

Nr. 361

Jakob Unger

Integrated Estimation of UAV Image Orientation
with a Generalised Building Model

Integrated Estimation of UAV Image Orientation with a Generalised Building Model

Von der Fakultät für Bauingenieurwesen und Geodäsie
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades

Doktor-Ingenieur (Dr.-Ing.)

genehmigte Dissertation von

Jakob Daniel Unger, M. Sc.

Prüfungskommission:

Vorsitzender: Prof. Dr.-Ing. Jürgen Müller

Referent: Prof. Dr.-Ing. Christian Heipke

Korreferenten: Prof. Dr.-Ing. Franz Rottensteiner

Prof. Dr.-Ing. Ingo Neumann

Prof. Dr.-Ing. Markus Gerke

Tag der mündlichen Prüfung: 12.06.2020

HANNOVER 2020

Statement

I state

- that I know and have complied with the rules of the doctoral regulations and that I agree to an examination according to the terms of the doctoral regulations,
- to have written the dissertation myself, not to have taken over any text sections from third parties or own examination papers without marking and to have indicated all the sources I used in my work,
- not to have provided third parties, either directly or indirectly, with monetary value for mediation activities or for the preparation of the content of the dissertation,
- not yet to have submitted the dissertation as an examination paper for a state or other scientific examination,
- not to have submitted the same or an essentially similar work to any other university as a dissertation, as well as not to have submitted any other work as a dissertation.

Hannover, 31st March 2020

Abstract

The estimation of position and attitude of a camera, addressed as image orientation in photogrammetry, is an important task to obtain information on where a platform is located in the world or relative to objects. Unmanned aerial vehicles (UAV) as an increasingly popular platform led to new applications, some of which involve low flight altitudes and specific requirements such as low weight and low cost of sensors. Image orientation needs additional information to retrieve not only relative measurements but position and attitude in a world coordinate system. Given the requirements on sensors and especially for flights in between obstacles in urban environments classically used information of Global Navigation Satellite Systems (GNSS) and Inertial Measurement Units (IMU) or specially marked ground control points (GCP) are often inaccurate or unavailable.

The idea addressed within this thesis is to improve the UAV image orientation based on an existing generalised building model. Such models are increasingly available and provide ground control that is helpful to compensate inaccurate or unavailable camera positions measured by GNSS and drift effects of image orientation. Typically, for UAV applications in street corridors, the geometric accuracy and the level of detail of such models is low compared to the high accuracy and high geometric resolution of the image measurements. Therefore, although the building model differs from the observed scene due to its generalisation, relations of the photogrammetric measurements to the building model are formulated and used in the determination of image orientation.

Three approaches to assign tie points to model planes in object space are presented, and a sliding window as well as a global hybrid bundle adjustment are set up for image orientation aided by a generalised building model. The assignments lead to fictitious observations of the distance of tie points to model planes and are iteratively refined by bundle adjustment. Experiments with an image sequence captured flying between buildings show an improvement of image orientation from the metre range with purely GNSS measurements to the decimetre range when using the generalised building model with the simplest assignment method based on point-to-plane distances. No improvement by searching planes in the tie point cloud to indirectly find the relations of tie points to model planes is observed. The results are compared to a building model of higher detail and systematic effects are investigated.

In summary, the developed method is found to significantly improve UAV image orientation using a generalised building model successfully.

Kurzfassung

Die Schätzung von Position und Lage einer Kamera, die in der Photogrammetrie als Bildorientierung bezeichnet wird, ist eine grundlegende Aufgabe, um Informationen darüber zu erhalten, wo sich eine Plattform in der Welt oder relativ zu Objekten befindet. Zunehmend führen unbemannte Luftfahrtsysteme (UAV) als Plattform zu neuen Anwendungen, die zum Teil geringe Flughöhen und spezifische Anforderungen wie Gewicht und Kosten der Sensoren mit sich bringen. Für die Bildorientierung werden zusätzliche Informationen benötigt, um nicht nur relative Messungen, sondern auch Position und Lage in einem Weltkoordinatensystem bestimmen zu können. Angesichts dieser Anforderungen und insbesondere für Flüge zwischen Hindernissen in städtischen Gebieten sind die klassisch verwendeten Informationen von Navigationssatelliten- (GNSS) und Intertialmesssystemen (IMU) oder auch speziell markierten Passpunkten (GCP) oft nicht verfügbar oder zu ungenau.

Die hier behandelte Idee ist daher, die Bildorientierung von UAVs auf der Grundlage eines bestehenden generalisierten Gebäudemodells zu verbessern. Solche Modelle sind in zunehmendem Maße verfügbar und bieten eine Möglichkeit, ungenaue oder nicht verfügbare GNSS-Kamerapositionen und Drifteffekte der Bildorientierung zu kompensieren. Bei UAV-Befliegungen in Straßenschluchten sind die geometrische Genauigkeit und der Detaillierungsgrad solcher Modelle im Vergleich zur hohen Genauigkeit und hohen geometrischen Auflösung der Bildmessungen typischerweise gering. Obwohl das Modell also aufgrund seiner Generalisierung von der beobachteten Szene abweicht, können Beziehungen der photogrammetrischen Messungen zum Gebäudemodell formuliert und in der Bildorientierung verwendet werden.

Es werden drei Ansätze zur Zuordnung von Verknüpfungspunkten zu Modellebenen im Objektraum sowie eine hybride Bündelausgleichung zur Bildorientierung mit Hilfe eines generalisierten Gebäudemodells, die global oder fensterbasiert abläuft, vorgestellt. Die Zuordnungen führen zu fiktiven Beobachtungen für den Abstand von Verknüpfungspunkten zu Modellebenen und werden während der iterativen Bündelausgleichung verfeinert. Experimente mit einer zwischen Gebäuden aufgenommenen Bildsequenz zeigen eine Verbesserung der Bildorientierung vom Meterbereich rein mit GNSS-Messungen in den Dezimeterbereich bei Verwendung des generalisierten Gebäudemodells mit der einfachsten Zuordnungsmethode auf Basis von Punkt-zu-Ebene-Distanzen. Eine Verbesserung der Punkt-zu-Ebene-Zuordnungen durch die Suche von Ebenen in der Punktwolke wird nicht beobachtet.

Zusammenfassend lässt sich sagen, dass die entwickelte Methode die UAV-Bildorientierung mit Hilfe eines generalisierten Gebäudemodells signifikant verbessert.

Contents

1	Introduction	7
1.1	Motivation	8
1.2	Problem Statement and Contributions.....	9
1.3	Structure	10
2	State-of-the-art	11
2.1	Integration of Object Knowledge in Image Space	12
2.2	Integration of Object Knowledge in Object Space.....	16
2.3	Discussion	20
3	Photogrammetric Pose Estimation with a Generalised Building Model.....	23
3.1	Overview	23
3.2	Hybrid Bundle Adjustment	25
3.2.1	Modelling Relations of Object Points to Model Planes	27
3.2.2	Functional Model	29
3.2.3	Stochastic Model	30
3.2.4	Robust Estimation	31
3.2.5	Determination of Initial Values.....	32
3.3	Workflow	33
3.3.1	Global Adjustment	33
3.3.2	Sliding Window Adjustment.....	33
4	Assignment Under Generalisation Effects	37
4.1	Generalisation Effects	38
4.2	Direct Assignment: Point-Plane-Matching	39
4.3	Indirect Assignment: Plane-Plane-Matching	40
4.3.1	Indirect Assignment without ROIs.....	41
4.3.2	Indirect Assignment with ROIs.....	43
4.4	Summary of the Assignment Parameters	44

5	Experiments.....	47
5.1	Setup of the experiments	47
5.1.1	Scenarios	48
5.1.2	Sequences	50
5.1.3	Evaluation.....	51
5.1.4	Structure of the Experiments.....	52
5.2	Dataset	53
5.2.1	Hardware	53
5.2.2	Data	54
5.3	Parameter Settings and Implementation.....	59
6	Results and Discussion.....	63
6.1	The Short Sequence: Generalisation & Systematic Effects	64
6.2	The Long Sequence: Generalisation & Systematic Effects, Block Deformations ...	71
6.3	Check Point Errors versus Estimated Standard Deviations	77
6.4	Sliding Window versus Global Adjustment.....	80
6.5	Assignment Strategies	82
6.6	The Full Sequence	89
6.7	Parameter Variation.....	93
6.7.1	Fictitious Distance Observations of Tie Points	94
6.7.2	Maximum Distance of Tie Points to Model Planes.....	95
6.7.3	Estimation of Vertex Coordinates	96
6.7.4	Window Size N_{ws} and Overlap N_{new}	97
7	Conclusion and Outlook.....	99
	References	103

1 Introduction

The estimation of position and attitude of a platform is an essential task in many areas such as robotics, autonomous driving, aviation, surveying and also for mobile phones. Position and attitude are also referred to as pose or exterior orientation and typically are described by six parameters: The 3D coordinates of the origin of the platform coordinate system and a 3D rotation that describes the attitude of the platform. The retrieval of these parameters is addressed as pose estimation. The goal is to gather information on where a platform is located in the world or relative to objects. Exploiting this information is also essential for unmanned aerial vehicles (UAV) which are used as a flexible platform in a broad range of civil applications. Applications of UAVs include 3D reconstruction for visualisation and planning, monitoring, inspection, cultural heritage, agriculture, search and rescue, security and logistics.

While the retrieval of pose information has been studied for a long time and applied in aviation and for example in aerial surveying, specific limitations occur when using UAVs. Requirements concerning low weight, low power consumption and low cost limit the choice of sensors. In addition, new application scenarios with new challenges for pose estimation such as the possibility of very low flying heights in between obstacles arise.

An approach to pose estimation often used in aerial applications but also on ground platforms is the direct orientation using Global Navigation Satellite Systems (GNSS) and Inertial Measurement Units (IMU). As a result of the cost considerations and limited payload capability of UAVs and consequently of inferior sensor quality, and also due to higher flight dynamics, such directly measured data for the pose are typically not accurate enough for precise positioning. As an alternative to direct orientation or in combination with it, optical sensors, e.g. a camera or a laser scanner, can be used to observe ground control information such as Ground Control Points (GCP) to derive the platform pose indirectly. This indirect orientation is based on the knowledge about the position of the control information.

The characteristics of GNSS and GCPs lead to limitations in the pose estimation for UAVs. As GNSS depend on the visibility of satellites, in contrast to aerial applications, problems comparable to those of ground platforms like in mobile mapping or autonomous driving occur: If the UAV flies at low altitude in urban canyons the buildings can limit or completely block satellite signals. Considering control information, GCPs have to be marked and defined in a coordinate frame, and their use is therefore infeasible for many scenarios typically due to cost or time considerations.

Due to their low weight and low cost, cameras are often used as sensors to capture the surroundings. A camera on a UAV can be considered as an instrument to derive the pose relative to objects in its field of view. However, from camera observations only, the scale of the scene and the pose in a world coordinate system cannot be derived without the aid of additional sensors or ground control information. In addition, purely camera-based orientation procedures like visual odometry are affected by unfavourable error propagation leading e.g. to block deformations if loops are not closed for single flight strips. The orientation of images relative to each other one after the other in a sequence suffers from the accumulation of errors from uncertain image coordinates of homologous points. This type of error propagation is addressed as “drift” and is also well known from, e.g., wheel odometry in robotics or the tachymetric measurement of open traverses in surveying.

1.1 Motivation

Many scenarios face the problems mentioned above, GNSS signal loss, inaccurate measurements from low-cost GNSS/IMU sensors, drift effects and unavailable GCPs. As an example, the following scenario requires these problems to be addressed: a flight of a UAV equipped with a camera in an urban environment from a starting point to a target location, e.g. to make a delivery or carry out measurements. Typically, there will be no ground control information available, GNSS signal quality will be low and the reduction of drift effects by flying loops is undesirable. Given the scale and the start pose of the UAV and the target in the same coordinate frame, relative measurements of the camera could be used in principle to navigate from start to target. If there were no drift effects, this navigation would not require additional sensors or specially placed ground control information. The idea addressed within this thesis is to reduce drift effects and improve the UAV pose based on an existing generalised building model. Here, generalisation of the building model typically refers to generalisation during data acquisition, e.g. of building footprints for cadastre that are used to derive the building model. Whereas both the geometric accuracy and the level of detail of such models may be limited, the integration of this information into the pose optimisation is helpful to compensate inaccurate camera positions measured by GNSS and drift effects of purely image-based pose estimation.

More and more generalised building models are produced for cities all over the world. For instance, Germany is on the way to store all buildings at a certain level in a generalised building model that gradually becomes free of charge and is relatively small in terms of storage needed. Therefore, such models provide a promising source to improve pose estimation if neither GCPs nor GNSS measurements are reliable or available. The result of a bundle adjustment that integrates a generalised building model is not limited to an improved pose but further supports applications that make use of the reconstructed tie points. An example is the 3D reconstruction of buildings in several epochs for monitoring purposes, in which the epochs can only be compared if the results are available in the same coordinate system or are related to the same

model. Another possibility is the refinement and rendering of the building model based on the captured data.

1.2 Problem Statement and Contributions

Two central problems that result from the task of integrating a generalised building model into camera-based pose estimation are addressed in this thesis:

- Can relations of tie points to building models be found although the model deviates from the observed scene due to the generalisation of the model?
- How can found relations be considered in pose estimation despite the deviations of the simplified model from the observed scene and do they improve the estimated poses?

The first problem leads to the objective to find suitable features to relate the photogrammetric measurements and the building model to each other. The second problem tackles the mathematical description of these relations, taking the deviations due to generalisation effects into account, and the integration of these found relations between images and building model into pose estimation.

Bundle adjustment as the standard photogrammetric procedure to retrieve image poses and object points is chosen as the basis. A hybrid bundle adjustment is set up that integrates a generalised building model by fictitious observations. These fictitious observations require object points to lie on building model primitives and are considered by soft constraints. Two types of points in object space are used: Tie points are the object points reconstructed from observed homologous points in the images. The second type of object points are vertices, i.e. the 3D corner points given by the building model.

While the vertices' relation to model planes is given by the model itself, the relation of tie points to model primitives has to be found. Three strategies to relate reconstructed tie points to planes as primitives of the building model are introduced and evaluated. Instead of directly matching features observed in the images, e.g. points or lines, with features of the model, the assignments are found in object space. The reason for this approach is the generalisation of the model: Its representation differs from the observed scene. Vertices or edges of the generalised building model do not have to correspond to vertices and edges present in the real scene. Generalisation leads to simplification and aggregation of model structures that, in reality, are more detailed and might consist of several vertices and edges.

The process of finding the relations of tie points to model primitives is addressed as "assignment" within this thesis. Tie points are assigned to planes as primitives of a building model. Differences between planes derived from tie points and model planes due to generalisation effects of the model are taken into account during the assignment and in the hybrid bundle adjustment. Tie points found to be related to a model plane add a soft constraint in the form of a fictitious observation with a certain standard deviation. In the assignment, thresholds on point-to-plane distances are used, which accounts for the generalisation effects.

Assignment takes place in object space. The first assignment method uses a simple distance criterion to directly relate tie points to model planes. The second and third methods indirectly relate tie points to model planes: They aim at finding individual planes in the tie point cloud and relate them to model planes based on certain criteria. In the second method, planes are found in the tie point cloud independently from the building model. In the third method, planes are only searched for in a subset of the tie points that is inside a region of interest (ROI) defined by each existing model plane.

In difference to the classical bundle adjustment as a post-processing step where all poses are optimised after the flight, a sliding window hybrid bundle adjustment is set up to allow for pose estimation during data capturing, e.g. to correct poses already during a UAV flight. In contrast to a filter approach, in which only the current pose is estimated, sliding window approaches estimate a “window” of several poses until the current pose and “slide” this window along the sequence of poses. The concept of sliding windows is realised as a sequence of hybrid bundle adjustments along the image sequence and the building model.

1.3 Structure

Chapter 2 analyses existing work related to the addressed problems. Then, the proposed method to integrate a generalised building model into photogrammetric pose estimation is described. Chapter 3 introduces the relevant entities of the scenario and describes the setup of a hybrid bundle adjustment with workflows for global and sliding window optimisation. In chapter 4, the focus is set on the assignment of 3D points to generalised model planes. Experiments based on different scenarios calculated with a captured image sequence are set up to investigate the performance of the developed method (chapter 5). The scenarios start with a best-case where a detailed building model is available, proceed with the generalised building model and end with a worst case where no building model is used. The results of the experiments are presented and discussed in chapter 6, structured by subsets of the captured image sequence for the analysed scenarios. The outcome of the thesis is summarised in chapter 7, which ends with an outlook on future work.

2 State-of-the-art

Trajectory estimation using a single camera is a fundamental problem in photogrammetry as well as in the fields of robotics and computer vision. The trajectory regularly is described by the parameters of the exterior orientation, the pose. Whereas approaches exist to estimate continuous trajectories, e.g. for satellite line scanners, this thesis aims at the orientation for frame camera images of an image sequence and uses the term *trajectory* to address sequences of image poses. The standard method for estimating the pose parameters as well as a typically sparse representation of the reconstructed scene in the form of tie points with their object coordinates is bundle adjustment. Traditionally, bundle adjustment is applied offline to blocks of captured images to estimate the parameters given image observations in post-processing. For online pose estimation, sequential methods have been developed that range from filter approaches, which predict and estimate only the current pose, to so-called smoothing or sliding window approaches that reduce the optimisation to only the most recent poses. These approaches are used in robotics to solve problems like visual odometry, where a camera is used for online trajectory estimation by dead reckoning, or simultaneous localisation and mapping (SLAM). The term *visual odometry* often used in robotics and computer vision corresponds to *image orientation* and *reconstruction* used in photogrammetry to address the task of retrieving the pose parameters of images and the retrieval of object coordinates of tie points, respectively. In SLAM, image orientation and 3D reconstruction are carried out simultaneously to retrieve camera poses and a map of the environment. Förstner and Wrobel (2016) give a comprehensive insight and details of the statistics and geometry for image orientation and reconstruction from both a photogrammetric and computer vision point of view. Strasdat et al. (2012) discuss differences in filtering and sliding window approaches for visual SLAM. In this thesis the classical offline bundle adjustment is used first. Then, it is extended to work in a sliding window manner; filtering is not addressed and remains for future work on improving processing speed while sacrificing geometrical accuracy of the hybrid optimisation.

Bundle adjustment typically does not only minimise residuals of image measurements to estimate the parameters of the exterior orientation and tie point 3D coordinates. To solve for the missing scale and to add absolute datum information, additional observations and unknowns must be considered. Bundle adjustment typically integrates other sensor measurements, e.g. from GNSS/IMU, and/or additional information, like GCP coordinates, constraints on planarity, parallelism, and so on. In this thesis, additional information comes from a building model of the captured scene. In general, one can distinguish between *soft constraints* and *hard constraints* that are used to model additional knowledge in an adjustment procedure (e.g. Rottensteiner, 2006). Hard constraints refer to constraints between the unknowns that have to be fulfilled exactly. McGlone et al. (1995) provide the generic mathematical framework to

integrate constraints into bundle adjustment: Soft constraints are set up as observation equations for so-called fictitious observations with an a priori standard deviation. Therefore, the constraints must not be fulfilled exactly but rather their standard deviation sets their strictness. Also, the residuals of these observations can be analysed to detect wrong constraints as outliers.

Camera-based pose estimation is addressed in different fields of research and application for various platforms from indoor robots over self-driving cars and flying platforms to satellites. This work aims at UAVs as a platform that recently received growing attention. Colomina and Molina (2014), Pajares (2015) and Nex and Remondino (2014) give an overview of UAV applications, platforms and sensors in photogrammetry. This thesis aims at a scenario where a UAV operates in an area not fully covered by other platforms: Images are captured from low altitude looking at building façades while the platform flies between the buildings. This scenario leads to the aforementioned problems in pose estimation.

In the following, existing approaches to pose estimation in conjunction with additional object knowledge are investigated, and open problems for the scenario of this thesis are identified. First, approaches are described, which integrate object knowledge by finding correspondences between features from image space and object knowledge (section 2.1). Then, work is analysed in which correspondences between image features and object knowledge are established in 3D object space (section 2.2). The related work in the sections is structured according to the type of object information, type of observations, application scenarios and methods to match measurements to object information. Tables give an overview of the related work. The chapter ends with a discussion of the related work in the context of this thesis.

