a structure-from-motion pipeline in which homologous points and initial values for image poses and 3D object point coordinates are derived. This is done using an image sequence and, if available, GNSS observations for the projection centre positions of the images as inputs. Subsequently, we run a robust bundle adjustment including the images and GNSS observations without considering the building model (step 2). In this adjustment process, gross errors in the image observations are identified and eliminated. Image poses and tie point coordinates are refined. In

subsequent steps, image observations are not tested for more gross errors. In this way, we separate outliers in the image coordinates of homologous points from those in the fictitious observations of tie points that are found in subsequent steps. Thus, we make sure that wrong fictitious observations do not lead to the elimination of potentially correct image observations. We assume the GNSS observations to be sufficient to define the datum of the image block in these first two steps.

Step 1	Image matching and SFM to derive tie points and image poses.
Step 2	Robust bundle adjustment including only images and, if available, direct observations of projection centres.
Step 3	Initial establishment of relations between tie points and model planes with one of the assignment methods.
Step 4	Hybrid bundle adjustment including the planes, robust with respect to fictitious observations for tie points on model planes.
Step 5	Establishment of relations between tie points and model planes using one of the assignment methods.
Step 6	Final hybrid adjustment.

Table 1: Workflow of pose estimation.

In step 3, the building model is considered for the first time. It must be available in the same coordinate system as the results of step 2. In our workflow this is the GNSS coordinate system, but the datum of the initial image block might be inaccurate due to using the low-cost GNSS receiver. Step 3 has to account for this and consists of the first assignment of tie points to model planes based on their estimated 3D positions from step 2 and the initial plane parameters to set up the fictitious observations. This is done following one of the approaches from section 4. For the direct approach, the threshold for the Euclidean distance of tie points to model planes is set to the size of the expected initial datum differences. For the indirect approaches, the same thresholds as in subsequent steps are used. The known relations of vertices to planes form the second group of fictitious observations which are not changed during subsequent steps.

Steps 4 and 5 are then applied in a repeated way: In step 4, one iteration of the hybrid robust adjustment is carried out. In step 5, tie points are assigned to model planes based on the estimated parameters of step 4 using one of the approaches described in section 4. For the direct approach, the threshold is now set to the a-priori standard deviation of the fictitious observations that relate tie points to model planes. The indirect approaches use the same thresholds as in step 3. The assignments are then used to update the fictitious observations. Steps 4 and 5 are carried out in an alternating fashion until convergence of the adjustment. Note that in step 4 only planes containing more than a pre-defined minimum number of tie points are considered in the adjustment.

Finally, in step 6, fictitious observations identified as outliers are removed and the remaining fictitious observations are kept to run the hybrid adjustment with the original weights to compute the final results.

6. EXPERIMENTS

The sequence for our experiments was captured by a gimbal-stabilised camera attached to our UAS, a manually controlled DJI Matrice 100 quadrocopter, during a flight in between buildings of our campus. The buildings are 4 to 30 m high and the flying height above ground varied between 20 m in the beginning and 2 m at the end of the flight.

We used a Zenmuse X3 camera. It has a fixed focus, 3.61 mm focal length and a 1/2.3" CMOS sensor having 4000x3000 pixels and a pixel size of 1.5 µm. Images were taken automatically every 2 seconds. The image sequence consists of 183 images with an average ground sampling distance of 6 mm/pix. In the process we use direct GNSS observations for the projection centres of the first 30 images only to show that our method can cope with such a configuration.

A 3D city model with Level of Detail 2 (LOD2), freely available for the whole city of Hanover, is used as ground control information. Like the GNSS observations, the model is given in WGS84/UTM Zone 32. The GNSS observations delivered by the copter are pre-processed internally, most probably with a filter that includes IMU and barometric measurements (details of this process are not available).