The focus of this literature review is on geometry rather than on radiometry, brightness or methods that require visual object knowledge like feature descriptors or texture information, e.g. street view imagery or textured building models.

2.1 Integration of Object Knowledge in Image Space

The integration of a priori known information about the captured objects, other than GCPs, is addressed by many authors by using features like points or lines that are detected in captured images and are matched to the same features known from object information. Often, known object information is projected into 2D image space to find corresponding features observed by a camera. The methods described in the following are listed as an overview in Table 1.

A basic case of integrating object knowledge is the consideration of general information about captured objects to retrieve constraints, e.g. knowledge about parallelism, orthogonality, right angles or vertical and horizontal lines. Gerke (2011) presents such an approach: The knowledge that building outlines detected in oblique aerial images in urban areas feature vertical and horizontal lines and right angles is integrated into camera pose estimation and self-calibration by additional fictitious observations as soft constraints.

Elevation models and building models provide knowledge about a captured scene. Avbelj et al. (2015) address the orientation refinement of hyperspectral aerial images using a digital surface model (DSM). They extract building outlines from the DSM and match them to lines detected in an image. Line pairs are statistically tested and counted in an accumulator for a 2D shift and a rotation between the DSM and the images projected onto the same plane. Line pairs of the accumulator cell with the highest number of correspondences are used to derive constraints for a Gauss-Helmert adjustment of the parameters of an affine transformation between the projected DSM and the image.

Several authors consider building models as object knowledge that is projected into image space to match features like points and lines. Läbe and Ellenbeck (1996) match lines between an aerial image and a projected wireframe building model using pose clustering and robust spatial resection to filter outliers. Their goal is to enable the use of such models as ground control feature. Hsu et al. (2000) alternate pose prediction and model-based orientation refinement. Their pose prediction is based on tracked image feature points in frames of a video. For the pose refinement, a cost-function is minimised that compares projected model lines to oriented gradient images derived from the images.

Frueh et al. (2004) align an oblique aerial image to a building model to texture the model. Exhaustive search is used to adjust the image pose parameters and the focal length by rating matches of lines. Lines are projected from the building model based on GNSS and IMU measurements for the initial image pose and compared to lines extracted from the image. Texturing of a building model is also the goal of Hoegner et al. (2007), but they use a terrestrial infrared image sequence and relate corner points calculated from horizontal and vertical lines detected in the images to vertices of the building model. The exterior orientation of an image is then calculated based on three 2D/3D corner point pairs. The authors mention the problem that due to generalisation, the level of detail of the building model is limited, and they propose a second method for the case that an image does not show enough corner points. In this case, the pose is not calculated per image but for a sequence of images that are assumed to observe a common plane in object space that covers most of each image. This plane is used to calculate relative homographies between the images and the façade based on homologous points.

Ding et al. (2008) determine a coarse image pose using vertical vanishing points to estimate roll and pitch angles using measurements of a compass for the yaw angle and GNSS for the position. Points as corner features from intersecting orthogonal edges are matched between image and model, assuming many horizontal and vertical lines from buildings in the captured scene. Matching candidates are first reduced by a Hough transform, and then Random Sample Consensus (RANSAC) (Fischler and Bolles, 1981) is used to find the best homography between image and model features. The image pose finally is estimated using the corresponding point features. Wang et al. (2013) find line segments close to projected model edges and refine image pose parameters in an adjustment with the coplanarity equation. In this way, they avoid the matching of line endpoints.

Reference	Object Knowledge	Matching Primitives	Matching Method	Pose Estimation
Gerke (2011)	Generic	Lines/points	Manual	Bundle adjustment
Avbelj et al. (2015)	DSM	Lines	Accumulator	Image-wise
Läbe and Ellenbeck (1996)	Wireframe	Lines	Pose clustering	Image-wise
Hsu et al. (2000)	Wireframe	Lines	Energy field	Sequentially
Frueh et al. (2004)	Wireframe	Lines	Exhaustive search	Image-wise
Hoegner et al. (2007)	Wireframe	Points	Corner point pairs/ homography	Image-wise
Ding et al. (2008)	Wireframe	Points	Hough transform/ RANSAC	Image-wise
Wang et al. (2013)	Wireframe	Lines	Geometrical reliability measure	Image-wise
Jung et al. (2016)	Wireframe	Points	Geometric hashing/scoring	Image-wise
Iwaszczuk et al. (2012)	Wireframe	Lines	Counting/RANSAC	Image-wise
Arth et al. (2015)	Wireframe	Lines	RANSAC/scoring	Image-wise
Urban (2016)	Wireframe	Lines	Nearest neighbour/ particle filter	Sequentially
Eugster (2012)	Wireframe	Lines	Line features/ relational matching/ energy function	Sliding window adjustment
Li-Chee-Ming and Armenakis (2018)	Wireframe	Lines/points	Random sampling/ resection	Image-wise

Table 1: Overview of related work that integrates object knowledge in image space.

Points and their edge context are used as corresponding features between an image and a building model by Jung et al. (2016), but they focus on corners at building roofs as satellite images are used. The context of the so-called edged corner features is taken into account in a geometric hashing to find corresponding corners and adjust the image pose parameters. In geometric hashing, a representation of model objects is stored in a hash table. This table is then used in a vote-counting scheme for observed objects to find correspondences. The authors replace the vote-counting by a scoring function that takes into account corner position distances as well as length and angle differences of edges at the corner features.

The uncertainty of lines of the building model is taken into account by Iwaszczuk et al. (2012). They set up a line based least squares adjustment to estimate the pose parameters of one aerial infrared image. The uncertainty of building model vertices is set in their covariance matrix and propagated to a covariance matrix of line parameters. The uncertainty of detected Canny lines (Canny, 1986) in the image takes into account line length and edge strength to calculate line weights. Lines with high weight are taken into account in a matching procedure

in image space based on buffers around lines as search space. The buffer size depends on the accuracy of given initial image poses from GNSS and IMU measurements. Before the adjustment, outliers in line correspondences are eliminated using RANSAC. In the follow-up work (Iwaszczuk and Stilla, 2017), neighbouring images of an image sequence are taken into account to restrict line matching by tracking lines. The adjustment is still done per image only.

Arth et al. (2015) use building outlines and height information to refine the pose of a mobile device to initialise the first frame of a visual SLAM system. Several sensors are used to retrieve a rough initial pose of the frame, and the height over ground is set manually. The scene is assumed to consist of horizontal and vertical edges. Image segmentation is used to restrict detected edges to façades in the image and exclude windows to support the matching of detected and projected edges. The SLAM system is only initialised using the building information and does not further integrate the model information to refine the pose iteratively.

Urban (2016) integrates a building model into indoor pose estimation of a multi fisheye camera platform. Possible platform poses are sampled on a plane within the model, and synthetic views are rendered offline to extract and cluster edge-based descriptors. These descriptors are then used to find the best matching pose from the sampled poses for captured fisheye images with two approaches online: Nearest neighbour search or a particle filter. The model is taken into account to refine the pose of key frames using point-based orientation between frames. The key frame pose parameters are refined by minimizing the reprojection error of edge points that are projected from the building model to the images and are found to correspond to detected edges.

Eugster (2012) addresses the model-based refinement of direct orientation using low-cost sensors for a UAV. Three methods to relate observed images to a building model are described: Line-based feature matching, line-based relational matching and optimisation of an energy function on distance images. Due to an inaccurate pose from direct sensor orientation, relational matching is chosen to find corresponding image and projected building model edges. Relational matching evaluates the mapping of two model descriptions. These model descriptions take into account line features and their topology and are derived for a captured image and for the projected building model. Tree search is used to optimise the mapping between image and projected model descriptions. The found optimal mapping leads to matches of image lines to model lines that are known in 3D space from the building model. These matches are used as hard constraints in a sliding window adjustment with homologous points between key frames of an image sequence. Such refined pose parameters of a current key frame are then input to a Kalman filter that fuses them with the GNSS/IMU measurements to estimate the final pose parameters.

Li-Chee-Ming and Armenakis (2018) present an approach for UAV pose estimation using mono camera image sequences in- and outdoor. Given an approximate camera pose, they apply a line-to-line space resection for pairs of horizontal and vertical lines. Lines are detected in a frame and projected into image space from a building model using a simulation environment. The space resection makes use of the known 3D model lines corresponding to the lines in image space to estimate the camera's pose parameters. The estimated pose is further refined by the

resection of line endpoints and estimation of their object space coordinates followed by point-to-point resection with these points. A SLAM system's current pose is reset for frames where the model-based pose refinement is successful. Experiments show a sub-meter positioning accuracy for in- and outdoor images.

2.2 Integration of Object Knowledge in Object Space

Some authors directly use the object knowledge in object space instead of projecting it into image space to find correspondences. The methods described in the following are listed as an overview in Table 2. Rodriguez and Aggarwal (1990) propose to use so-called cliff maps to refine the pose parameters of an aerial image. The problem of finding image features corresponding to the cliff map is reduced from 3D to 2D as the 3D object knowledge is reduced to 2D assuming a nadir image. The cliff map is derived as a 2D map of edges and points of high curvature extracted from a digital elevation model (DEM). Edges and points are matched to image features.

A DEM is integrated by Strunz (1993) and by Spiegel (2007) into the orientation of satellite line-camera images. The height of reconstructed object points is restricted to lie on the DEM surface using soft constraints in a hybrid bundle adjustment to improve the parameters of the continuous satellite trajectory. Briskin et al. (2017) intersect image rays with a mesh approximation of a DEM to generate soft constraints for a bundle adjustment to refine the pose parameters of a UAV mono camera image sequence.

Matches of lines in 2D image and 3D object space are used by Talluri and Aggarwal (1996) and Stamos and Allen (2001) in terrestrial applications to estimate the pose parameters of single images that capture buildings. Talluri and Aggarwal (1996) reduce the pose parameters to be estimated for a robot to the 2D position in the ground plane and one angle, and use only building roof edges, assuming flat roofs and vertical façades. So-called edge visibility regions are used to restrict the assignment of edges. Candidate regions are identified by a Hough transformation. Stamos and Allen (2001) exploit constraints like parallelism and orthogonality that are assumed to occur at buildings in urban areas to retrieve the relative orientation of single optical to depth camera images. Lines as common features of images and a model are also used by Aider et al. (2005), but they address the localisation of an indoor robot. They match lines by searching an interpretation tree that originally holds all possible combinations of two sets of lines. Before the search, correspondences in the tree are reduced by introducing knowledge about edge visibility from an approximate robot pose, and the robot is assumed to move on a plane.

A matching that takes place in object space completely is realised by Zhao et al. (2005). Their goal is to align oblique aerial images taken from a height of 90 m over ground to a building model which they manually derive from a laser scanning point cloud. They manually retrieve an initial pose for the first two key frames based on corresponding image points and model vertices and image and model edges. Homologous points are then used for relative orientation

of further key frames sequentially. A point cloud calculated by dense stereo matching is matched against the laser scanning point cloud using the iterative closest points (ICP) algorithm (Besl and McKay, 1992) to retrieve a refined camera pose for key frames. Relative image orientations and refined orientations per key frame are input to an adjustment to estimate refined and smoothed pose parameters for each frame. They apply the line based method of Hsu et al. (2000) described in section 2.1 to their data and show an example where the line matching fails due to a bad initial image pose while their ICP-based approach is successful. Another experiment shows that the usage of a reference point cloud and ICP allows for image pose corrections at irregular structures like a bridge and trees.

The integration of object knowledge is also done for laser scanning data. Kager (2004) does not integrate a DEM or a building model but instead uses homologous planar patches in aerial laser scanning point clouds as tie features for strip adjustment. Planar features between flight strips are used to derive fictitious observations as soft constraints that require points to lie on the planes in the adjustment. Hebel and Stilla (2012) address the pose refinement of a helicopter equipped with an oblique forward-looking line laser scanner. They propose to find corresponding planar patches between flight strips or between captured data and given planes from a database. They apply their method to the adjustment of the relative orientation of the scanner to GNSS/IMU sensors (mounting calibration) if these sensors provide sufficiently accurate measurements. They also use the plane database to refine the helicopter pose if GNSS/IMU measurements are of low accuracy or unavailable. Planar patches are matched to planes concerning distances of their centre of gravity (COG), angle between normal vectors and similarity of eigenvalues of covariance matrices. Plane matches are filtered using RANSAC, and a least-squares adjustment that weights planar shapes depending on their size is carried out.

Laser scanning with a mobile mapping platform in an urban environment is examined by Monnier et al. (2013). They relate scanned object points to vertical façade planes of a 3D building model. A point is assigned to a façade plane if their distance is below some threshold and if the point's and the plane's normal head in a similar direction. A point normal is calculated based on neighbouring points. Assuming good initial pose parameters and stable attitude, drift parameters for the 3D platform position are estimated offline using a non-rigid ICP: ICP between scanned points and model planes is carried out on time-dependent parts of the trajectory assuming slowly varying translational drift. Non-façade points are filtered out by a geometric descriptor as well as points at which the measurement ray does not intersect a façade plane. Non-rigid ICP drift parameter adjustment and the assignment of points to planes are iterated alternately. Monnier et al. (2013) mention that the problem could be reduced to 2D as only vertical façades are used. This is what Vysotska and Stachniss (2017) propose in their constraint sliding window adjustment. They align terrestrial laser scans to building façades from a given map to estimate the 2D position and the heading of a robot moving on the ground. A new window is issued every 25 m. Correspondences of points at lines identified in the laser scan and lines from building outlines are found using ICP. They also analyse the localisability of possible robot positions in the map to avoid localisation problems before moving there.

The combination of laser scanner and camera observations for pose estimation with object knowledge for an aerial platform is addressed by Glira (2018). He describes a hybrid adjustment that can take into account correspondences in object space (strips from laser scanning, known 3D point clouds like a DEM, photogrammetric tie points) and in image space (homologous points, GCPs) to estimate the pose parameters as a continuous trajectory of an aerial platform. A trajectory from direct orientation using a Kalman filter with GNSS and IMU measurements is assumed to be given. Correction parameters for this trajectory as well as for the sensor mounting calibrations are estimated in the hybrid adjustment. Correspondences in object space are found for laser scanner points as well as for sparse photogrammetric tie points using adapted ICP. The point-to-plane distance is used in ICP instead of relaxing the requirement of finding point-to-point correspondences. The author calculates a tangent plane for a laser point and its neighbouring points using a principal component analysis of the covariance matrix of their coordinates. Points are rejected during ICP if their normal vector is unreliable in terms of a high measure of roughness calculated using the neighbourhood of the point, and if angles of normal vectors of corresponding points or the point-to-plane distance exceed a threshold. The distance threshold is set dynamically as a tolerance around the median of the point-to-plane distances. Correspondences are weighted based on their roughness and measurement precision and lead to soft constraints in the robust hybrid adjustment. Correspondences are searched for iteratively in each iteration of the adjustment.

Lothe (2011) improves the pose parameters of a car moving along urban streets using a forward-looking wide-angle camera mounted on the car. The results of a SLAM approach are corrected in two steps. The first step is a non-rigid ICP that estimates concatenated transformations of linear parts of the trajectory placing joints at curves of the trajectory. The height of the camera is kept at a fixed height above the ground, assuming a horizontal ground plane for the whole image sequence as an elevation model. An approximate initial pose is set manually. The metric used for ICP is the orthogonal distance of 3D tie points to model planes. Points are assigned to their closest model plane. M-estimation (Huber, 1981) is used for a robust estimation by reducing the weights of wrong assignments of tie points to model planes per trajectory segment. In the estimation of the transformation parameters, segments are weighted depending on an estimated threshold and the number of corresponding tie points. As the non-rigid ICP leads to discontinuities between segments and cannot correct for drift effects on linear trajectory segments, a second step refines the results. An adjustment is set up that minimises residuals of image observations for back-projected intermediate points. Intermediate points lie in an associated model plane. The classical reprojection error is combined with an error induced by triangulated points that lie exactly on the model planes. Lothe (2011) uses M-estimation to identify outliers concerning the combined reprojection criterion. As in the first step, tie points are assigned to their closest model planes, and the assignments are kept fix during optimisation. Therefore, the assignment and the consecutive adjustment is iteratively repeated. Points

Reference	Measurements	Object Knowledge	Matching Primitives	Matching Method	Pose Estimation
Rodriguez and Aggarwal (1990)	Image	DEM	Edges/points	Alignment/ consistency checks	2D, resection
Strunz (1993); Spiegel (2007)	Line image	DEM	Point/plane	Height over DEM	Bundle adjustment
Briskin et al. (2017)	Image	DEM	Ray-to-plane	Surface intersection	Bundle adjustment
Talluri and Aggarwal (1996)	Image	Wireframe	Edges	Hough transformation	2D, resection
Stamos and Allen (2001)	Image	Depth camera	Edges	Vanishing points, line clustering	Resection
Aider et al. (2005)	Image	Wireframe	Edges	Interpretation tree	Resection
Zhao et al. (2005)	Image	Point cloud	Points	ICP	Bundle adjustment
Kager (2004)	Laser	Point cloud	Point/plane	Overlapping patch candidates	Adjustment
Hebel and Stilla (2012)	Laser	Point cloud	Point/plane	Overlapping patch candidates, RANSAC	Adjustment
Monnier et al. (2013)	Laser	Wireframe	Point/plane	Non-rigid ICP	Adjustment
Vysotska and Stachniss (2017)	Laser	Building outlines map	Point/line	ICP	2D, adjustment
Glira (2018)	Image, laser	Point cloud	Point/plane	ICP	Hybrid adjustment
Lothe (2011)	Image	Wireframe	Ray-to-plane	non-rigid ICP	Adjustment
Tamaazousti et al. (2011)	Image	Wireframe	Ray-to-plane	non-rigid ICP	Bundle adjustment

Table 2: Overview of related work that integrates object knowledge in object space.

identified as outliers or not related to any plane are excluded from the adjustment. The height of the image projection centres is again kept at a fixed height above ground. The approach is reported to have problems in areas with too many objects occluding building façades, where the robust elimination of tie points fails, and in areas where there are not enough model planes, because only assigned tie points are part of the adjustment. Unassigned points are excluded due to unsolved problems in weighting two types of reprojection errors for assigned and unassigned tie points.

This problem is addressed by Tamaazousti et al. (2011), who extend the work of Lothe (2011). They integrate tie points not related to model planes in a bundle adjustment by combining two cost functions that are weighted according to a rejection threshold of the robust estimator. Tie points related to a plane are estimated with two degrees of freedom whereas unrelated tie points are classically represented by three degrees of freedom.

2.3 Discussion

The collection of related work shows that the topic of integrating a priori knowledge of the captured scene for pose estimation is addressed for different sensors on platforms operating from ground to space. Geometric object knowledge taken into account includes height models, wireframe models and point clouds. Constraints are retrieved from a variety of features such as points, lines, planes and information about their topology. Correspondences between features are found using various strategies from simple geometrical assignments to voting schemes, hashing and tree search. The methods aim to refine a platform pose for several goals like navigation, model texturing, sensor calibration, strip adjustment, error identification, loop closing, drift reduction and reduction of the need for other ground control information.

The most limiting problem for the application of these existing methods to the scenario addressed here is the impact of model generalisation effects given a much higher geometrical resolution and accuracy of the captured photogrammetric data compared to the model representation of the captured scene. It cannot be assumed that enough building model corners or edges are visible in a captured image to estimate its pose without considering additional observations. Due to the proximity of the camera to the façades, it is likely that there will be images not observing any model vertex or edge, as only a part of one façade is captured. This excludes methods that estimate single image poses based on such an assumption to produce usable results. The closer the façades are to the sensor, the larger the difference between relatively accurate photogrammetric measurements and the generalised building becomes. Methods are less affected by the generalisation effects if they are applied to data captured from a larger distance to buildings, given the same geometrical resolution and degree of generalisation.

It cannot be assumed that building corners or edges observed in an image of relatively high resolution correspond to model features due to the generalisation of the model. Typically, the

complexity of a real building will be reduced to an aggregated simpler representation. The camera captures the complex structure while the generalised model only shows the simpler representation. For this reason, direct matching of vertices as point or edges as line features between image and model is avoided here. Therefore, many of the listed methods, especially those that match features in image space, are not suitable to solve the given problem.

Addressing a UAV as sensor platform means that several assumptions which are applied for other platforms cannot be used: The UAV does not move on a ground plane or in a fixed height, nor can the images be assumed to be taken with some constant attitude although the camera is gimbal-stabilised, so that the roll angle is close to zero.

The type of generalised building model considered in this thesis consists of planes of arbitrary orientation. Assumptions on e.g. orthogonality, verticality, horizontality and parallelism of planes are avoided (although façades theoretically are known to be vertical).

Only the sparse photogrammetric tie point cloud is used here. As object points are assigned to model planes, the related work using laser scanning data, especially Glira (2018), is close to the model addressed here. However, the irregular distribution and spacing of points from images likely prevent normal estimation. Therefore, no point normal, as described by some authors for laser scanning point clouds, is employed in the assignment of points. As the goal is to reduce drift effects and block deformations, no rigid transformation, like ICP, is conducted here. Instead, soft constraints that relate tie points to model planes in object space are set up in a hybrid adjustment. This adjustment estimates pose parameters for each image, tie point coordinates and other parameters. Related papers close to the model described here are Lothe (2011) and Tamaazousti et al. (2011), but the authors use the information that their platform moves on a ground plane and assume locally rigid transformations in their trajectory estimation. In addition and also in contrast to Glira (2018), the amount of generalisation is explicitly handled in this thesis during the assignment of tie points to model planes and in the adjustment by a standard deviation of the soft constraints.

3 Photogrammetric Pose Estimation with a Generalised Building Model

At the beginning of this chapter the mathematical model of photogrammetric pose estimation with a single camera using a priori knowledge of the observed scene in the form of a generalised building model is described. The method uses a hybrid bundle adjustment integrating planes as primitives of the building model. An overview of the mathematical model is presented in section 3.1. In section 3.2, the functional and stochastic models of the hybrid bundle adjustment for pose estimation are described. The integration of a generalised building model into the hybrid bundle adjustment in terms of relating observed object points to primitives of the generalised building model is covered in detail in section 3.2.1. In section 3.2.5, the retrieval of initial values for the optimisation is described. In section 3.3, the workflow is described and the method is extended to a sliding window approach to be applied to image sequences.

3.1 Overview

Pose estimation using prior object knowledge consists of an optimisation that integrates a building model. Entities of the model relevant to the method are described in the following.

The model consists of a single camera that observes objects in multi-view stereo configuration while moving along the scene. Buildings of that scene are known in the form of a generalised building model represented in boundary representation. The building model consists of planar elements that are represented as faces defined by vertices. The vertices are given as a list of 3D coordinates. The topology of the model is represented by a list of the vertex indices that belong to each model plane. The vertices of a plane describe the boundary of the face. A face thereby is a bounded portion of the plane.

Figure 1 depicts relevant entities to describe image poses and the building model as well as the relation of tie points to model planes with respect to their orthogonal distance. This geometric distance of tie points to model planes in object space describes the relation between points observed in the images and model planes without the need to recognise the model vertices in the images.

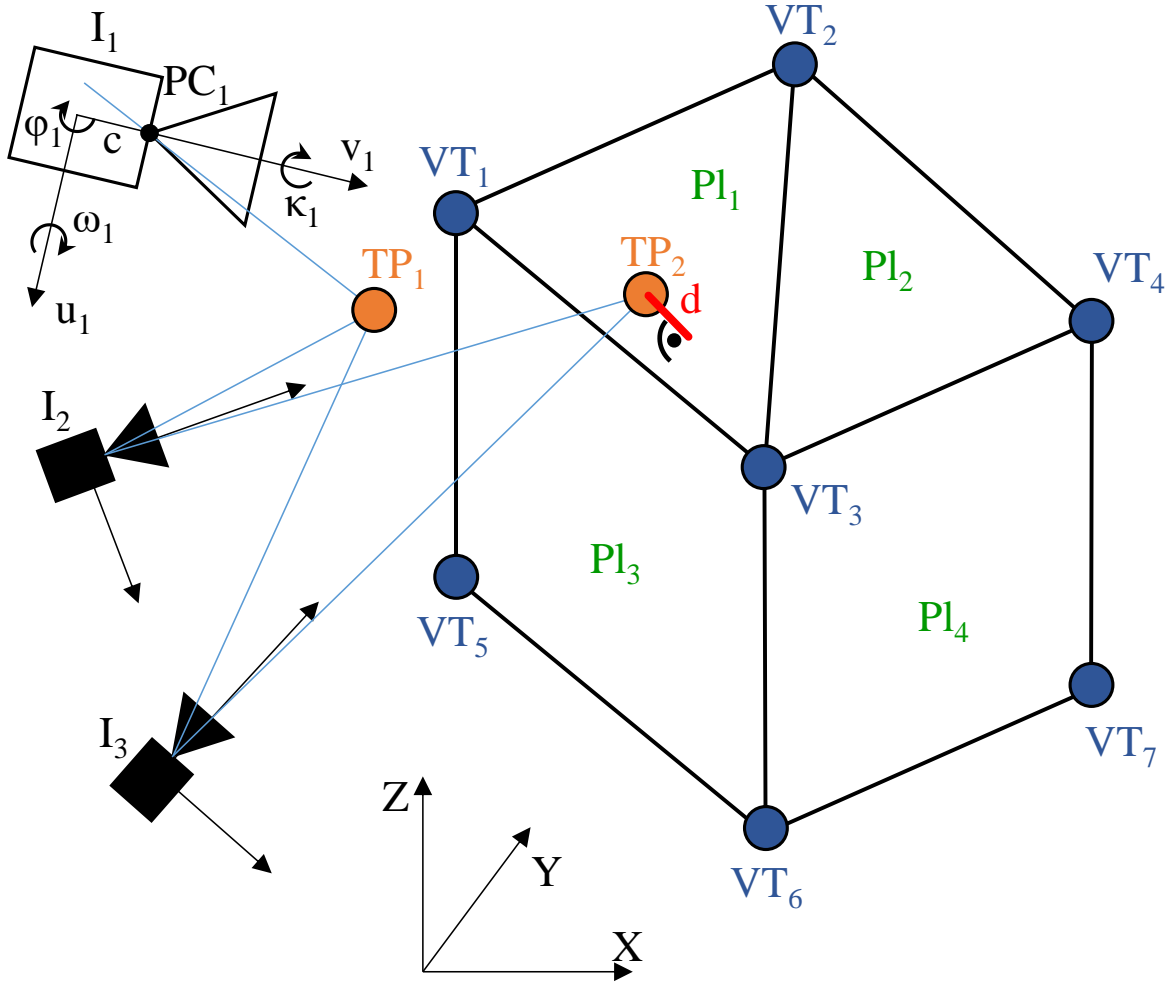


Figure 1: Relevant entities in the model: Three poses I_i of the same camera with focal length c , orthogonal image coordinate axes (u_i, v_i) , projection centres PC_i and three rotation angles $(\omega_i, \phi_i, \kappa_i)$ with $i \in \{1, 2, 3\}$ (shown only for image 1) represent the multi-view model. Two captured tie points are denoted as TP_m (orange) with $m \in \{1, 2\}$. The generalised building model is represented by its vertices VT_k (blue) with $k \in \{1, 2, \dots\}$ and by j planes Pl_j (green), the VT_k are situated in, with $j \in \{1, 2, 3, 4\}$. Camera poses, tie point and vertex coordinates are described in the world coordinate system with axes X, Y, Z . TP_1 is not assigned to any plane while TP_2 is related to Pl_1 by the orthogonal distance d shown in red.

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 ω ,

φ, κ . 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.

The mathematical description of the relations of tie points to model planes and the parameterisation of the planes is covered in section 3.2.1.

3.2 Hybrid Bundle Adjustment

Pose estimation of an image sequence and a building model is based on nonlinear relations between observations and unknowns and requires an iterative solution. Bundle adjustment is an established method that iteratively estimates unknowns using the underlying nonlinear functional relations (Eq. (1)). The adjustment is hybrid because it uses different observation types, not only image coordinates. The hybrid adjustment uses the Gauss-Markov model.

The entities shown in Figure 1 are interpreted as observations and unknowns. For the sake of clarity, indices that differentiate between single realisations of entities of variables are not used in the following.

The following observation types that lead to observation equations f are used as inputs to the adjustment:

- f_{TP} : image coordinates of tie points $(u, v)_{TP}$
- f_{CP} : image coordinates of check points $(u, v)_{CP}$
- f_{pose} : direct observations $(X_0, Y_0, Z_0)_{pos}$ and $(\omega, \varphi, \kappa)_{att}$ for the pose of the images
- f_{VT} : direct observations for the vertices of the building model $(X, Y, Z)_{vertex}$
- $f_{d_{TP}}$: fictitious observations d_{TP} relating object space coordinates of a tie point to one plane of the building model
- $f_{d_{VT}}$: fictitious observations d_{VT} relating object space coordinates of a vertex to planes of the building model

These observations are used to estimate the following unknowns:

- $X_0, Y_0, Z_0, \omega, \varphi, \kappa$: the pose parameters for each image I (projection centre coordinates and rotation angles)
- $(X, Y, Z)_{TP}$: the object space coordinates of each tie point TP
- $(X, Y, Z)_{CP}$: the object space coordinates of each check point CP
- α, β, δ : the parameters of each plane Pl of the building model
- $(X, Y, Z)_{VT}$: the object space coordinates of each vertex VT of the building model

The fictitious observations f_d and the plane parameterisation are covered in detail in section 3.2.1.

A simplified version of the model is shown in Figure 2 with a corresponding graph aside. This example will be used to describe the workflow of pose estimation we apply. The upper part of the graph represents images I and tie points TP . Homologous point observations are represented by blue edges between key frames and reconstructed 3D tie points. Check points are handled similar to tie points and are not represented in the figure. The lower part represents the vertices VT of the building model that span the model planes Pl .

Figure 2 follows the concept of a factor graph (Kschischang et al., 2001). The graph consists of unknown random variables (large circles) and probabilistic information (small circles). Measurements (e.g. image coordinates, GNSS positions) and prior knowledge (coordinates of vertices of the building model) are probabilistic information that corresponds to observations in the adjustment.

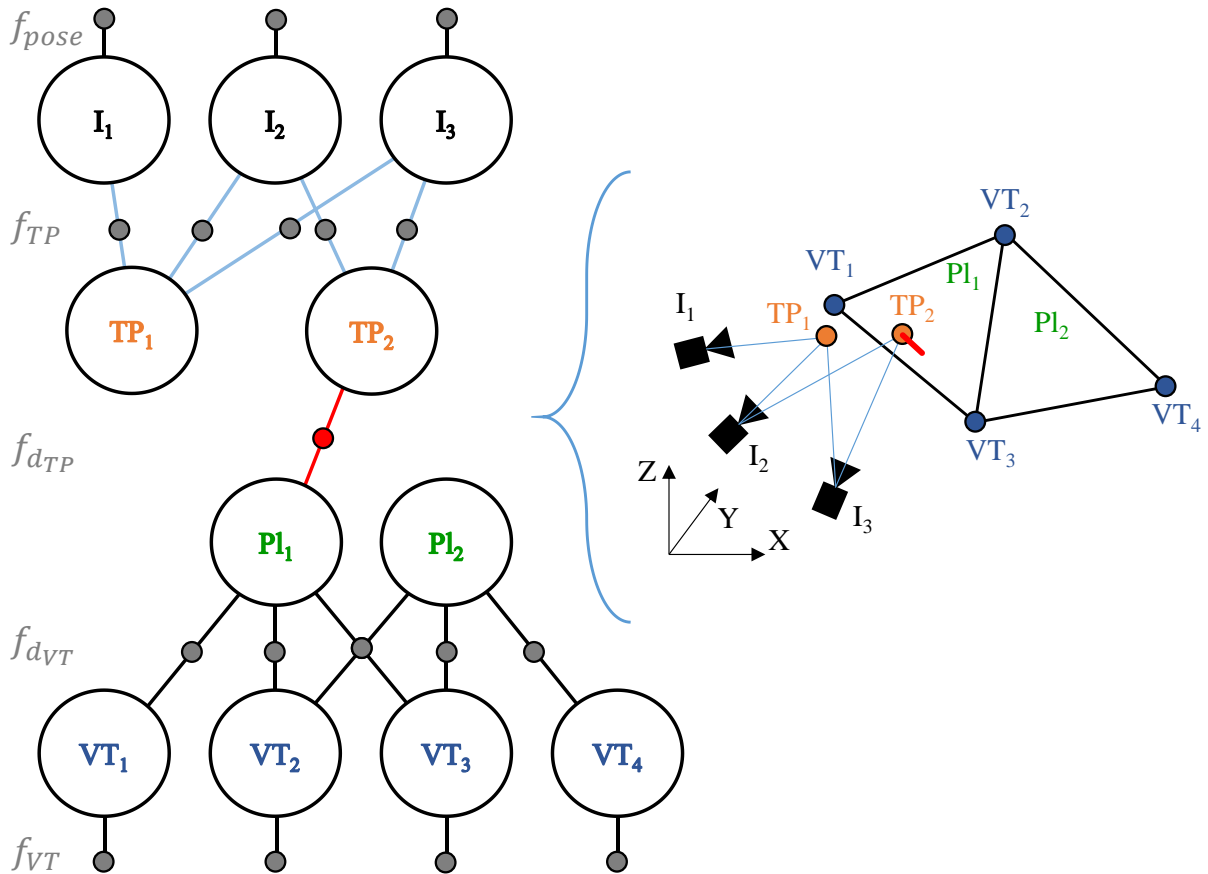


Figure 2: Example model from Figure 1 with corresponding graph representation (for simplification, only two planes of the building model are considered). The graph consists of the scene entities shown as circular nodes and edges that indicate which entities are related to each other. TP_2 is related to Pl_1 by the orthogonal distance d , indicated in the graph as a red edge between TP_2 and Pl_1 . TP_2 is observed (blue edges) by the images I_2 and I_3 . The vertices VT_k are connected (black edges) to the planes they belong to. No tie point is assigned to Pl_2 which means that Pl_2 is defined only by its three vertices $k = 2, 3, 4$. Small circles on edges represent the observations f_{pose} , f_{TP} , $f_{d_{TP}}$, $f_{d_{VT}}$ and f_{VT} .

Observations are represented by the small circles on the edges. The underlying observation equations f_{pose} , f_{TP} , $f_{d_{TP}}$, $f_{d_{VT}}$ and f_{VT} mathematically describe the relations between observations and unknowns visualised by the edges and thereby represent the functional model of the adjustment (section 3.2.2).

3.2.1 Modelling Relations of Object Points to Model Planes

The relation of an object point to a plane of the building model occurs twice ($f_{d_{TP}}$, $f_{d_{VT}}$) in Figure 2. The mathematical description of both types is identical. The relation is described as a so-called fictitious observation of the orthogonal distance of a point to a plane. As this distance is “observed” to be 0 if a point lies in the related plane, the observations are called “fictitious”.

The relation shown in the graph in red stands for the group of fictitious observations that relate object space coordinates of tie points to planes of the building model $f_{d_{TP}}$. The relations between model vertices and model planes constitute the group of fictitious observations relating object space coordinates of a vertex of the building model to the planes of the building model $f_{d_{VT}}$.

Figure 3 shows two planes and a related tie point. The relationship between object point and model plane is expressed in a local coordinate system. Similar to Kraus (1996), a local coordinate system (x_j, y_j, z_j) is attached to each plane in which the fictitious observation equations are formulated. Six parameters describe the pose of this local plane coordinate system in the object

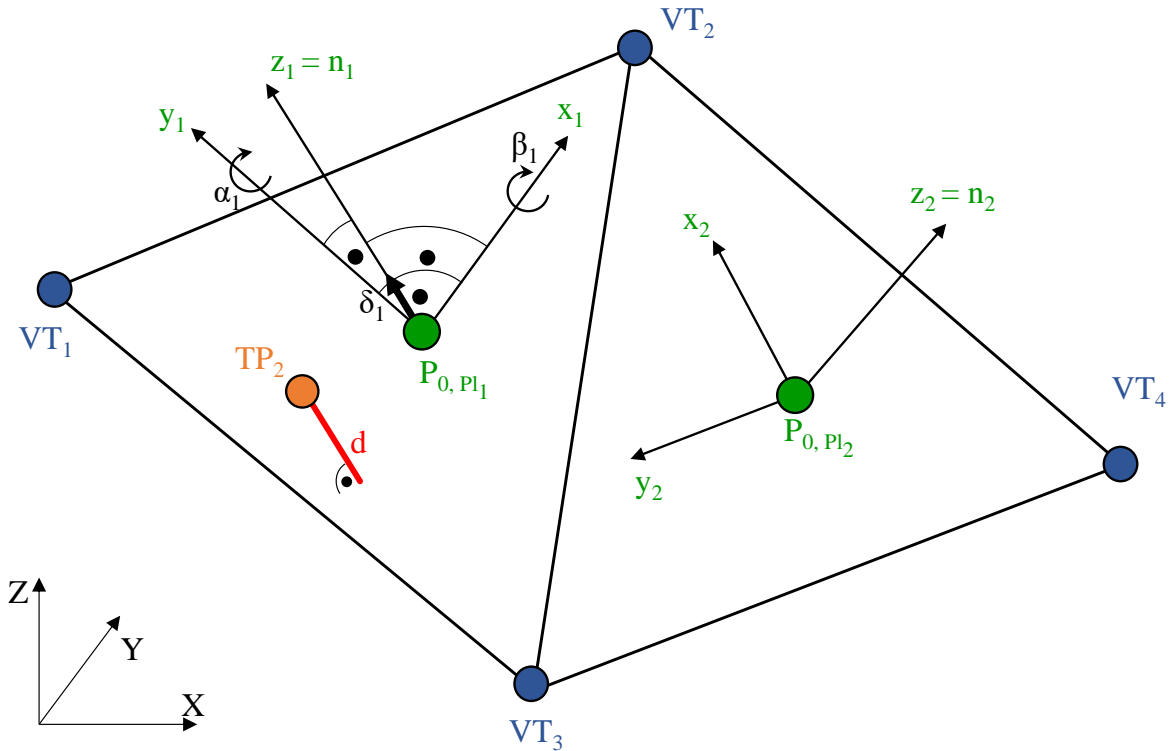


Figure 3: Each plane j has a local coordinate system (x_j, y_j, z_j) where initially the local z_j -axis is the plane normal n_j and x_j, y_j are axes in the plane. The origin of the coordinate system of plane j is P_{0,Pl_j} ; each plane coordinate system is rotated relative to the world coordinate system (X, Y, Z) by a rotation matrix R , not shown in the figure. Two angles α, β and a shift δ (bold arrow) along the local z -axis (only shown for Pl_1) represent the parameterisation of the local plane. TP_2 is shown in orange with its distance d to Pl_1 (red).

coordinate system (X, Y, Z) . These parameters are a 3D-shift P_0 from the object coordinate system to the local one for each plane and three rotation angles (used to parameterise a 3D-rotation matrix R). 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 stated before, vertex coordinates per model plane represent the building model. The local coordinate system per plane is defined as follows. The origin of the local system P_0 is the centre of gravity of all vertices that belong to a plane. The matrix of second moments (variance-covariance matrix) M of the plane vertices VT is calculated according to Eq. (3) interpreting the vertex coordinates per coordinate axis as observations (random variables).

$$\text{cov}(A, B) = \frac{1}{N-1} \sum_{i=1}^N (A_i - \mu_A)(B_i - \mu_B)^T \quad (2)$$

with the mean μ_A and μ_B of two random variables A and B of length N

$$M = \begin{pmatrix} \text{cov}(VT_X, VT_X) & \text{cov}(VT_X, VT_Y) & \text{cov}(VT_X, VT_Z) \\ \text{cov}(VT_Y, VT_X) & \text{cov}(VT_Y, VT_Y) & \text{cov}(VT_Y, VT_Z) \\ \text{cov}(VT_Z, VT_X) & \text{cov}(VT_Z, VT_Y) & \text{cov}(VT_Z, VT_Z) \end{pmatrix} \quad (3)$$

The eigenvalue decomposition of M (Eq. (4)) is used to retrieve the rotation matrix R that describes the rotation from the object coordinate system of the vertices to the local coordinate system. R results as the matrix of eigenvectors sorted along columns according to the size of the eigenvalues in the diagonal matrix D . As R is a rotation matrix, $R^{-1} = R^T$.

$$M = R \cdot D \cdot R^{-1} = R \cdot D \cdot R^T \quad (4)$$

The third column of R corresponds to the smallest eigenvalue and represents the normal vector which points in the direction of the z -axis of the local coordinate system.