Image distortion was eliminated prior to processing based on available interior orientation parameters. We used the commercial software package Agisoft PhotoScan[1] for steps 1 and 2, considering only the first 30 GNSS observations for the positions of the projection centres. On average, each tie point was observed in almost six images. The viewing direction is almost orthogonal to the flight direction and the facades. The image coordinates of the homologous points serve as observations and the estimated camera poses and object point coordinates as initial values for our adjustment. We only consider observations of points that are observed in at least three images. Although the adjustment carried out by PhotoScan is robust, we found that some of the exported observations still fall in the group of outliers, since we use stricter constraints. Therefore, we repeat step 2 with our own robust bundle adjustment before going on with the remaining steps.

The a-priori standard deviations of the observation types that form the stochastic model are set as follows:

Image coordinates	σ_{img}	± 0.75 pix
GNSS obs. of projection centres	σ_{GNSS}	± 3 m
Building model vertices	σ_{VT}	± 0.5 m
Fictitious distance for tie points	$\sigma_{d_{TP}}$	± 0.3 m
Fictitious distance for vertices	$\sigma_{d_{VT}}$	± 0.01 m

σ_{VT} reflects the accuracy and generalisation effects of the vertices of the building model. $\sigma_{d_{TP}}$ describes the deviation of the model planes due to the generalisation.

In the iterations of the hybrid adjustment (steps 4 and 5), for the direct approach of tie point assignment, we choose to take into account fictitious distances for points to planes only if the distance is smaller than 0.3 m. As described before, this threshold is chosen in accordance with the expected degree of generalisation of the model $\sigma_{d_{TP}}$. In step 3, the initialisation, we use a higher threshold of 2 m, because of the low GNSS accuracy to obtain as many correct assignments as possible with some outliers only.

[1] http://www.agisoft.com/

For the indirect approach of tie point assignment, the following parameters are set. The distance for points to planes found using MLESAC is again set to 0.3 m according to $\sigma_{d_{TP}}$. We use this value to be left with as few wrong fictitious observations as possible accepting, that only 68% of the points of a plane w.r.t. $\sigma_{d_{TP}}$ might be found. The radius to find connected components within extracted planes is set to $\vartheta_\alpha = 3$ m to allow for some larger plane regions with few tie points. The choice of ϑ_α is not critical as points outside the boundary polygon of model planes are rejected later. The maximum angle between normal directions of extracted and model planes is set to 15°. We thus allow for some differences in case of inaccurate initial values and especially for small planes or in case only small parts of planes are extracted. For the indirect approach without the reduced search space, the maximum distance of an extracted plane's COG to a model plane is chosen to be 1 m. For the variant with reduced search space, the maximum distance to consider points for extracting a corresponding plane also is set to 1 m. Both distance thresholds are used also in the initialisation phase (step 3), which is sufficient for our test data. With the 1 m distance at least some corresponding plane points are found even with inaccurate initial values. This distance is assumed to be large enough to find planes that are inaccurate in the model and have to be corrected within the adjustment based on tie point observations. For both approaches, planes are used only if at least 15 points are assigned to them. In our experience, the pose of planes described by fewer tie points often is not stable.

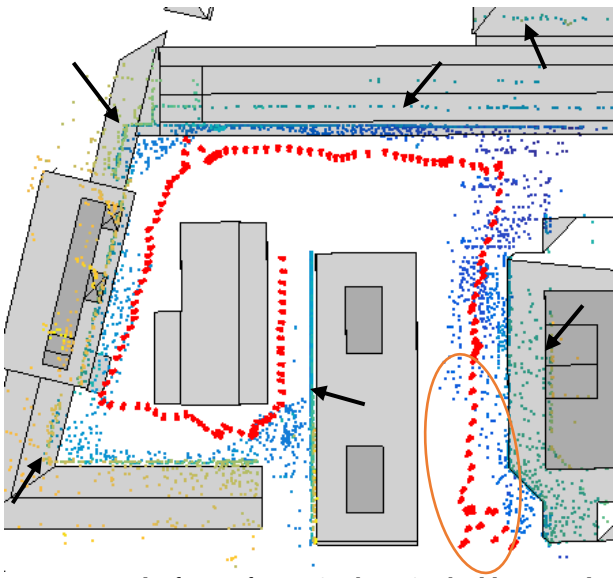

Figure 2: Result of step 2 for a 150 m by 120 m building complex: LOD2 model (grey), camera poses (red), tie points color coded by height from blue (low) to yellow (high). Differences of the tie point cloud to the model are highlighted by black arrows. GNSS was used only within the orange ellipse.