The graph in Figure 2 at the bottom shows the vertices and, $f_{d_{VT}}$, the type of fictitious observations that relate the vertices to the planes. Direct observations of the object coordinates of the vertices are introduced. Vertex coordinates are part of the adjustment as unknowns. The change of a plane due to adjusted vertex and tie point coordinates is described within the local plane coordinate system: P_0 and R are constant during an iteration of the adjustment and instead, the plane is parameterised in the local system by two angles α, β defining the direction of the normal and a translation δ along the (local) z -axis. Consequently, α, β and δ are the entities to estimate as unknowns. Using this plane parameterisation, the relation between a point and a plane is described following Eq. (5).

$$d = N(\alpha, \beta)^T \cdot \bar{P}(X, Y, Z) + \delta \quad (5)$$

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

$N(\alpha, \beta)$ is the normal vector of the plane as a function of the two angles. $\bar{P}(X, Y, Z)$ is the object point expressed in the local coordinate system defined by R and P_0 . α, β and δ are initialised as

zero. Whenever the parameters α, β and δ are changed in the iterative adjustment, these values are used to update R and P_0 . After the parameter update, the adjusted plane again corresponds to the (slightly shifted and rotated) x-y coordinate plane of the local system. α, β and δ are reset to zero after updating R and P_0 for each iteration.

As stated before, the distance d between a point and a plane is assumed to be zero, i.e. the point is assumed to lie in the plane. For the vertices of the building model, the model defines in which plane they are situated in. In contrast, relations between tie points and model planes must be established prior to carrying out the adjustment (see chapter 4).

3.2.2 Functional Model

The Gauss-Markov model, according to Eq. (6), is used to minimise the sum of squared residuals v of the given observations l . l and v are $n \times 1$ vectors. n refers to the number of observations; u is the number of unknowns. The $n \times u$ design matrix A consists of the derivatives of the observation equations with respect to the unknowns. The observation equations mathematically describe how the observations contribute to the estimation of the parameters \hat{x} . The hat symbol $\hat{\cdot}$ denotes estimated variables.

$$l + \hat{v} = A\hat{x} \quad (6)$$

The following observation equations are the mathematical formulation of the adjustment:

- f_{TP} : $\hat{v}_{(u,v)_{TP}} = f(\hat{X}, \hat{Y}, \hat{Z}, \hat{X}_0, \hat{Y}_0, \hat{Z}_0, \hat{\omega}, \hat{\phi}, \hat{\kappa}) - (u, v)_{TP}$ and
 f_{CP} : $\hat{v}_{(u,v)_{CP}} = f(\hat{X}, \hat{Y}, \hat{Z}, \hat{X}_0, \hat{Y}_0, \hat{Z}_0, \hat{\omega}, \hat{\phi}, \hat{\kappa}) - (u, v)_{CP}$
 for the image observations of tie points and check points according to the classical collinearity equations (Eq. (1)). The interior orientation of the sensor is considered as known and constant.
- f_{pos} : $\hat{v}_{(X,Y,Z)_{pose}} = (\hat{X}_0, \hat{Y}_0, \hat{Z}_0) - (X, Y, Z)_{pos}$ and
 f_{att} : $\hat{v}_{(\omega,\phi,\kappa)_{att}} = (\hat{\omega}, \hat{\phi}, \hat{\kappa}) - (\omega, \phi, \kappa)_{att}$
 for the direct observations of the image poses.
- f_{VT} : $\hat{v}_{(X,Y,Z)_{VT}} = (\hat{X}, \hat{Y}, \hat{Z})_{VT} - (X, Y, Z)_{vertex}$
 for the direct observations of the building model vertices.
- f_d : $\hat{v}_d = N(\hat{\alpha}, \hat{\beta})^T \cdot \bar{P}(X, Y, Z) + \hat{\delta}$
 for fictitious observations according to Eq. (5).

The collinearity equations represent the photogrammetric model. Absolute information can be added by including observations for the parameters of exterior orientation f_{pos}, f_{att} , e.g. GNSS and/or IMU measurements. The offset of the GNSS receiver to the camera pose is neglected in this thesis due to the expected relatively low accuracy of the receiver. An IMU is not used in this thesis, but all six parameters of the exterior orientation for some images are set to be directly observed based on already estimated parameter values in the sliding window approach to transit from one window to the next (see section 3.2.5).

The direct observations for vertex coordinates of the building model f_{VT} add the building model, which must be available in the same coordinate system, to the adjustment. The building model is formed by the second type of fictitious observations $f_{d_{VT}}$ that denote the topology of the model by relating vertices to model planes. The first type of fictitious observations $f_{d_{TP}}$ is the connection between the building model and the photogrammetric model.

3.2.3 Stochastic Model

The stochastic model represents the stochastic information of the different observations. Uncorrelated observations and a constant a priori level of accuracy for each observation type are assumed. The a priori variance factor σ_0^2 is set to 1. This leads to a diagonal cofactor matrix Q_{ll} of the observations that is equal to the variance-covariance matrix Σ_{ll} (Eq. (7)).

$$\begin{aligned}\Sigma_{ll} &= \text{diag}(\sigma_{TP}^2, \sigma_{CP}^2, \sigma_{pos}^2, \sigma_{att}^2, \sigma_{VT}^2, \sigma_{d_{TP}}^2, \sigma_{d_{VT}}^2) \\ \Sigma_{ll} &= \sigma_0^2 Q_{ll} = Q_{ll}\end{aligned}\tag{7}$$

$$P = Q_{ll}^{-1}\tag{8}$$

The weight matrix P is used in the parameter estimation to minimise the weighted squared residuals $v^T P v$ according to Eq. (9).

$$\hat{x} = (A^T P A)^{-1} A^T P l\tag{9}$$

In Eq. (7), the variances of the measured image coordinates of tie points and check points are denoted by σ_{TP}^2 and σ_{CP}^2 . The variance of direct measurements of the camera poses is reflected by σ_{pos}^2 and σ_{att}^2 . The variance σ_{VT}^2 is related to the accuracy of the coordinates of the building model vertices. For the two groups of fictitious distance observations different variances, namely $\sigma_{d_{TP}}^2$ for tie points and $\sigma_{d_{VT}}^2$ for vertices, are used. As stated before, the mathematical model of both groups of fictitious observations is equal, but the parameters of the stochastic model are different.

According to the building model, the vertices are known to lie exactly on the planes to which they belong. Therefore, their fictitious distance observations conceptually should be zero (for numerical reasons a small variance $\sigma_{d_{VT}}^2$ resulting in high weights of these observations is used). On the other hand, the variance $\sigma_{d_{TP}}^2$ of the observed distance of tie points to their related planes mainly depends on the degree of generalisation of the building model and has to be selected accordingly. The geometrical accuracy of the building model is represented by σ_{VT}^2 while generalisation effects affect $\sigma_{d_{TP}}^2$.

Depending on the scene and the window size, not enough planes with different orientations might be visible. At least four planes with different orientations are needed to fix the 7 parameters of the datum. To correct drift effects of the image block or block deformations, even more planes might be needed. If the planes are inaccurate, redundancy in the planes is a requirement. Redundancy in the planes is also needed to improve not only the trajectory but also the model

planes. Therefore, σ_{VT}^2 by default is set to a small value, to almost fix the building model. The photogrammetric block then is adjusted with respect to the model. This means that model planes cannot move towards the tie points instead of correcting the potentially erroneous trajectory. On the other hand, correspondences of tie points to erroneous model planes directly influence the trajectory instead of changing the coordinates of the vertices of the model planes.

The assumption of a constant level of accuracy for the individual observation groups simplifies the model. This is especially the case for σ_{dTP}^2 , which in reality is likely to vary between the fictitious observations for tie points. These variations are not known and they are mainly caused by generalisation effects of the building model. The reconstructed tie points represent the real scene without generalisation and therefore differ from the generalised planes. In addition, the differences between generalised model and captured real scene are not covered by the assumption of random errors as they likely consist of systematic offsets of groups of points from a plane. Therefore, this kind of stochastic model poses a simplification of the real situation. The possible influences of systematic effects are neither covered by the functional nor by the stochastic model. They have to be taken into account when interpreting the outcome of the hybrid adjustment.

3.2.4 Robust Estimation

The number of outliers is already reduced when generating the initial values for the adjustment. Outliers in the image observations are largely reduced beforehand by using RANSAC to retrieve relative image orientations based on a set of inlier homologous points removing outliers of the image matching procedure. The assignment of tie points to planes of the building model also has the goal to minimise the number of outliers in the fictitious observations. Only tie points close to model planes are assigned, which means that no large residuals are accepted already before and also between iterations of the adjustment.

Remaining outliers have to be detected by the adjustment procedure. This is done following the iteration scheme for robust estimation described by Förstner and Wrobel (2016). The iteration scheme allows for the separation of inliers and outliers first, before removing observations based on a threshold on their residual. This prevents the removal of observations that initially show large residuals, although being inliers.

Assuming good initial values, the weights of observations originally retrieved using Eq. (8) are varied during iterations of the iteration scheme based on the residuals. The initial Σ_{ll} is not changed. The residuals are normalised by dividing them by the a priori standard deviation of their observation type. The normalisation allows for the usage of a weight function for all observation types.

Outliers can only be detected if their residual is large enough and if there is enough redundancy to distinguish between outliers and inliers. The implemented hybrid adjustment allows to turn on robust estimation for every observation type individually.

3.2.5 Determination of Initial Values

Initial values for the adjustment are retrieved by sequential structure from motion. The goal is to derive homologous points in the images and to estimate image poses and 3D object coordinates of tie points in an arbitrarily chosen coordinate system. Not every frame of the image sequence is taken into account: At the beginning, based on a given time interval key frames are chosen along the image sequence. Points are extracted for the first and the second key frame. Image matching for the found points is carried out to identify homologous points. Based on those point pairs, the parameters of the relative orientation of the two key frames are retrieved and the base length is set to 1 to fix the scale.

One key frame after the other is added along the sequence using matches to already reconstructed tie points to estimate relative image poses by spatial resection. New tie points are triangulated as soon as they are found in at least three images. To reduce the number of outliers already before the adjustment, tie points with reprojection errors exceeding a threshold are rejected. Distant tie points with nearly parallel image rays are also rejected based on a low base-to-height ratio. Frames are skipped based on several criteria such as the number of inliers in image resection and the total number of found tie points (see section 5.3 for details).

To be able to integrate the building model, the image poses and object points must be available in the same coordinate system as the building model. Therefore, the results of the first step must be transformed from the arbitrarily chosen coordinate system into the desired object coordinate system using a 3D similarity transformation (Helmert transformation) with 7 parameters.

Due to the low-cost sensor setup, direct sensor orientation can be used only to some extent: Measurements of the on-board GNSS receiver for positions of image projection centres can be used to roughly define the object coordinate system.

At least three pairs of projection centres that are not situated on a straight line and have sufficient base length are needed to calculate the seven parameters of the Helmert transformation. As the measurements are noisy, pairs of projection centre positions of the first few key frames, measured by GNSS and estimated during image orientation, are used to rescale the local coordinate system. Then, frames of the image sequence are only accepted as key frames if after resection their base length to the previous image of the sequence exceeds a certain threshold. This prevents the usage of frames captured from a close-by position in situations where the flying speed is low. Only after a certain number of key frames has been oriented, the parameters of the Helmert transformation are calculated using these image positions.

3.3 Workflow

3.3.1 Global Adjustment

Table 3 shows the processing steps for the simultaneous estimation of all available key frames. The results of the coarse image orientation (step 1) described in section 3.2.5 are optimised in step 2. This bundle adjustment does not include the building model and is robust with respect to outliers in the image coordinates of homologous points (see section 3.2.4).

Step 1	Image matching between key frames and retrieval of 3D tie point coordinates and parameters of image poses. This step includes the transformation into the coordinate system of the building model.
Step 2	Optimisation using only tie point coordinates and, if available, GNSS/IMU observations of pose parameters (upper part of the graph of Figure 2 only).
Step 3	Establishment of relations between tie points and model planes (edges between TP_m and Pl_j in the graph).
Step 4	Hybrid adjustment including images, tie points, vertices and the planes (complete graph).

Table 3: Workflow of global pose estimation. Steps 3 and 4 are carried out in a loop for the adjustment iterations.

In step 3, tie points are assigned to planes of the building model based on their 3D positions estimated as described in section 3.2.5. Step 4 consists of the hybrid adjustment, including the images, the building model and the found assignments. Steps 3 and 4 are iterated: In each iteration of the hybrid adjustment, the assignment of the tie points to the planes of the building model is recomputed (step 3) based on the current optimisation result. In contrast, the known relations of vertices to planes are not changed.

In step 2 of the workflow, the ability of the adjustment to turn on robust estimation for each observation type individually is used to do a robust estimation with respect to the observations of homologous points only. Outliers found are excluded directly before step 3. Step 4 then is carried out robustly with respect to the first type of fictitious observations only. The separation and elimination of outliers in image coordinates and fictitious observations is done under the assumption that the image block has enough redundancy in image observations to identify outliers. The relatively inaccurate and generalised building model is assumed not to add information that helps in identifying additional outliers in the image observations. In contrast, the separation of outlier detection assures that no image observation is eliminated where wrong assignments to model planes lead to large residuals of a tie point at an image coordinate.

3.3.2 Sliding Window Adjustment

The workflow for global adjustment determines the orientation parameters of all key frames, the tie points and the building model at once. Therefore, in terms of computational load the solution grows with the number of observations and unknowns, which makes it unfeasible for large

numbers of images. An often-seen solution is to split the optimisation into smaller parts. The so-called sliding window approach splits a sequence of N images into smaller subsets corresponding to windows that overlap each other in terms of images and tie points. This simplification of the optimisation is based on the assumption, that only neighbouring images show the same scene and the same planes and therefore are connected. This connection consists of homologous tie points between the neighbouring images and commonly observed model planes. Note, that the simplification to only taking into account neighbouring images typically prevents to close loops of the image sequence. However, with the integration of a building model into sliding window estimation, the results are still related to a common coordinate system, so that loop-closure is expected to become easier or even unnecessary.

Another motivation to process the image sequence and the building model with the sliding window approach are block deformations and drift effects that contradict the assumption (see chapter 4) of having good initial values for the assignment of tie points to model planes: Block deformations due to drift become larger with the length of the image sequence if no absolute information like GNSS is available. If only parts of an image sequence are processed as windows, drift effects are limited in each window before the next window is processed as long as valid correspondences of tie points to model planes are found.

The size of a window in terms of the number of key frames that are considered is denoted as N_{ws} . The sliding window approach computes local optimisations that overlap and therefore are assumed to lead to results close to the outcome of a global optimisation.

Table 4 lists parameters for the sliding window estimation. N_{new} new key frames are added per window. Therefore, the number of old, already adjusted key frames to keep per window is $N_{old} = N_{ws} - N_{new}$. Parameters and observations of older key frames are removed from the new window. Thereby, the parameters of the older key frames are assumed not to be affected by the new images. The N_{old} key frames are part of the new window estimation again. Their already estimated parameters might potentially benefit from additional observations thanks to the N_{new} key frames.

N	Total number of key frames
N_{ws}	Number of key frames per window: window size.
N_{new}	Number of new key frames to add per window.
N_{old}	Number of recently adjusted key frames to keep for a new window: $N_{old} = N_{ws} - N_{new}$

Table 4: Parameters of windowed pose estimation.

Figure 4 shows an extension of the exemplary scene and the corresponding graph representation from section 3.2. Two new images observe two new points. Image I_3 observes the new tie point TP_3 . Image I_4 adds an additional observation of TP_2 . TP_3 is related to the second plane Pl_2 . Whereas images $I_1 - I_3$ from the previous section and the building model are shown as window 1, one old and the two new images, tie points $TP_2 - TP_3$ and relations to the building model are optimised as window 2. I_1 , I_2 and TP_1 do not belong to this new window. In this example $N = 5$, $N_{ws} = 3$, $N_{new} = 2$ and therefore $N_{old} = 1$.

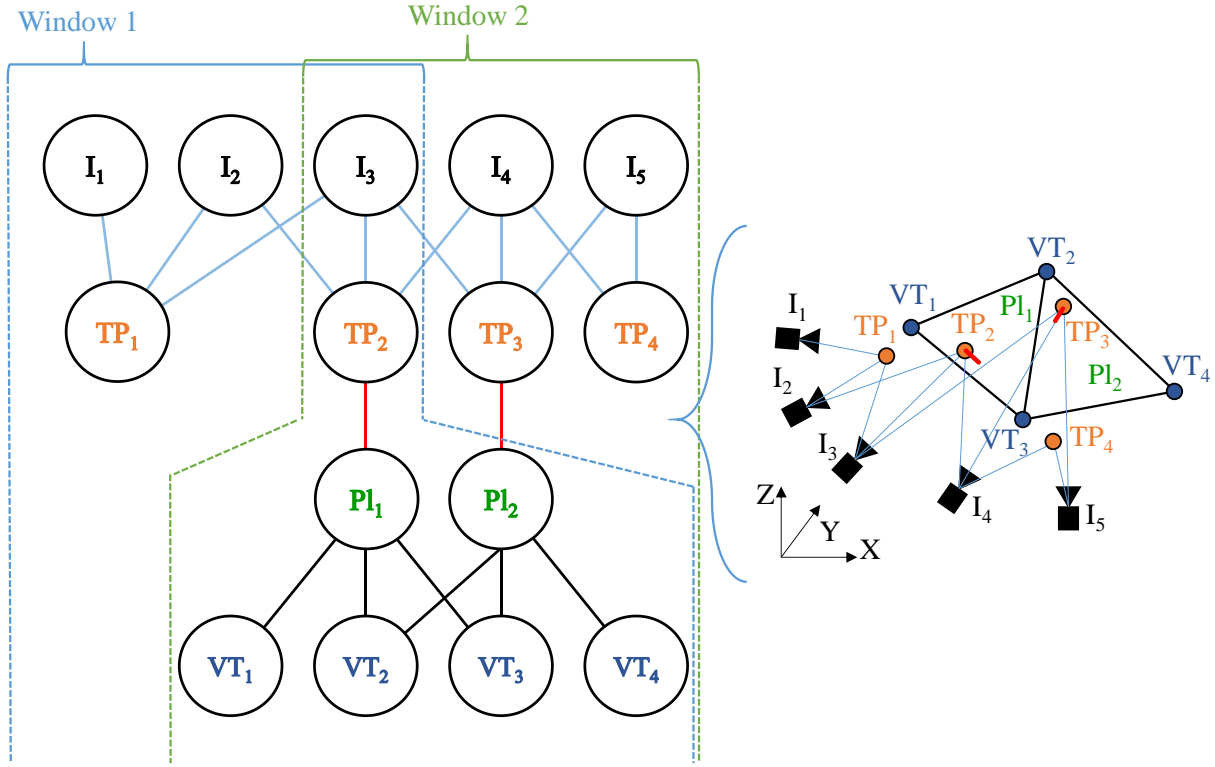


Figure 4: Extended example with corresponding graph representation from Figure 2. Two new images and two new tie points are added. Windows 1 and 2 denote the entities that are part of a sliding window.

Table 5 describes the steps of the sliding window pose estimation. Steps 1 and 2 are similar to those of the global workflow (Table 3) but only use the first N_{ws} images. In the example, the first window refers to images 1-3. The images of the first window are oriented, and image poses and tie point coordinates in the world coordinate system are retrieved. Based on these results, the iterative procedure starts which processes images in a sliding window manner. Steps 3 to 6 are repeated for every window of the sequence.

Step 3 consists of relating the reconstructed tie points of the current window to the building model (red connections in the graph in Figure 4). Step 4 is the hybrid bundle adjustment on the current window including the found relations and the building model. Steps 3 and 4 are iterated together, as for each iteration of the hybrid optimisation the assignments of tie points to model planes are re-established.

After convergence of the hybrid optimisation, the current window has been processed, and step 5 follows: N_{new} new images are added to the adjusted images to carry out image orientation and reconstruction. The N_{new} and the N_{old} images form the new window. In step 6, this new window is then input to an optimisation like in step 2 without considering the building model. The model is added again in step 3, which restarts the cycle for a new window.

The estimated parameters of the exterior orientations of some oldest images with their estimated variances are used in steps 3 and 6 as direct observations for the parameters of the exterior orientation of those images. This is due to two reasons: To define the datum of the block in step 6 if no GNSS measurements are used and to transfer some information from the previous window to the new window. This assumes N_{old} to be at least three, and if it is larger, the exterior orientation

Step 1	Image matching between key frames and retrieval of 3D tie point coordinates and parameters of image poses for the first window of key frames (images 1 to N_{ws}). This step includes the transformation into the coordinate system of the building model.
Step 2	Optimisation using only tie point coordinates and, if available, GNSS/IMU observations of pose parameters.
Step 3	Establishment of relations between tie points and model planes (edges between TP_m and Pl_j in the graph).
Step 4	Hybrid optimisation including images, tie points, vertices and the planes.
Step 5	Based on the last N_{old} adjusted image poses, add N_{new} key frames of the next window and retrieve image poses and tie point coordinates.
Step 6	Optimisation for a window of N_{old} remaining and N_{new} key frames without considering the building model. The already optimised parameters of the exterior orientation of some of the oldest images are included as direct observations with their estimated variances and covariances from the window before.
Step 7	If all N key frames are optimised, a post-processing step can follow to refine the results: Final global hybrid adjustment including all key frames, the planes and found relations of tie points to model planes.

Table 5: Workflow of the sliding window pose estimation. Steps 3 and 4 as well as steps 3 to 6 are carried out in loops for adjustment iterations and per window.

parameters of the old images after the oldest three are estimated again. This allows the block to be corrected even if the parameters for the old images were estimated inaccurately. If no GNSS measurements are available for a window and no corresponding model planes are found within the window, these direct observations for the parameters of the exterior orientation are the only absolute information and prevent the normal equation system from becoming singular.

Steps 2 and 6 are made robust with respect to outliers in the observations of homologous points. Step 4 keeps image coordinates that are inliers in step 2 or step 6 and it is made robust only with respect to outliers within the fictitious observations of tie points lying on model planes. Direct observations for parameters of the exterior orientation or for coordinates of model vertices and fictitious observations for the connection of model vertices and model planes are not checked for outliers, assuming they are not affected by gross errors.

After all images have been processed in windows, a global hybrid optimisation can be carried out as an optional post-processing step. This can have two goals: It does a global optimisation, and it can be used to find optimal shift and rotation parameters of the model planes by increasing the a priori standard deviation of the direct observations of model vertices.

The final global optimisation includes all optimised image poses, all image observations and the building model and is done robustly with respect to outliers in image observations and fictitious observations of tie points to model planes. The relations of tie points to model planes are updated in each iteration of the global adjustment and are subject to outlier identification and elimination.

4 Assignment Under Generalisation Effects

The second important problem addressed by this thesis is how the relation of tie points to building model planes can be established. This problem corresponds to step 3 in the global (Table 3) as well as in the sliding window workflow (Table 5). Main factors to consider when setting up methods to find the relations are the inaccuracy of image pose parameters and tie point coordinates as well as the generalisation effects and inaccuracy of the building model. An established relation of a tie point to a model plane leads to the assignment of this point to the plane and is represented in the optimisation as a fictitious observation $f_{d_{TP}}$ (section 3.2.1).

In this chapter, first, the generalisation effects that are to be expected for a typical generalised building model are described. Requirements for the assignment of tie points to model planes are derived, and three different strategies to relate tie points to planes of a generalised building model are presented. The building model and the tie point cloud are inputs to the assignment process.

The first method, abbreviated as *method (a)*, uses a simple distance criterion to assign points to model planes directly. The second *method (b)* is based on finding planes in the entire tie point cloud independent of the building model and matching these planes to the planes of the building model. The third *method (c)* uses the model planes to generate regions-of-interest (ROIs) around the model planes and thereby considers only points in the vicinity of a model plane in order to find a corresponding extracted plane; explicit plane matching is thus avoided, and the search of planes in the tie point cloud is restricted to the ROIs near the given model planes. Method (a) is set up as the simplest and fastest method, while (c) potentially adds robustness and (b) has the highest potential for situations where the assumption of a good initial position of the object points that are matched to the building model is violated.

The fictitious observations for vertices are given as a part of the model's topology and are not changed during adjustment. The strategies to establish fictitious observations for tie points on model planes require that both, the tie points and the building model vertices, must be given in the same coordinate system.

Note that the assignment strategies are not limited to tie points observed by a camera but can be applied for any sensor that delivers 3D points. In addition, other primitives of a 3D model than planes could be used in matching if a suitable distance function is set up for them, and, for method (b), if these primitives are identified in the tie point cloud.

4.1 Generalisation Effects

The building model consists of planes that are defined by vertices which are often not visible in the images. The reason for this is the generalisation of the real world that is employed for the building model. The building model is a simplified representation of the scene: vertices do not have to correspond to real features of the captured scene and most model planes are a simplified representation of more complex structures, e.g. a façade with windows and doors or a roof that consists of tiles. In contrast, the observed tie points directly correspond to features of the captured scene.

The amount of generalisation of building models is often defined by their level of detail (LOD). The widely used City Geography Markup Language (CityGML) encoding standard, version 2.0 applies 5 levels of detail (Open Geospatial Consortium, 2012). While level 0 represents building footprints or roof edge polygons, level 1 is a simple 3D block model. Level 2, in addition, represents roof structures. Higher levels add wall and roof details, e.g., windows or doors (LOD3), and building interiors (LOD4). Typically, the minimal dimension of modelled objects becomes smaller, and the geometrical accuracy increases from lower to higher levels of detail. More and more cities are entirely modelled in LOD1 or LOD2.

The level of detail of a given building model and specification of the minimal dimension of objects and their geometrical accuracy are needed to quantify the degree of generalisation. Often, information about methods used to capture and process the data are also given. Generalisation effects of the models typically stem from generalisation during data capturing rather than from simplifying a model of higher detail. LOD1 can be generated from building footprints out of a 2D GIS database and building height information. Uncertainties from digitising such data from paper maps have to be considered. LOD2 often is produced by adding parameterised roof structures from a catalogue of standardised roof types. These roof structures are reconstructed based on detailed height information, e.g., captured by aerial photogrammetric or laser scanning flights. Therefore, the degree of generalisation and the geometrical accuracy can differ within the same model between wall and roof surfaces.

In this work, the difference between the captured scene and its generalised representation is taken into account by the way relations between captured tie points and the model are designed. Intersections of two planes of the generalised building model do not have to correspond to an edge of the captured scene. As mentioned before, an intersection of three (or more) planes leads to a vertex that might not correspond to a corner of the real building. Therefore, direct use of point features for corners or line features for edges is avoided in contrast to most related work (chapter 2).

Instead, the relation of tie points to generalised model planes is addressed. A tie point is related to a plane by its orthogonal distance as described in section 3.2.1. Instead of using a constraint that restricts a point to lie on a plane by forcing the distance to be zero, fictitious observations, as soft constraints, allow to set an accuracy for this requirement. Ideally, the distance of a tie point to a model plane should be zero, but due to generalisation and other measurement uncertainties, it is

allowed to deviate in a specific range. The a priori standard deviation $\sigma_{d_{TP}}$ of the fictitious observations $f_{d_{TP}}$ represents the degree of generalisation whereas the a priori standard deviation of the direct observations of vertex coordinates σ_{VT} represents the geometric accuracy of the model.

Each tie point is allowed to be related to only one plane in one iteration of the adjustment: A maximum of one fictitious observation per tie point is established. Therefore, tie points on plane intersections and corners are not explicitly modelled, as they must be related to only one of the planes. On the contrary, model vertices usually are related to more than one plane and, therefore, have several fictitious observations per point.

4.2 Direct Assignment: Point-Plane-Matching

A direct way to establish relations of tie points to model planes is the assignment of a tie point to the closest plane in terms of a distance criterion. Consequently, tie points are assumed to be related to the closest plane provided that their Euclidean distance from the plane is below a given threshold. Figure 5 exemplarily shows the tie points (black), the model planes with their vertices (blue) and their distances. The distance threshold (shown in green) has to be selected in accordance with the accuracy and degree of generalisation of the building model and the accuracy of the tie points including potential datum problems of the point cloud.

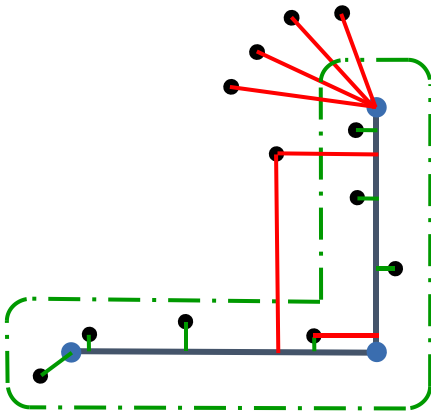


Figure 5: Sketch of the direct assignment (a). Black points are assigned to the closest blue plane if their distance is below a threshold (green dashed-dotted line). Red lines show rejected assignments.

Only if the distance of a tie point to the nearest plane is below the threshold, the relation is established. The larger this threshold is chosen, the more points are assigned to planes, but the higher is the probability to set up wrong assignments. During iterations of the optimisation, the threshold is recalculated and lowered to a pre-defined minimum value. This minimum value is related to the a priori standard deviation of the fictitious observations that relate tie points to model planes $\sigma_{d_{TP}}$: The minimum distance threshold is set to $2 \cdot \sigma_{d_{TP}}$. As described in section 3.2.3, $\sigma_{d_{TP}}$ is set in accordance with the size of the generalisation effects that are expected to occur. The setting of the minimum threshold reflects the assumption, that tie points exceeding a distance of $2 \cdot \sigma_{d_{TP}}$

to a model plane are not related to these planes as soon as a potential overall datum error is corrected.

In the first iteration, the distance threshold has to account for potential datum problems of the point cloud, which are assumed to become smaller during adjustment iterations. Causes for such datum problems can be bad initial values and/or inaccurate GNSS measurements for the camera positions. The threshold is recalculated in each iteration by calculating the mean distances of assigned tie points to the model planes per model plane. If the mean distance over all plane mean distances is smaller than the specified minimum value, the minimum value is set as the new threshold. If the mean distance is above the minimum value, the threshold is set to the calculated mean distance.

Small planes are likely to be removed in the generalisation of the building model. Therefore, the assignments are only kept if a minimum number of tie points is found to potentially correspond to a plane. To avoid assignments of points not forming a valid plane the planarity is checked. An example of assignments that do not form a valid plane are points at the top of a façade that are erroneously assigned to the roof plane and therefore are distributed almost along one line (see Figure 6). The planarity is calculated using the eigenvalues λ_i of the variance-covariance matrix of these points as $(\lambda_2 - \lambda_3)/\lambda_1$ where $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and $\lambda_i \geq 0$ (West et al., 2004). Planes should have relatively large values of λ_1 and λ_2 and a small λ_3 . Only if the planarity is below a threshold, these assignments are accepted, and the corresponding fictitious observations are added to the optimisation.

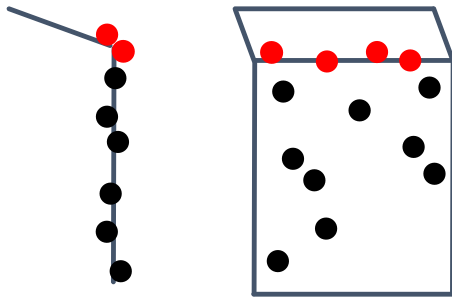


Figure 6: Sketch of a façade and a roof plane as cross-section (left) and in side view (right) with façade points (red) erroneously assigned to the roof plane, resulting in detections for the roof plane distributed almost along a straight line.

4.3 Indirect Assignment: Plane-Plane-Matching

With the indirect strategies, points that form a plane are searched in the tie point cloud instead of assigning single tie points to model planes. Two indirect strategies have been developed: The first searches for planes in the whole tie point cloud. The aim is to find model planes that correspond to the extracted planes. The other indirect method uses ROIs around model planes to restrict the plane search directly to a corresponding plane in the ROI of each model plane. Planes found in the

point cloud are referred to as *extracted planes*. Based on the detected 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. Compared to the direct assignment, this indirect approach only relates tie points to the model which together form a plane. The first indirect method requires the building model only in a second step to match the extracted planes to the model. The indirect assignment has been developed to check if it improves robustness of the assignment of tie points to model planes compared to the direct method.

Planes are detected in the tie point cloud using M-estimator SAmple Consensus (MSAC) (Torr and Zisserman, 2000). MSAC is a variant of RANSAC that weights inliers based on their fit to the model. Points closer to a plane, therefore, have a larger weight than points further away. 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.

4.3.1 Indirect Assignment without ROIs

This first variant of the indirect assignment applies MSAC to the whole point cloud of tie points to sequentially extract planes without considering the building model. 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. Alpha shapes are a generalisation of the convex hull: If $\vartheta_\alpha = \infty$, the resulting connected component is the convex hull of all points. Only points which are part of the largest connected component are kept as inliers to exclude points belonging to other structures. The inliers 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.

Inlier tie points with respect to the largest connected component are excluded from the tie point cloud, and in the remaining point cloud the next plane is searched. This is done until no more planes are found, or the number of points found per plane is repeatedly smaller than a threshold.

Figure 7 shows the tie point cloud (black) with a detected vertical plane (red) in top view. In stage 2, the connected component analysis of the plane points is shown, where two points are removed due to the gap to the other plane points. The connection (yellow in Figure 7) to the removed point is marked by a black cross. Stage 3 shows three found planes.

Only after the planes have been found, the building model is considered to find correspondences between the extracted planes and the model planes (stage 4). Multiple extracted planes per model plane are allowed. This reflects the fact that only parts of a larger generalised model plane might be found in the tie point cloud.

Each combination of an extracted plane and a model plane is checked for correspondence. For each such pair, an angle and a distance are computed that must be smaller than predefined thresholds for this pair to become a candidate for a correspondence. The angle is computed between the normal vectors, and the distance is the orthogonal distance of the centre of

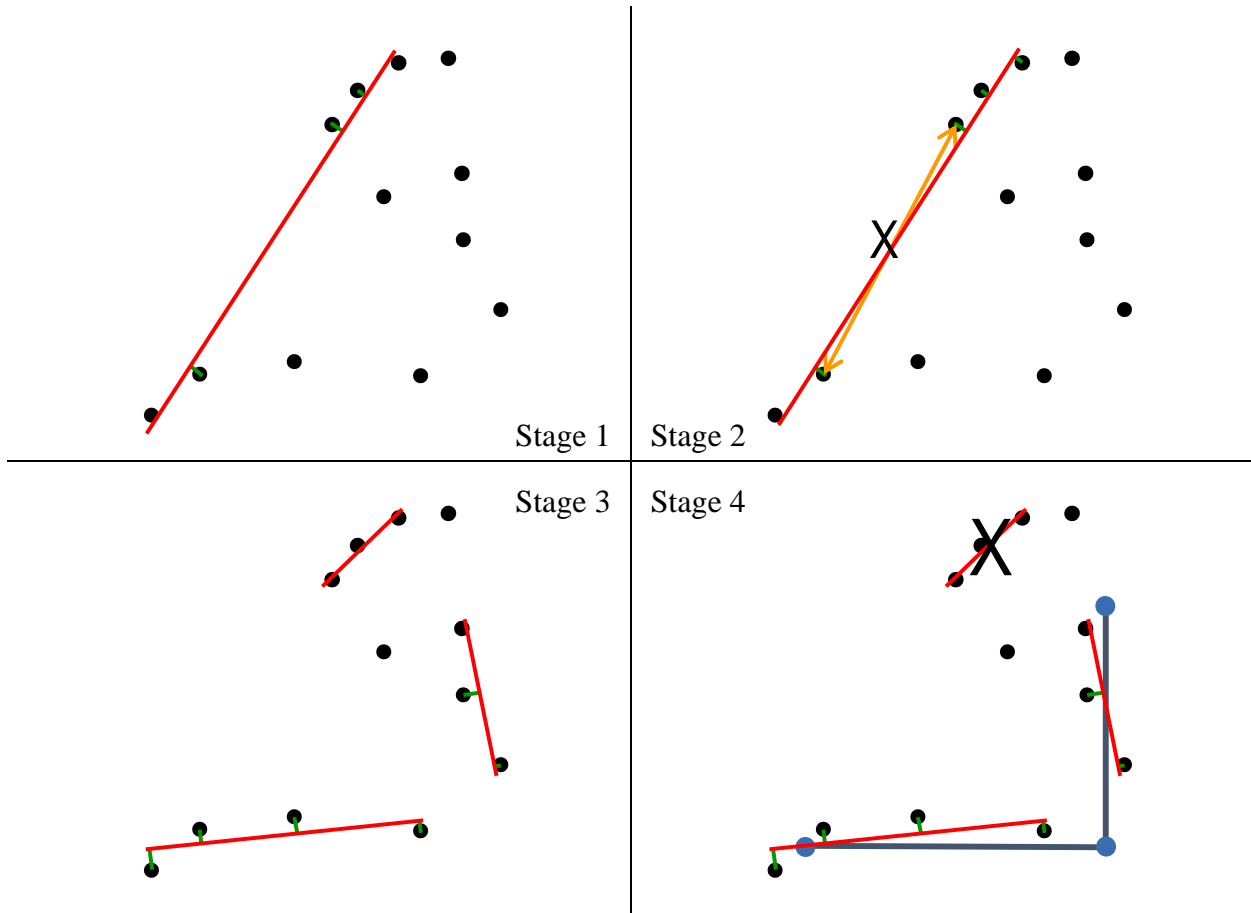


Figure 7: Sketch of the indirect assignment with the plane search without ROIs. Stage 1: First found plane (red) in the point cloud. Stage 2: Check of gaps between points of the plane using alpha shapes (yellow) where two points are removed from the plane. The parameters of this plane are recomputed based on the inlier points. All found plane candidates are shown in stage 3. In stage 4, the candidates are matched to the blue planes of the building model. One plane is removed as no corresponding model plane is found.

gravity (COG) of the boundary points of the extracted plane from the model plane. In Figure 7, two planes are shown that correspond to model planes, whereas one extracted plane is not assigned to any model plane (black cross, stage 4).

As an additional criterion, the polygon and the model plane must overlap. To check this requirement, the COG of the boundary points of the extracted plane is projected orthogonally to the model plane. If the projection of the COG lies outside the boundary polygon of the model plane the candidate is eliminated.

In the 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. Then, the remaining points are checked for the minimum number of points per plane and for their planarity as described in assignment method (a). If enough tie points are found inside the boundary of the model plane which fulfil the planarity constraint, they finally lead to fictitious observations that relate those points to the corresponding model plane.

4.3.2 Indirect Assignment with ROIs

In the second variant of the indirect assignment, for each plane of the building model, a corresponding extracted plane is searched for directly in the points in its vicinity (ROI). Therefore, all candidate points defined by the ROI are assumed to be potential candidates for the model plane. Figure 8 shows two model planes (blue) with their ROIs (green). No separate matching step between extracted planes and model planes is required. For each model plane, all points having a distance to that plane smaller than a given threshold are selected as being inside the ROI. Again, MSAC is employed to the points that are inside the ROI of each model plane individually with the configuration to only find planes with normal vectors that do not exceed a given angular distance to the normal of the model plane to reduce the search space. Similar to the indirect assignment without ROIs, only points inside the plane's boundary polygon that fulfil the planarity constraint are kept, and their number has to exceed the threshold for the minimum number of points per plane. No check for connected components is done, as the ROI already restricts the points to the vicinity of the model plane. Points of the extracted plane then lead to the fictitious observations relating tie points to model planes.

Points inside the ROIs are identified in the same way as in the direct assignment method, but the ROI potentially can be chosen larger than the distance threshold assuming that outliers are removed due to the subsequent plane search.