Figure 2 shows the captured scene with the 183 camera poses, 5701 tie points after step 2 and the generalised building model. GNSS for the projection centres was used only for the first 30 images in the beginning of the image sequence (lower right corner, orange ellipse). Shift, rotation and scale differences of the tie point cloud to the building model are highlighted with red arrows: Tie points forming vertical planes are off the model planes they should correspond with.

Table 2 shows some statistics of the robust adjustment (step 2) with just the images and GNSS observations (1. column) and the robust hybrid adjustment with the direct (2. column) and both indirect approaches (3. and 4. column; 4. column with reduced search space). With respect to the estimated mean standard deviations of the tie points $\hat{\sigma}_{TP}$, the integration of the building model in all variants leads to an improvement from meter range to about a decimetre. As planes are found for at least some tie points of each image, the $\hat{\sigma}_{TP}$ values are not split into tie points with and without a fictitious observation. There is almost no difference in the $\hat{\sigma}_{TP}$ as the whole image block profits from the planes. The numbers of found fictitious observations for tie points per approach show that the direct approach finds the most assignments. The indirect approach that searches the whole tie point cloud can only assign points to planes if it is able to extract the candidate plane. As not all planes are found by the MLESAC, the approach yields the lowest number of fictitious observations. Also, the indirect approach with reduced search space finds less points then the direct approach. The values for $\hat{\sigma}_{TP}$ for all three approaches are virtually identical.

Approach	Images, GNSS	Direct	Indirect	Indirect, Reduced
#images	183	183	183	183
$\hat{\sigma}_0$	0,70	0.67	0.68	0.68
$\hat{\sigma}_{TP_X}$ [m]	3.10	0.10	0.10	0.09
$\hat{\sigma}_{TP_Y}$ [m]	3.33	0.09	0.10	0.09
$\hat{\sigma}_{TP_Z}$ [m]	4.16	0.08	0.10	0.08
#fict. obs. tie pts	0	3431	3026	3256
#object points	5701	6423	6423	6423
#observations	67485	74630	74225	74455
#unknowns	18201	21192	21192	21192
#outliers	487	487	487	487

Table 2: Results of the (hybrid) adjustment on the test sequence.

The number of outliers does not change after the adjustment of step 2. This means that no detectable outliers were present in the fictitious observations of any of the approaches. Potentially remaining wrong assignments can thus not be detected.

Figure 3 shows the resulting tie points and the estimated building model for the direct and for the indirect (reduced) approach. Points with the same colour belong to the same plane, while grey points are not assigned to any plane. Both results show no visible offsets like in figure 2. The initial building model is not shown as the differences can only be seen in higher zoom levels. Orange ellipses highlight planes that were found only using the indirect approach with reduced search space.

The indirect approach without reduced search space, which is not shown in the plot, finds fewer planes then the direct approach. The missed planes are either small or the points are scattered and are therefore not considered as part of the largest connected component. While fewer correspondences are found, the remaining correspondences are more likely correct as they must be part of a plane and its largest connected component. This means that isolated points are not accepted to correspond to a plane. Even if they are close to a plane, they have a higher probability to belong to structures not represented in the model in comparison to points which are part of a connected plane component. This behaviour can also be seen for the indirect approach with reduced search space as shown in figure 3, e.g. in the areas highlighted by black ellipses.