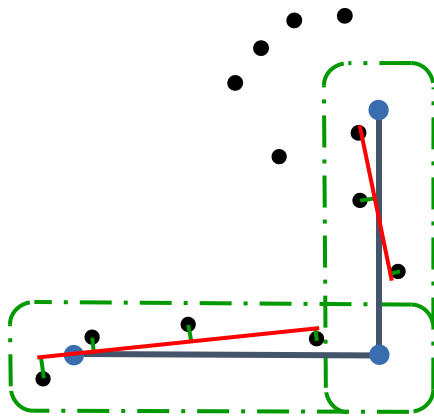


Figure 8: Sketch of the indirect assignment with plane search in ROIs. Planes similar to model planes are directly searched for only within the points closer than a threshold (ROI, green dashed-dotted line) to each plane. Points falling in ROIs of more than one plane are only used in the ROI of their closest plane.

4.4 Summary of the Assignment Parameters

The parameters for the assignment procedures are documented in the following overview.

Parameters common to all three assignment methods:

Minimum number of tie points per plane:

The indirect assignment methods (b) and (c) conceptually depend on groups of points describing a plane, as the assignment is based on the identification of planes in the point cloud. It would be possible to only assign one tie point to a model plane, as the plane is already defined by its vertices. To limit the influence of false assignments of tie points to model planes, assignments are only accepted if at least a minimum number of tie points is found to correspond to a model plane. A higher minimum number of points per plane prevents planes to be related to tie points if the planes are too small or are not covered by enough tie points. Setting this threshold too high prevents correct assignments.

Planarity:

The planarity is calculated using the eigenvalues of the variance-covariance matrix of the tie points assigned to a model plane. Regardless of the assignment method, the planarity constraint is used to only accept assignments to planes if the object points form a plane themselves. Cases in which the assigned points lie almost on a line or their orthogonal deviations from the plane are large are filtered out. Such cases would add correspondences that have a high probability of insufficiently describing a plane; see also the example of false detections along a line in Figure 6, section 4.2. The constraint threshold has to be chosen adequately to neither accept unstable planes nor exclude too many correct assignments.

(a) Parameters for the direct assignment:

Maximum distance:

This is the maximum distance to assign a point to a plane. The maximum distance in the first search for point-plane assignments is set to an upper value that has to be chosen according to the accuracy of the initial solution and the accuracy and degree of generalisation of the building model. In further iterations, the maximum distance is calculated as a multiple of the overall mean distance of all distances of assigned tie points to the model planes. If a lower limit is reached, the maximum distance is set to this lower limit. The lower limit has to be set according to the differences of the scene observed by tie points and the building model. Differences originate from the inaccuracy of tie point measurements and the building model, but are assumed to mainly depend on the amount of generalisation effects present in the building model. Setting the maximum distance too large increases the number of false assignments. If the distance is set too small, correct assignments are more likely to be missed. As described in section 4.2, the lower limit is set as $2 \cdot \sigma_{d_{TP}}$. The upper limit can be related to the generalisation effects of the building model and the accuracy of the

GNSS receiver, if direct observations for projection centre positions are used as absolute information.

(b) Parameters for the indirect assignment: Plane Search without ROIs

Maximum inlier distance:

A parameter of the plane search is the maximum distance of a point to a plane to be accepted as an inlier by MSAC. The maximum inlier distance has to be set according to the same considerations as for the lower limit of the maximum distance in the direct assignment.

Radius ϑ_α of alpha-shapes:

The radius ϑ_α defines how close tie points must be to each other to belong to a connected component of an extracted plane. The larger ϑ_α is set, the more points are part of the detected plane by accepting larger gaps between tie points. The parameter has to be set smaller than gaps between different planes with similar orientation: A gap due to a window in a façade should be bridged, but a gap from one building façade to the façade of a neighbouring building should lead to separate planes. The larger ϑ_α is chosen, the higher the risk becomes to include points of neighbouring buildings or structures that do not belong to the same model plane. Smaller values increase the risk to erroneously exclude points of the same plane due to gaps because of low texture areas or openings like doors, windows, etc.

Maximum distance of P_0 to a model plane:

Tie points as part of extracted planes are only assigned to model planes if the orthogonal distance of the extracted COG P_0 to the model plane is smaller than a threshold. The higher this threshold is chosen, the more corresponding planes are accepted, but the higher will be the risk of accepting wrong plane correspondences.

Maximum angle of normal vectors:

Tie points as part of extracted planes are only assigned to model planes if the angle between the normal vectors of the extracted and the model plane is smaller than a threshold. The higher this threshold is chosen, the more corresponding planes are accepted, but the higher will be the risk of falsely accepting plane correspondences.

(c) Parameters for the indirect assignment: Plane Search with ROIs

Maximum search distance:

Only points closer to a model plane than the maximum search distance are considered for the plane search. Comparable to the maximum distance of assignment method (a), the maximum search distance starts at an upper limit that has to be set according to the accuracy of the initial solution and ends at a lower limit that mainly corresponds to the size of generalisation effects to be expected. Compared to the maximum distance in assignment method (a) the setting of the maximum search distance does not directly define assignments but only acts as a threshold to

define points that are considered for the plane search with MSAC. If this threshold is chosen too small, plane points are falsely excluded from the plane search. Too large values are critical if points of non-corresponding nearby planes become dominant in the ROI and are identified as corresponding plane points.

Maximum inlier distance:

This is the same parameter as in method (b).

Maximum angle of an extracted normal vector to the normal of a model plane:

An additional parameter of MSAC is the maximum angle between normal vectors when searching for a plane with an orientation constraint in form of a roughly known normal vector: It only searches for planes with a normal vector that does not differ more than the specified angle from the given normal. The same considerations as for the maximum angle of normal vectors of method (b) apply.

5 Experiments

In section 5.1, this chapter defines the goals and setup of the investigation of the method introduced in chapters 3 and 4. The experiments are structured as scenarios with different setups of integrating building models and are run on subsets of one image sequence. Data acquisition is described in section 5.2, including an overview of the platform and sensors used to capture the data. Section 5.3 outlines relevant implementation details, including parameter settings. The results of the experiments are presented and discussed in chapter 6.

5.1 Setup of the experiments

The goal of this thesis is to integrate a generalised building model into the orientation of a monoscopic image sequence captured by a flying platform and to investigate if this integration improves the estimation of the platform pose and how generalisation effects influence the results. No model texture or visual descriptors of the captured scene and no ground control points are used in the process.

Experiments are set up to evaluate if an improvement of drift effects and deformations of the image block is achieved using the described hybrid bundle adjustment compared to not using a building model. In general, ground control points or ground control planes of, e.g., a detailed building model, improve the estimation of an image block. However, positive effects when using a building model with generalisation effects have not yet been shown. The experiments will show if pose errors can be reduced to the accuracy range and size of generalisation effects of the integrated generalised building model.

The experiments are set up with different scenarios, described in the next section, starting from a best-case with the best available building model. This best-case scenario is then degraded to scenarios with generalisation and systematic effects to a worst-case scenario that does not integrate any building model. The best-case scenario corresponds to a scenario given a detailed building model. The worst-case scenario on the other hand represents the application of a UAV flight along a street corridor given a starting pose and image measurements. The intermediate scenarios are the scenarios of interest according to the motivation of this thesis, as they integrate a widely available type of generalised building models.

For the purpose of comparison of all scenarios, check points are used to evaluate errors in terms of coordinate differences at known object points. From these check points, errors of the pose are inferred indirectly. For each scenario, systematic effects of the whole image block are analysed.

All experiments use subsets of the same image sequence (see section 5.1.2). One small subset is used to analyse all scenarios with respect to the adjustment as a post-processing step. The results of this adjustment are then compared to those of the sliding window workflow that potentially allows for pose estimation already during data capture on a longer subset of the image sequence. These experiments are done using the direct assignment method. In further experiments, the three introduced assignment strategies are compared to quantise their ability to set up the relations of tie points to model planes.

5.1.1 Scenarios

The following inputs are used in the experiments:

- Image coordinates of homologous points retrieved by image matching and orientation for an image sequence.
- Synchronised GNSS measurements for the 3D coordinates of image projection centre positions, if available.
- A generalised building model of the captured scene represented by the vertices of the buildings and a list of indices of vertices that define boundary polygons of the model planes. Due to systematic offsets of the original model, a corrected version of it is produced.
- A detailed building model: A model (in the same representation as the generalised building model) that is assumed to be at least as detailed and accurate as the photogrammetrically reconstructed object points. Ideally, all details of the scene represented by reconstructed object points should be represented in the building model, too.
- Check points: Well-marked points in the captured scene distributed all over the image sequence with known object coordinates a magnitude more accurate than the photogrammetrically reconstructed object points.

#	Image coordinates	GNSS Image Positions	Generalised Building Model	Detailed Building Model	Check Points
Scn1 (a, b, c)	x	all		x	x
Scn2 (a, b, c)	x	all	x (corrected)		x
Scn3 (a, b, c)	x	all	x		x
Scn4 (a, b, c)	x	beginning	x		x
Scn5	x	all			x
Scn6	x	beginning			x

Table 6: Inputs of the scenarios.

Table 6 lists the scenarios with their inputs set up for the experiments and Table 7 lists potential comparisons of the scenarios. Some scenarios only use GNSS measurements for the coordinates of projection centre positions for the first 50 key frames. The other scenarios use such GNSS measurements for all key frames, which allows for an investigation of the influence of the GNSS measurements on orientation results. There are six scenarios:

Scn1: Best case with a detailed building model.

Scn2: Generalisation effects and reduced systematic effects using a corrected version of the generalised building model.

Scn3: Uncorrected generalised building model.

Scn4: Uncorrected generalised building model and GNSS observations for projection centre positions only at the beginning of the image sequence.

Scn5: Scenario without any building model.

Scn6: Worst case without any building model and with GNSS observations for projection centre positions only at the beginning of the image sequence.

Scn1 vs. Scn2	Detailed vs. corrected generalised building model	→ Influence of generalisation effects
Scn1 vs. Scn3	Detailed vs. original generalised building model	→ Influence of generalisation effects in combination with systematic effects
Scn2 vs. Scn3	corrected vs. original generalised building model	→ Influence of systematic effects
Scn2 vs. Scn5	Corrected generalised building model vs. no building model	→ Result of using a generalised building model with reduced systematic effects
Scn3 vs. Scn5	Original generalised building model vs. no building model	→ Result of using a generalised building model
Scn3 vs. Scn4	All GNSS observations vs. only the first ones	→ Influence of GNSS observations in combination with a generalised building model
Scn5 vs. Scn6	All GNSS observations vs. only the first ones	→ Influence of GNSS observations without the usage of any building model

Table 7: Comparison of the scenarios.

The first scenario (Scn1) as an optimal case uses a detailed building model to support image orientation. The second scenario (Scn2) adds generalisation effects by only using a generalised building model. To investigate systematic effects of the planes of the generalised building model, the given model is corrected for systematic effects as described in section 5.2.2. Note that the remaining generalisation effects still consist of remaining local systematics. The influence of the generalisation effects is investigated by comparing Scn1 and Scn2.

The next scenarios (Scn3 and Scn4) use the uncorrected original generalised building model, which means that systematic as well as generalisation effects are present. Differences between Scn3/Scn4 and Scn2 are due to systematic effects. The comparison of Scn3/Scn4 to Scn1 reveals differences between using the original generalised building model and a detailed one. To analyse the influence of the GNSS observations of projection centre positions in combination with the building model, Scn3 uses all GNSS observations, while Scn4 only uses them in the beginning of the sequence.

Scenarios that integrate a building model are processed with the three assignment strategies and are therefore called 1a, 1b, 1c, 2a, etc. The direct assignment is addressed as (a), the indirect assignment without using the building model ROIs for the plane search as (b) and the indirect assignment with reduced search space by using the ROIs as (c).

Finally, the image orientation without the aid of any model with all GNSS observations (Scn5) and with using them only in the beginning of the sequence (Scn6) is investigated and compared to the scenarios that use a building model as additional information.

5.1.2 Sequences

Key frames of the same image sequence are used in different subsets (Table 8). The term “full sequence” refers to all frames of the used sequence. The full sequence covers a trajectory of 400 m and consists of 10812 frames resulting in approximately 535 key frames. The number of key frames varies between runs due to the usage of RANSAC procedures in the retrieval of relative orientation parameters and the elimination of outliers of the image matching.

Sequence	Scenarios	#Key frames	Length [m]
short	1, 2, 3, 5	~90	70
long	2, 3, 4, 5, 6	~340	320
full	2, 3, 4, 5, 6	~535	400

Table 8: Scenarios and parameters of the sequences.

The detailed building model is visible only in a small part of the full sequence as the given detailed building model only covers one building. The frames in which the detailed building model is visible form a subset which is addressed as the “short sequence” and consists of approximately 90 key frames covering a trajectory of 70 m. Consequently, Scn1 is only considered with the short sequence, which includes the detailed building model. For this short sequence, the computations are carried out according to the global workflow (see section 3.3.1). For Scn2 all 90 GNSS observations are used and Scn4 and Scn6 are not considered for the short sequence as the relevance of the GNSS measurements is investigated with a longer sequence.

Scenarios 2 - 6 are also run with the full sequence and a second subset of this full sequence of around 340 key frames, called the “long sequence”. The long sequence consists of all images of the full sequence except those at the beginning and in the end. The images not considered in the long sequence are influenced by unfavourable coverage of model planes and insufficient observations of homologous points. The full sequence with these potentially unstable parts is

investigated in comparison to the long sequence to further analyse the limitations of the presented approach in combination with these effects. For the full and the long sequences, the sliding window workflow (see section 3.2.5) followed by a global optimisation is carried out.

5.1.3 Evaluation

For the evaluation, measures of internal and external diagnostics (Förstner and Wrobel, 2016) are used. Internal diagnostics refer to measures derived from the adjustment itself, typically from the variance-covariance matrix. These measures represent the precision. Internal diagnostics cannot reveal deviations that occur because of systematic effects and correlations not modelled by the employed stochastic model. Such deviations can only be revealed by using external reference information. Measures that are derived using such independent external information are related to the accuracy and represent the external diagnostics.

Coordinate differences at check points lead to external diagnostics whereas a posteriori standard deviations represent internal diagnostics. Theoretically, if functional and stochastic models are correct and no systematic errors exist, external and internal diagnostics should lead to identical results.

Check Point Error

For all scenarios check points allow for a comparison of their estimated tie point coordinates $(\hat{X}_{CP}, \hat{Y}_{CP}, \hat{Z}_{CP})$ to precisely known coordinates (X_{CP}, Y_{CP}, Z_{CP}) as independent reference values. The difference of estimated and known check point coordinates (Eq. (10)) is addressed as *check point error* Δ_{CP} in this thesis. The check point errors and their root mean square (RMS) (Eq. (11)) are the main indicators used in the evaluation to compare the scenarios to each other and for sliding window and global workflow.

$$\begin{aligned}\Delta_{X_{CP}} &= X_{CP} - \hat{X}_{CP} \\ \Delta_{Y_{CP}} &= Y_{CP} - \hat{Y}_{CP} \\ \Delta_{Z_{CP}} &= Z_{CP} - \hat{Z}_{CP}\end{aligned}\tag{10}$$

$$RMS \Delta_{XYZ_{CP}} = \sqrt{\frac{1}{N_{CP}} \sum_{i=1}^{N_{CP}} (\Delta_{X_{CP_i}}^2 + \Delta_{Y_{CP_i}}^2 + \Delta_{Z_{CP_i}}^2)}\tag{11}$$

with N_{CP} = number of check points

Estimated Standard Deviation

As a measure of internal diagnostics, a posteriori standard deviations of the coordinates of the estimated check points $\hat{\sigma}_{CP}$ are computed and compared to the check point errors. Analogously to the notation of check point errors, standard deviations are denoted as $\hat{\sigma}_{X_{CP}}, \hat{\sigma}_{Y_{CP}}, \hat{\sigma}_{Z_{CP}}$ for specific coordinate axes and as $\hat{\sigma}_{XYZ_{CP}}$ for the three axes combined. The values of $\hat{\sigma}_{CP}$ result from the variances $\hat{\sigma}_{CP}^2$ on the diagonal of the variance-covariance matrix $\Sigma_{\hat{x}\hat{x}}$ as $\hat{\sigma}_{CP} = \sqrt{\hat{\sigma}_{CP}^2}$. $\Sigma_{\hat{x}\hat{x}}$ is derived

from the cofactor matrix of the unknowns $Q_{\hat{x}\hat{x}}$ (Eq. (12)) by multiplication with the empirical variance factor $\hat{\sigma}_0^2$ (Eq. (13)), according to Eq. (14). $Q_{\hat{x}\hat{x}}$ and $\hat{\sigma}_0^2$ are calculated using the Gauss-Markov adjustment described in section 3.2.

$$Q_{\hat{x}\hat{x}} = (A^T P A)^{-1} \quad (12)$$

$$\hat{\sigma}_0^2 = \frac{\hat{v}^T P \hat{v}}{n - u} \quad (13)$$

$$\Sigma_{\hat{x}\hat{x}} = \hat{\sigma}_0^2 Q_{\hat{x}\hat{x}} \quad (14)$$

Helmert Transformation

To further analyse datum effects that affect the whole image block, pairs of estimated and reference check point coordinates are used to calculate a 3D similarity transformation with three shift (X_t, Y_t, Z_t) and three rotation (eX_t, eY_t, eZ_t) parameters and one scale parameter s_t (Helmert transformation). The COG of the estimated check point coordinates is set as the rotation centre of the transformation.

5.1.4 Structure of the Experiments

The evaluation is organised as follows. First, the short sequence is analysed with respect to generalisation and systematic effects using the detailed, the corrected generalised and the original generalised building model and only assignment method (a).

Then, the long sequence is used to analyse generalisation and systematic effects in combination with block deformations. Here, no detailed building model is available. Deformations occur due to drift effects and varying quality of the configuration of tie points and model planes along the captured sequence.

Check point errors of the scenarios of the short sequence are compared to their estimated standard deviations to investigate the influence of the stochastic model with its assumptions on error distribution especially of the generalisation effects. Afterwards, the results of the global optimisation are compared to those of the sliding window approach with the long sequence to examine the impact of partial optimisation along the trajectory.

Up to that point, only assignment method (a) is investigated,. In further experiments, the long sequence is then also used to compare the results of the three assignment strategies to each other.

Finally, the full sequence is analysed with respect to the influence of unstable configurations in tie points and model planes on the results.

5.2 Dataset

The dataset required for the evaluation consists of an image sequence of a calibrated camera, captured by a UAV flying between buildings. For buildings, a generalised building model is available. Check points and a detailed building model for parts of the captured buildings are also available, as well as synchronised measurements of a low-cost GNSS receiver for the images. As the method should be able to cope with unfavourable GNSS conditions or signal loss, some experiments do not use all GNSS measurements. No benchmark data were found, that meet these criteria; own hardware was used to capture data and to set up the needed control information.

5.2.1 Hardware

- UAV

The UAV used for the experiments is a Matrice 100¹ of the company DJI. The quad-copter includes a low-cost GNSS receiver, IMU and barometer and has a weight of 2.5 kg including batteries. The maximum take-off weight is specified to be 3.4 kg. DJI states a hovering (position hold) accuracy of 0.5 m in vertical and 2.5 m in horizontal direction. In addition, the copter is equipped with DJI guidance (Zhou et al., 2015), which contributes to the relative positioning with a downward looking camera and serves as an obstacle avoidance system using stereo cameras and ultrasonic sensors to detect obstacles around and below the copter, to warn the pilot and to stop automatically before colliding with other objects.

GNSS measurements are downloaded from the copter's on-board memory in post-processing. The output binary files, amongst other data, contain time stamps and position information with a 100 Hz sampling rate that are retrieved using the software tool DatCon². No exact documentation of the outputs is available. Investigations of the trajectories for several flights lead to the assumption that no raw positions of the GNSS receiver are given but only filtered positions for the platform that lead to a smoothed trajectory. The retrieved values are probably a fusion of GNSS observations with barometer and IMU measurements. For the height component relative heights w.r.t. the starting position are used. The approximate height of the starting position is applied as an offset to all relative heights to retrieve heights approximately in DHHN2016.

- Camera

The camera used is the DJI Zenmuse X3³ with an EXMOR 1/2,3" CMOS sensor. The camera has a pixel size of 1.5 μm and a fixed-focus lens with 3.6 mm focal length, resulting in a horizontal field of view of 90°.

The camera is integrated into a three-axis gimbal that stabilises the camera in all three coordinate axes and is mounted on the copter. The gimbal is configured to smoothly follow the copter's heading (yaw, rotation around vertical axis) and keep the camera stable in roll and pitch axis. The weight of the camera, including the gimbal, is 262 g. The camera is powered by the

¹ <https://www.dji.com/de/matrice100>, accessed January, 2020

² <https://datfile.net/DatCon/downloads.html>, accessed January, 2020

³ <https://www.dji.com/de/zenmuse-x3>, accessed January, 2020

copter batteries and is connected to the copter's flight controller which allows for remote configuration and commands.

Image sequences are captured in 24 Hz video mode, as single image shots can be triggered shortest every 2 seconds, which is not enough for the scenario. From the 4k Full HD video, two frames per second with a size of 4096 x 2160 pixels are extracted in post-processing. These frames are the input image sequence to the algorithm developed in this thesis. Experiments showed that key frames extracted from the video sequence typically were not closer to each other than 1 second.

5.2.2 Data

The data set used for the experiments consists of several entities described in the subsequent paragraphs:

Image Sequence:

A flight was carried out with the UAV and camera described in section 5.2.1 in an area that consists of several different buildings with varying façade and roof characteristics in a distance of approximately 5 to 15 m from the buildings. The camera was mainly looking sideways relative to flight direction and almost orthogonal to the building façades, slightly tilted downwards to also capture roof parts from above. The height profile of the flight is given in Figure 9.

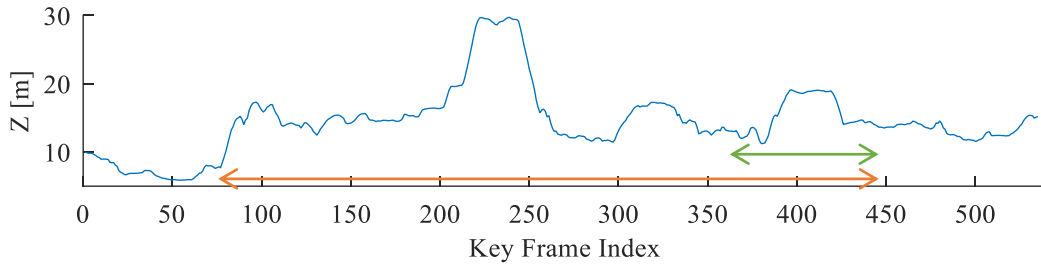


Figure 9: Height profile of the key frames of the full sequence. Key frames of the long sequence are marked in orange and key frames of the short sequence are marked in green.

The ground sampling distance (GSD) for 5 to 15 m object distance is 2 to 6 mm. Image overlap and base length depend on the choice of key frames. With a standard deviation of $\sigma_{TP} = 1 \text{ px}$ of measured image coordinates and a base length of 3 m, the precision of triangulated 3D tie point coordinates of an image pair in the photogrammetric normal case is 2 to 6 mm across viewing direction and 5 to 46 mm along viewing direction for 5 to 15 m object distance, respectively. This high precision shows that the image block can be expected to have a superior precision compared to the observations that define the datum, namely GNSS observations and planes of a generalised building model.

Without the building model, errors of the trajectory in the range of meters are likely to occur, due to only using images and GNSS measurements. Even larger errors can occur if GNSS becomes unavailable during data capture. For a generalised building model of LOD2, trajectory errors in the range of accuracy and generalisation effects of the building model - sub-metre down to decimetre level - are expected.

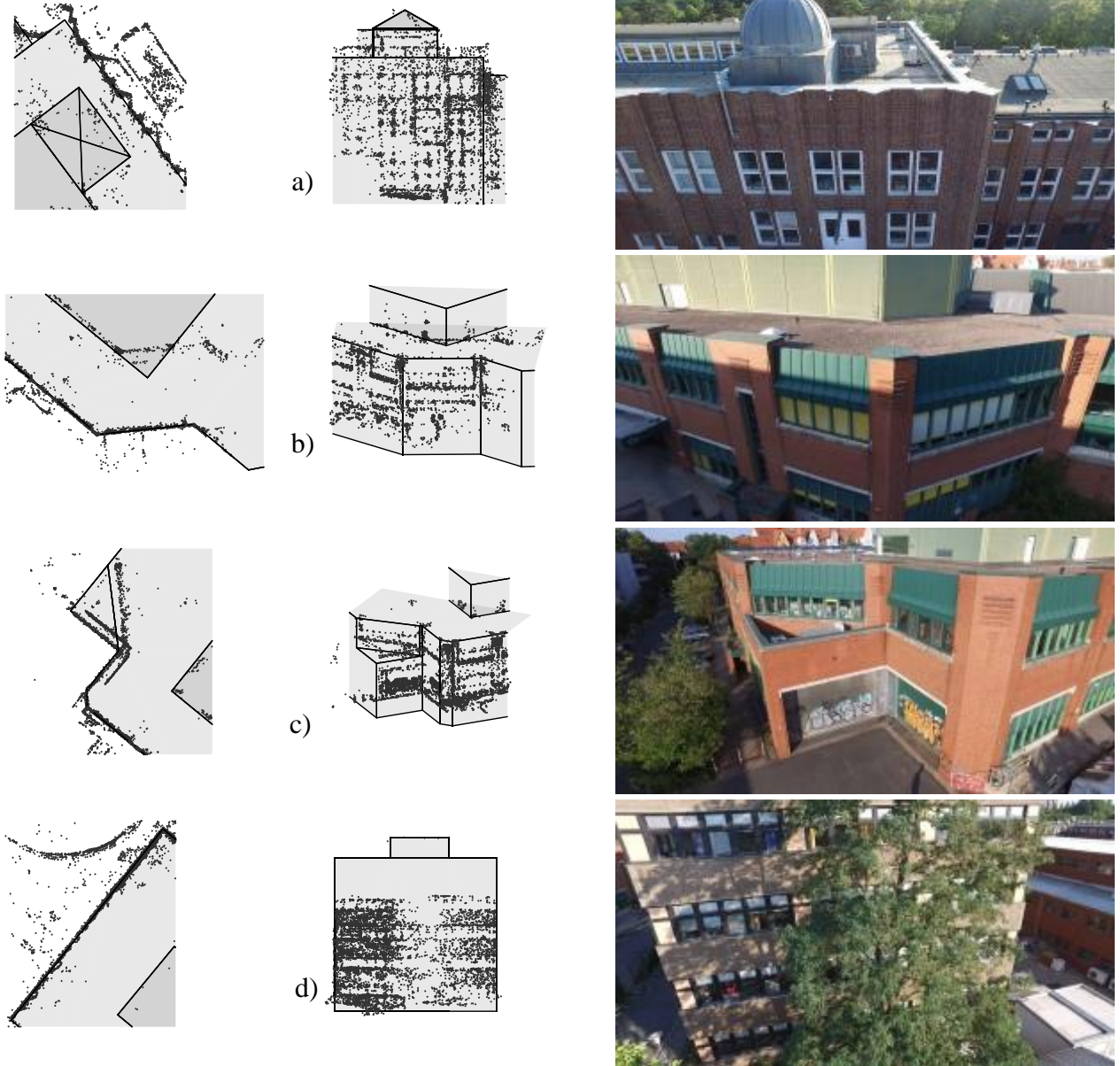


Figure 10: Examples of façade types covered by the image sequence with their corresponding representation in the generalised building model and reconstructed tie points (black points) in top (left column) and side views (mid column): a) a dome modelled in the generalised building model by only four triangles, a façade with spikes, windows and a flat roof with balustrades not present in the model; b) a building with various details not present in the model like roof top structures and roofs over building entries orthogonal to the façade; c) garage gates situated several decimetres behind the façade plane shown in the model; d) an office building with vegetation in front of the façade.

Figure 10 shows four examples of areas captured in the image sequence with corresponding images from the sequence and with the generalised building model and reconstructed tie points.

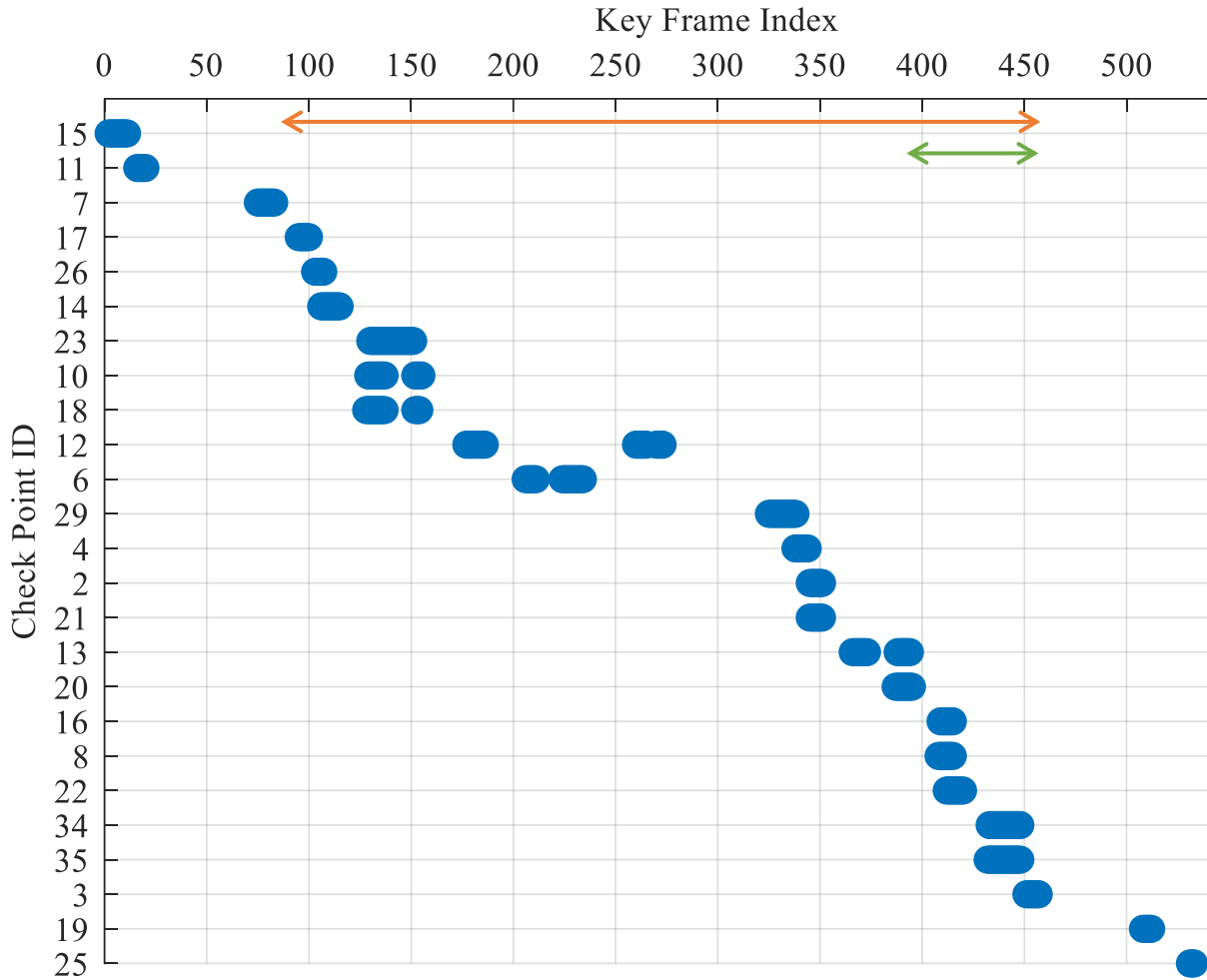


Figure 11: Visibility (blue) of check points in the key frames along the full image sequence. Check point 12 is observed by several key frames two times in the sequence, because a loop of 80 key frames was flown. Key frames of the long sequence are marked by an orange arrow and key frames of the short sequence are marked by a green arrow.

Check Points

37 check points are set up in the captured area in different heights on façades and at roofs. Mounts to magnetically hold corner cube reflectors or non-coded planar photogrammetric targets are glued to the buildings. The cubes were used in a tachymetric measurement campaign including precisely measured GNSS control points to retrieve check point coordinates in a network adjustment. The coordinate system is ETRS89/UTM zone 32N with heights in the DHHN2016. The coordinates are determined with an estimated mean standard deviation of 3.19 mm. During image data capture, the mounts are equipped with photogrammetric targets which are produced to have the same centre point as the corner cube reflectors in the mounts. The image coordinates of the check points were measured automatically using the commercial software Metashape⁴ (formerly PhotoScan). The results were refined manually by deleting obviously wrong measurements and adding some missed measurements. Check point identifiers were assigned manually to the targets, as these were non-coded.

⁴ <https://www.agisoft.com/>, accessed January, 2020

The visibility of check points within the full sequence consisting of about 535 key frames is shown in Figure 11. There are gaps in check point coverage of up to 50 key frames. Figure 12 shows the trajectory between key frames 170 and 270, where a loop along a large façade was flown, which is the reason why check point 12 is seen in the beginning and at the end of this loop. 25 of the 37 check points are seen in the full, 21 in the long and 8 in the short sequence. The orange arrow marks key frames belonging to the long sequence; the green arrow marks those belonging to the short sequence.

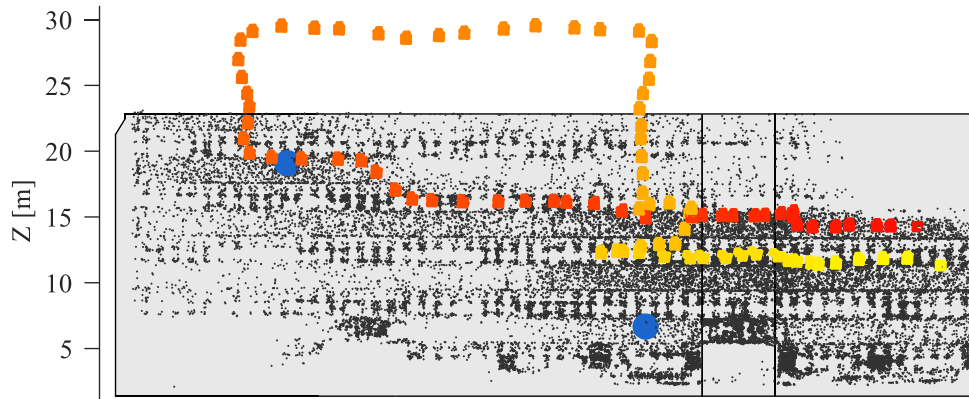


Figure 12: Orthogonal view of a façade (light grey). The positions of key frames 170 to 270 are shown in colour along the sequence from red to yellow, respectively. The façade is captured in a loop which is why check points 6 (blue, upper left) and 12 (blue, bottom) appear twice in the sequence. The tie points are shown in grey.

Generalised Building Model

The generalised building model is the publicly available LOD2 building model of the city of Hannover. It is published under Creative Commons 4.0 DE by the city of Hannover, Department of Planning and Urban Development, Department of Geoinformation. The model is given in the ETRS89/UTM zone 32N coordinate system with heights in DHHN2016, and the data format is CityGML 2.0. According to the documentation of the model, building footprints from the cadastre (Automatisierte Liegenschaftskarte (ALK)) are combined with an aerial laser scanning dataset for building heights and roof modelling. The height accuracy is specified as ± 1 m. The horizontal accuracy of the building footprints in ALKIS depends on the data acquisition method. The buildings used in this thesis were digitised from a paper map in a scale of 1:1000, which implies an accuracy in the decimetre range. Buildings with a footprint of at least 15 m^2 are represented in the model. A horizontal ground plane per building is set to the height of the lowest building corner. Therefore, per building, all vertices on the ground have the same height. The model vertex coordinates and the topology in form of a list of vertex indices per plane are extracted from the CityGML files to generate the input format of the building model described in section 3.1.

Figure 13 shows the generalised building model, the image positions of the full sequence and the check points seen by the full sequence in top and side view. In addition, the four example locations displayed in Figure 10 are indicated in the top view as a) to d).

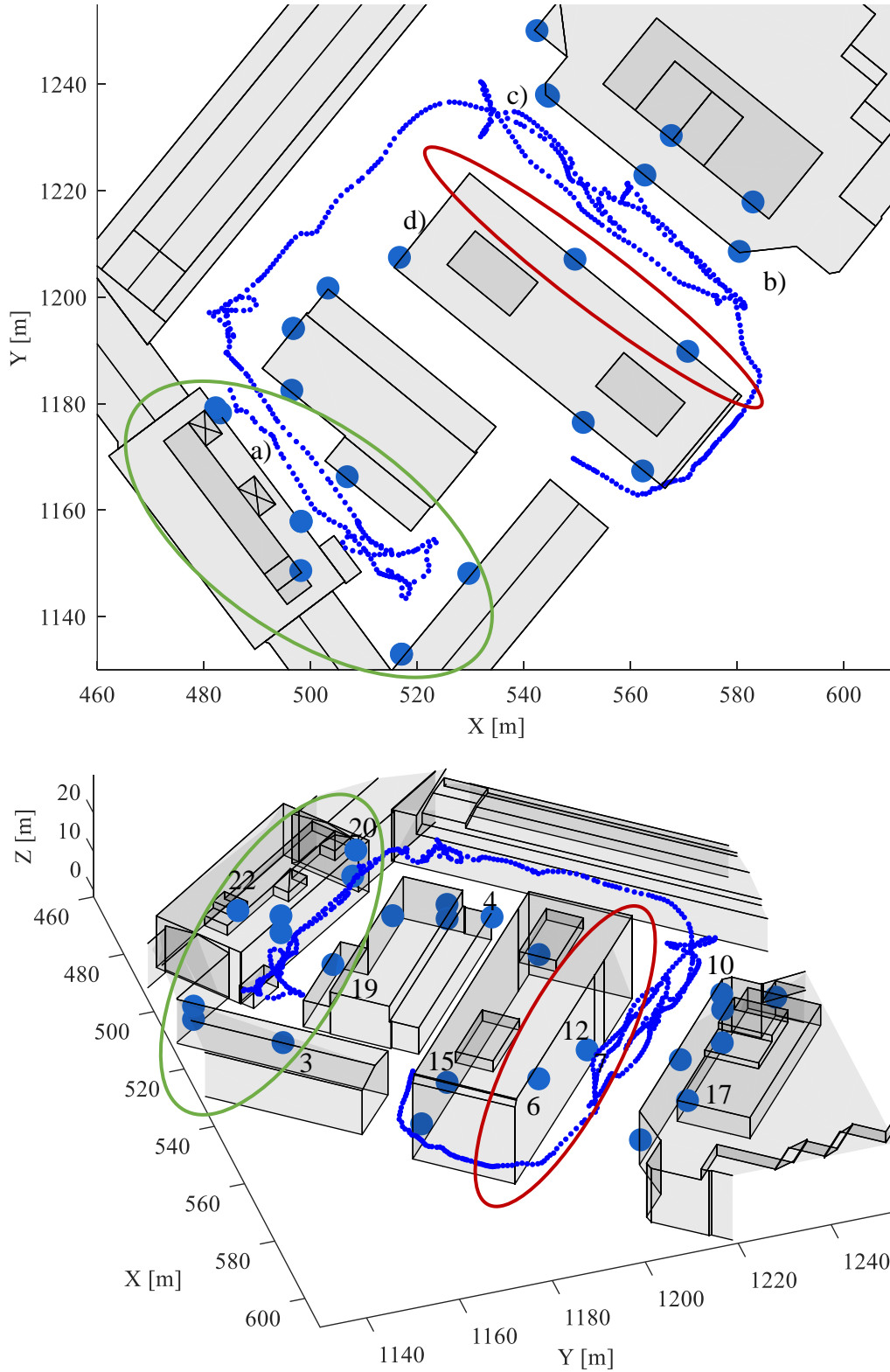


Figure 13: Key frame positions (blue) for the full sequence, generalised building model (light grey) and positions of check points (blue dots) in top and side view. A green ellipse marks the area selected as the short sequence. A red ellipse marks the façade shown in Figure 11. The example areas from Figure 10 are marked as a) – d) in the top view. The side view shows some check point IDs.