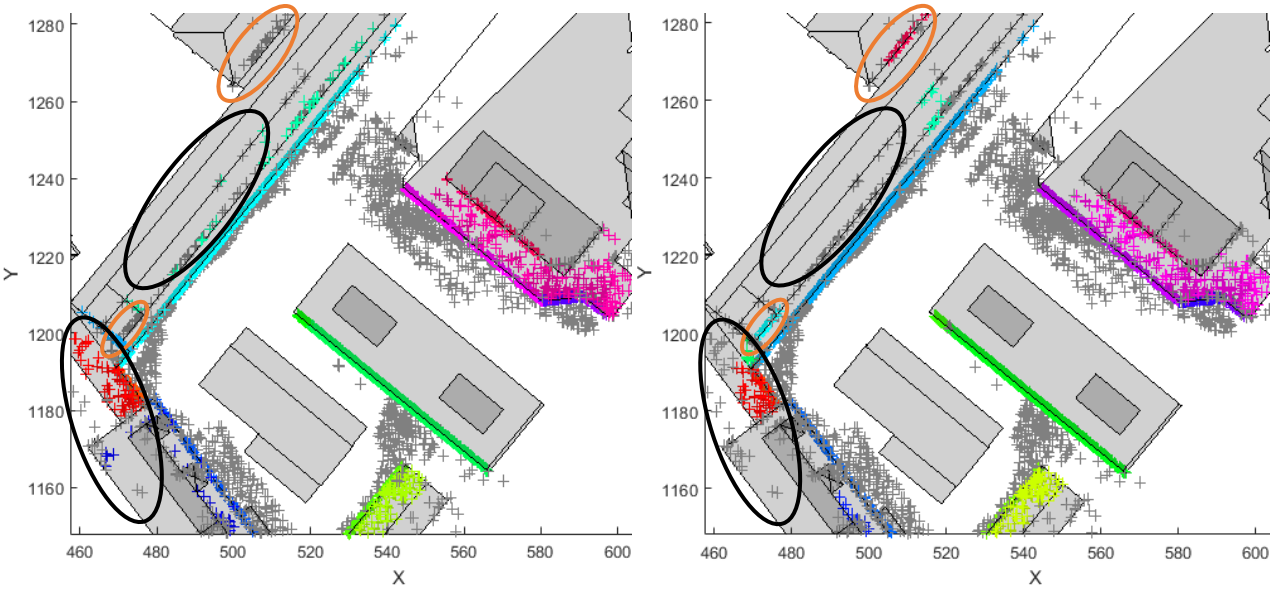

Figure 3: Estimated tie points of the hybrid adjustment with the direct (left) and the indirect (reduced) approach (right) plotted with the estimated building models in top view. Points with the same color were assigned to the same model plane. Tie points that are not matched to a plane are shown in grey. Orange ellipses highlight planes, that were found only using the indirect approach. Black ellipses denote areas, where correspondences were rejected by the indirect approach.

Another advantage of the approach without reducing the search space is that points which are part of extracted planes that were not matched to any model plane are not assigned to the model. This is shown in figure 4, where a façade is represented as one plane in the model. In reality, it consists of a roof above doors and a ramp used for deliveries. These are detected as planes and points assigned to these planes are not assigned to any model plane. The figure also shows that the estimated model plane is hardly affected by rotations due to wrong assignments of, e.g., ground points.

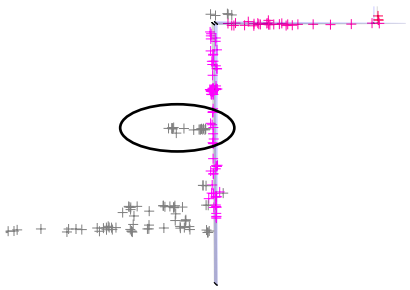

Figure 4: Side view on a slice of a building shown in the initial (blue) and adjusted model (grey). Structures not represented by the model are well covered with tie points (e.g. the roof plane highlighted by the ellipse) and are not assigned to the model plane. The initial and the estimated model planes have very similar parameters.

7. CONCLUSION AND FUTURE WORK

The presented method improves the pose estimation of an image sequence captured by an UAS by integrating a generalised building model into the adjustment. The integration is done based on fictitious observations that are found using a direct approach and two variants of an indirect method. The direct approach uses only a distance criterion, while the indirect approach extracts planes in the tie point cloud and matches them to the building model to generate fictitious observations. The pose estimation is carried out in a hybrid adjustment in which outliers in the homologous points and in the fictitious observations are handled separately.

Our experiments show that both approaches for generating the fictitious observations lead to an improvement of estimated tie points w.r.t. their estimated standard deviations. Significant differences between the different results were not found.

One advantage of the plane-to-model matching is that it finds planar structures not represented in the model. These might be helpful to generate additional tie features in future. In addition, points close to these planes can be assumed to not belong to a close by model plane.

In our future work, we will evaluate the method on longer image sequences and with different facades. To further analyse the method, we will compare the estimated tie points and the estimated model to reference data e.g. using independent check points, a more detailed and accurate point cloud or a building model with higher level of detail. Also, we plan to analyse combinations of the assignment methods.

In addition, the stochastic model will be refined to separate systematic errors of the GNSS receiver from smaller random errors. This will be done by introducing global shifts and rotations that cover a systematic offset of the GNSS. Its relative accuracy can then be assumed to be more accurate than the one used in our current stochastic model.

Finally, while in this paper, we describe the method as a post-processing step, we see the potential to apply it incrementally in a real-time or near-real-time scenario which is needed e.g. for augmented reality or search and rescue applications.

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.

Edelsbrunner, H., Kirckpatrick, D.G., Seidel, R., 1983. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory* (29), pp. 551–559.

Eugster, H., Nebiker, S., 2009. Real-time Georegistration of Video Streams from Mini or Micro UAS using Digital 3d City Models. *6th International Symposium on Mobile Mapping Technology*.

Förstner, W., Wrobel, B.P., 2016. Photogrammetric Computer Vision. Springer International Publishing, Cham, Switzerland.

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.

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. *zfv* (6), pp. 379–386.

Läbe, T., Ellenbeck, K.H., 1996. 3D-wireframe models as ground control points for the automatic exterior orientation. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXI-B2, pp. 218–223.

Lerner, R., Rivlin, E., Rotstein, H.P., 2006. Pose and motion recovery from feature correspondences and a digital terrain map. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (9), pp. 1404–1417.

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.

Li-Chee-Ming, J., Armenakis, C., 2017. Matching real and synthetic panoramic images using a variant of geometric hashing. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-1/W1, pp. 199–206.

Lothe, P., Bourgeois, S., Dekeyser, F., Royer, E., Dhome, M., 2010. Monocular SLAM Reconstructions and 3D City Models: Towards a Deep Consistency. In: Ranchordas, A., Pereira, J.M., Araújo, H.J., Tavares, J.M.R.S. (Eds.), *Computer Vision, Imaging and Computer Graphics. Theory and Applications*. Springer, Heidelberg, pp. 201–214.

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.

Pajares, G., 2015. Overview and Current Status of Remote Sensing Applications Based on Unmanned Aerial Vehicles (UAVs). *Photogrammetric Engineering & Remote Sensing* 81 (4), pp. 281–329.

Rottensteiner, F., 2006. Consistent estimation of building parameters considering geometric regularities by soft constraints. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXIV - 3, 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.

Torr, P., Zisserman, A., 2000. MLESAC. A New Robust Estimator with Application to Estimating Image Geometry. *Computer Vision and Image Understanding* 78 (1), pp. 138–156.

Unger, J., Rottensteiner, F., Heipke, C., 2016. Integration of a generalised building model into the pose estimation of UAS images. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B1, pp. 1057–1064.

Vysotska, O., Stachniss, C., 2017. Improving SLAM by Exploiting Building Information from Publicly Available Maps and Localization Priors. *PFG - Journal of Photogrammetry, Remote Sensing and Geoinformation Science* (85), pp. 53–65.