Corrected Generalised Building Model

To investigate systematic effects of the planes of the generalised building model, an image block with the measured (check) points as ground control points was set up as a reference. A hybrid adjustment with the generalised building model was then carried out, using all estimated tie points of this block as ground control points: Direct observations of the 3D tie point coordinates were introduced with high accuracy. The direct observations for coordinates of model vertices received a low a priori standard deviation of $\sigma_{VT} = 0.5$ m and assignment method (a) was used. This results in a corrected generalised building model where the planes of the generalised building model are fitted into the reference point cloud. The model's topology is not changed in that process as the vertices move according to plane corrections and as no planes are removed or added. Therefore, the level of generalisation stays unchanged, but systematic effects of the model planes are reduced. The correction of systematic effects mainly consists of shifts up to several decimetres per plane. The difference of normal vectors between original and corrected model planes mainly is below 1° .

Detailed Building Model

The detailed building model was produced based on architectural plans of the building footprints and façades to retrieve a 3D representation of much higher detail than the generalised building model. Part of the building roof was modelled based on laser scans and some distances along building façades were measured by tachymetry. The resulting detailed building model was shifted from a local coordinate system into ETRS89 coordinates using a dense point cloud from terrestrial laser scanning of parts of the building and the reference tie point cloud already used for the correction of the generalised building model. The laser scanning point cloud is given in ETRS89 with DHHN92 heights. The height difference of 3 mm between DHHN92 and DHHN2016 is negligible.

Remaining errors of the shift from local to ETRS89 coordinates and errors due to potential differences between the architect's plan and the real building were corrected using the hybrid bundle adjustment: The terrestrial laser scan and the reference tie point cloud were used to correct the building model analogously to the correction of the generalised building model. The parameters were set according to the accuracy and low degree of generalisation of the detailed building model. The topology was not changed. Remaining generalisation effects and uncorrected planes especially at the building roofs not covered by laser scanning or not modelled in detail due to missing details in the architectural data are to be expected in the produced detailed building model. Examples are non-planar roof surfaces and circular shapes of the domes shown in Figure 10a.

5.3 Parameter Settings and Implementation

The methods described in chapters 3 and 4 contain several thresholds and free parameters. They are summarised in the following, and standard settings used for the experiments are listed. The

first step is the retrieval of initial values for the adjustment by image orientation and point triangulation.

Image orientation (section 3.2.5):	Standard Value
Maximum number of best candidate feature points approximately uniformly distributed throughout the image to use for image matching	6000 points per frame
Minimum base length between last and current key frame	1 m
Maximum number of tie points per image (if more matches are found, the frame is skipped under the assumption to still have sufficient matches in the next frame, even if min. base is >1 m)	500 tie points
Minimum number of known tie points per new key frame (if less already known tie points are found, the minimum base length criterion is ignored to assure not to lose too many tie points due to skipping frames)	100 tie points

Tie point triangulation (section 3.2.5):	
Base to height ratio (excludes tie points that are too distant from the camera compared to the longest base length calculated for each tie point individually)	1/10
Minimum number of views (rejects tie points observed by fewer key frames due to redundancy considerations)	3 images

The parameters used for the sliding window processing and the adjustment are set as follows. The window size and overlap are chosen according to experience with several image sequences. They are set to be large enough to provide enough correspondences of tie points to different model planes in many situations. The smaller the window size is chosen, the higher will be the probability of not finding sufficient plane correspondence as absolute information to fix the datum and correct drift effects of the window image block.

Sliding window (section 3.3.2):	Standard Value
Window size (the more images are processed per window, the more plane correspondences potentially can be found, but the more processing resources will be needed)	$N_{ws} = 100$ images
Window overlap (the more already adjusted images are taken into account for a new window, the higher is the overlap and the lower is the risk of too large drift effects, but the more windows will have to be adjusted)	$N_{new} = 33$ images
Bundle Adjustment:	
Convergence limit (adjustment iterations stop if the relative change of the RMS of the residuals is smaller)	10^{-5}
Maximum number of iterations (adjustment iterations stop if this number of iterations is reached)	10 iterations

The setting of a priori standard deviations for the stochastic model follows assumptions about the accuracy of the different observation types. As no absolute accuracy of the GNSS observations is available, the standard deviation of observations of the projection centre positions is set to several meters with a lower assumed accuracy in Z-direction to give a relatively low weight to the GNSS observations.

In the experiments, the building model is nearly fixed by setting σ_{VT} to 1 cm. σ_{VT} is only released to .5 m to allow for model corrections to produce the detailed and the corrected generalised models.

The standard deviation of fictitious observations for tie points $\sigma_{d_{TP}}$ is set to 20 cm as this is the expected typical size of generalisation effects. $\sigma_{d_{TP}}$ is lowered to 5 cm for Scn1 as the detailed building model is assumed to have reduced generalisation effects.

Standard Stochastic Parameters (section 3.2.3)		
Image coordinates of tie points	σ_{TP}	± 1 pixel
Image coordinates of check points	σ_{CP}	± 1 pixel
Direct observations of projection centre coordinates	σ_{pos}	$\pm[3, 3, 5]$ m in X, Y, Z
Direct observations of building model vertex coordinates	σ_{VT}	± 0.01 m (± 0.5 m to correct the detailed and the generalised model)
Fictitious distances for tie points	$\sigma_{d_{TP}}$	± 0.2 m (± 0.05 m with the detailed model)
Fictitious distances for vertices	$\sigma_{d_{VT}}$	± 0.01 m

The parameters of the assignment method mainly depend on the accuracy and the amount of generalisation effects of the building model but also on the quality of absolute information used to set up the initial image block before the assignment is carried out. Therefore, the maximum point-to-plane distance, the maximum distance of P_0 and the maximum search distance of the ROIs of methods (a), (b) and (c) start at an upper and are reduced to a lower limit. The upper limit represents the larger differences between tie point cloud and model planes to be expected due to inaccurate datum information of the image block. The upper limit is set more restrictive, 2 m, for the direct assignment than for methods (b) and (c) as they already restrict the search to points forming a plane.

The lower limit mainly accounts for differences due to generalisation effects and is set to twice $\sigma_{d_{TP}}$, the standard deviation of the fictitious observations of distances that represent the assignments. In each iteration of the adjustment, the maximum distance is calculated as 3.5 times the overall mean distance of all distances of assigned tie points to the model planes. If this mean distance becomes smaller than the lower limit of the distance threshold, the maximum distance is set to its lower limit. The factor of 3.5 results from experience with several different image sequences.

In method (a), the assignment of tie points to model planes is directly limited by the maximum distance. Methods (b) and (c) have a maximum inlier distance that defines, how close to an extracted plane tie points have to be to become inliers of the MSAC-based plane search. As method (b) searches for planes in the whole tie point cloud, the threshold is set to 30 cm, which is more restrictive than the 40 cm for method (c).

For all three assignment methods, a minimum number of 15 tie points per model plane is set. Experiments with varied minimum numbers of points per plane were carried out. They suggested

that planes with fewer assigned tie points more often consist of erroneous assignments and do not positively influence the results of the hybrid adjustment, while larger thresholds might lead to missing correct assignments of planes covered by fewer points. Thresholds of the maximum angle of normal vectors, alpha shape radius and planarity are set based on the investigation of their influence with some example planes and plane parts that had to be removed as false assignments compared to a solution without the thresholds.

Point-Plane Assignment:	
(a) Direct (section 4.2):	
Maximum distance (upper and lower limit)	2 m (upper), 0.4 m (lower)
Minimum number of points per plane	15 points (5 for Scn1)
Planarity	0.001
(b) Indirect without ROIs (section 4.3.1):	
Maximum inlier distance	0.3 m
Radius of alpha-shapes	1 m
Maximum distance of P_0 to a model plane (upper/lower limit)	3 m (upper), 0.4 m (lower)
Maximum angle of normal vectors	10°
Minimum number of points per plane	15 points
(c) Indirect with ROIs (section 4.3.2):	
Maximum search distance (upper/lower limit)	3 m (upper), 0.4 m (lower)
Maximum inlier distance	0.4 m
Maximum angle of normal vectors	10°
Minimum number of points per plane	15 points

The method developed in this thesis was implemented in *Matlab*⁵ and *C++* code. Data handling, image matching, orientation and reconstruction as well as the assignment of object points to model planes are done in Matlab. Image matching, orientation and reconstruction are implemented based on the *Matlab vision* package. *Matlab Engine* is used to interoperate with the *C++* code. The hybrid bundle adjustment is based on a robust bundle adjustment as part of an in-house developed toolbox of the Institute of Photogrammetry and GeoInformation (IPI) at Leibniz University Hannover. As part of this thesis, the bundle adjustment was extended to handle a building model and the fictitious observations and to run in a sliding window manner. For the bundle adjustment, the toolbox uses the *SimplicialLLT* sparse matrix solver of the *Eigen*⁶ library.

⁵ <https://www.mathworks.com/>, accessed January, 2020

⁶ <http://eigen.tuxfamily.org>, accessed January, 2020

6 Results and Discussion

In section 6.1, this chapter compares the results for the short sequence referring to RMS coordinate differences of check points addressed as RMS check point errors Δ_{CP} . The influence of the building models of different quality onto pose estimation with a single building is investigated. In section 6.2, the results of the global adjustment of the long sequence with several different buildings are compared to the results of the short sequence. The influence of the generalised building model on pose estimation is further discussed. Block deformations due to systematic effects and insufficient absolute information are investigated with the estimated Helmert transformation parameters. Table 9 gives an overview of the sections that deal with the results of different combinations of scenarios, sequences and processing modes.

Section	6.1	6.2	6.3	6.4	6.5	6.6	6.7
Seq. Topics	Generalisation, Systematics	Generalisation, Systematics, Block Deform.	Int. / Ext. Diagnostics	Window versus Global	Assignment Strategies	The Full Sequence	Parameter Variations
Short Seq. Global	Scn1a, Scn2a, Scn3a, Scn5		Scn1a, Scn2a, Scn3a, Scn5				
Long Seq. Global		Scn2a, Scn4a, Scn5, Scn6		Scn2a, Scn4a, Scn5, Scn6	Scn2 (a,b,c), Scn4 (a,b,c)		Scn4a
Long Seq. Windows				Scn2a, Scn4a, Scn5, Scn6	Scn2 (a,b,c), Scn4 (a,b,c)		
Full Seq. Global						Scn2a, Scn3a, Scn4a, Scn5, Scn6	

Table 9: Overview of scenarios, sequences and sections of chapter 6.

In section 6.3, for the short sequence, the estimated standard deviations of check points are discussed in comparison to the check point errors to put measures of internal and external diagnostics into perspective. In section 6.4, results of the sliding window processing are presented and discussed in comparison to the results of the global adjustment shown in section 6.2. Until here, only the results of the direct assignment strategy are considered. In section 6.5, the three assignment strategies are investigated with respect to coordinate differences at check points. In section 6.6, the results for the full sequence with additional challenges are presented and discussed. Finally, variations of the essential parameters are analysed in section 6.7.

6.1 The Short Sequence: Generalisation & Systematic Effects

The short sequence, i.e. the part of the full sequence that consists of the images that observe the building represented by the detailed building model, is used to analyse the influence of the different types of building models on pose estimation. Only results with assignment method (a) are shown in this section.

Scn1: The best case with a detailed building model

The best case investigated for pose estimation is the usage of the detailed building model with reduced generalisation and systematic effects. This Scn1a leads to an RMS of check point errors $\Delta_{XYZ_{CP}}$ at the eight check points of this sequence of 10 cm.

Scn2: Generalised building model

In comparison to a generalised building model the detailed building model is expected to lead to improved results of the adjustment. Adding generalisation effects in Scn2a results in an RMS $\Delta_{XYZ_{CP}}$ of 19 cm. This means the errors almost double because of the generalisation effects. This improvement is assumed to be mainly related to generalisation effects.

Scn3: Generalised building model including systematic effects

Using the original generalised building model, which does not only have generalisation effects but also systematic effects at model planes (Scn3a), results in a higher RMS $\Delta_{XYZ_{CP}}$ of 25 cm for the short sequence. Note that generalisation effects that can be corrected as systematic effects per plane are likely adjusted in the corrected generalised model. This means that systematic and generalisation effects cannot be strictly separated in this discussion.

Scn5: Scenario without any building model

Figure 14 depicts these RMS $\Delta_{XYZ_{CP}}$ and additionally shows them in relation to the RMS check point errors of the scenario using GNSS observations for the image projection centre positions, Scn5. As expected, check point errors are largest when only using GNSS as absolute information; with <1 m they are smaller than the GNSS accuracy of 3 metres set by the a priori standard

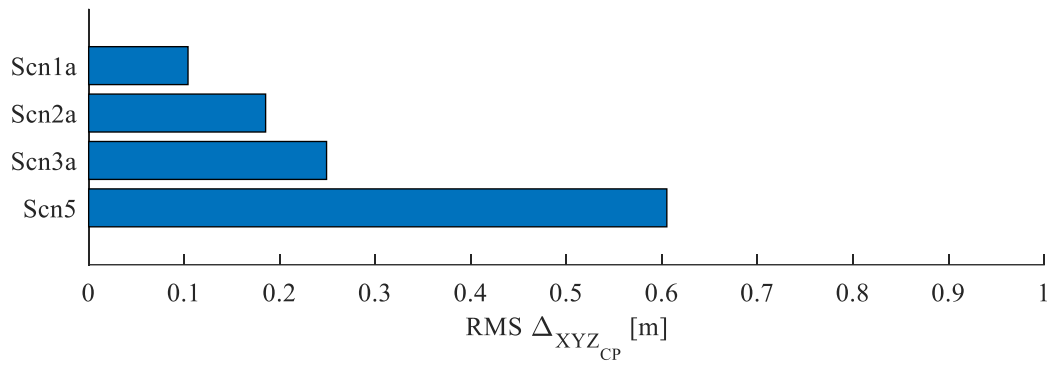


Figure 14: Comparison of the check point errors of the 8 check points per scenario of the short sequence.

deviation σ_{pos} . In the area of the short sequence, the flight was done approximately at building height, which leads to a relatively good GNSS configuration.

Figure 15 shows an example of the differences between the captured scene and the generalised building model. While the local detailed structure is preserved in the tie points, the points and thus also the image poses are moved closer to the building model as a result of the adjustment. In contrast to using hard constraints, the soft constraints described through the standard deviation of the fictitious distance observations prevents tie points e.g. at the prongs of the façade to be forced to lie exactly in the façade plane.

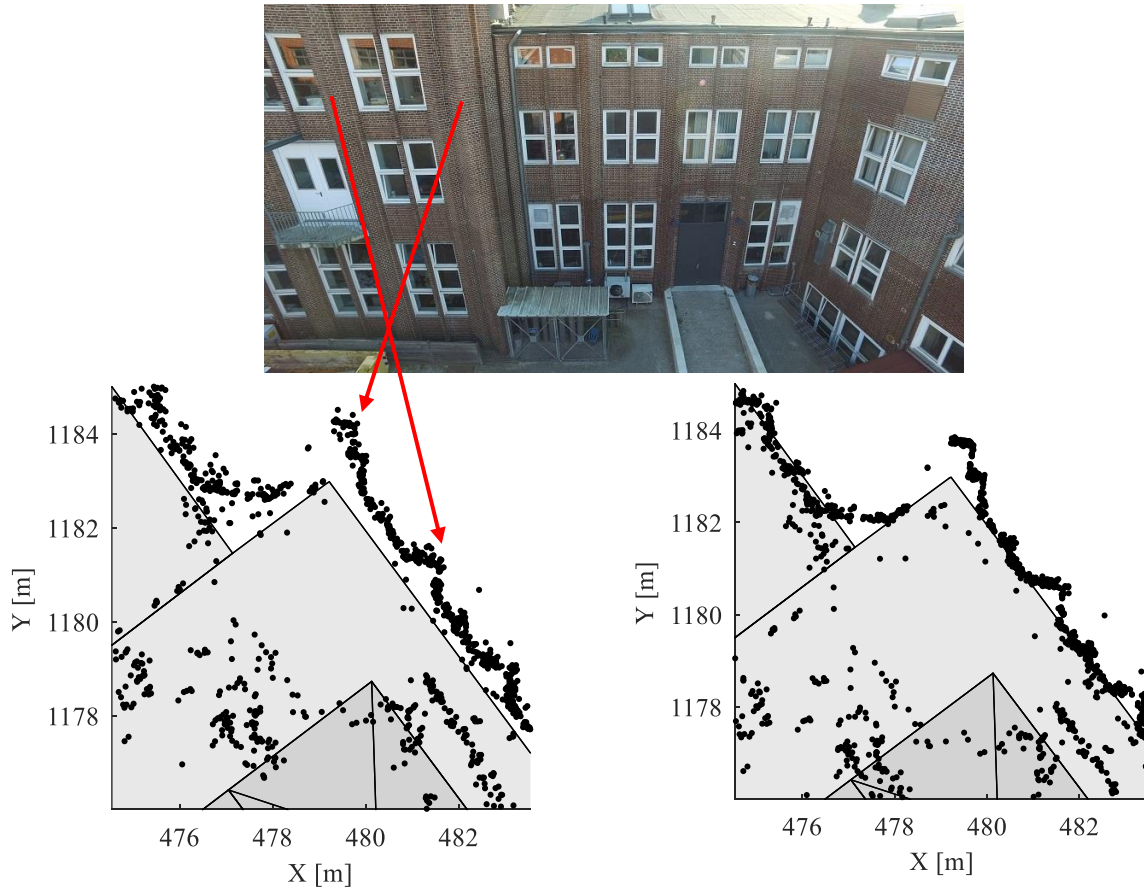


Figure 15: Combination of the detailed scene (black tie points) and the generalised building model (grey). Top: Frame of the image sequence. Bottom: Corresponding top view before (left) and after (right) adjustment. Prongs of the façade visible in the image highlighted by red arrows.

Discussion

Check point errors of around 25 cm with the generalised building model are in the expected range given the amount of generalisation and accuracy of the building model. The improvement to errors below 20 cm using the corrected version of the generalised model consequently is not very large but still shows that deviations of the generalised building model were lowered by correcting the model planes.

The decimetre level of check point errors with the detailed building model is not in the range of errors at the centimetre level predicted for the flight configuration. The detailed building model does not cover all details down to the accuracy level that can be expected for the photogrammetric block and the topology is not refined in the correction of the detailed model using the reference point cloud. Moreover, the details of the building present in the captured tie points is limited: Many planes of the detailed model are covered by less than three tie points, which is not sufficient to reconstruct those planes based on the observed tie points. The direct assignment method (a) used here theoretically can relate a single point to one plane, but the low coverage of the model with tie points still can pose problems: The smaller the planes are, the more likely a tie point is assigned to a close-by wrong neighbouring plane. Furthermore, as mentioned before, the lower the number of points per plane is, the higher is the influence of an outlier due to low redundancy. Assuming a regular distribution of tie points, for larger planes, fewer tie points might fall on the edges of the planes where the potential of wrong assignments to neighbouring planes is highest. Consequently, also for method (a), a minimum number of points per plane is established but lowered to 5 points for Scn1 instead of the 15 set for Scn2 and Scn3.

As described in section 5.2.2, the accuracy of the detailed building model at roof structures is limited. If split into the different coordinates, the main RMS check point error of Scn1 turns out to be in Z-direction with 9 cm while the X- and Y- directions are smaller with 3 and 4 centimetres respectively. Roof planes are the primary source of absolute support for image orientation in Z-direction. The captured façade consists of mainly vertical model planes which do not contribute to the Z-direction, i.e. the assigned tie points can freely move along vertical planes in Z-direction. Small planes in the façades of the detailed model, like window sills, theoretically could contribute to the Z-direction, but in most cases are not covered by enough tie points to obtain assignments that lead to fictitious observations of such planes. This behaviour can be seen in Figure 16, where the assignment results for one façade are shown with the detailed building model, the corrected and the original generalised building models.

As the assignment distance threshold is set smaller for the detailed building model than for the generalised model, fewer points are assigned to the horizontal roof plane (e.g. magenta coloured tie points for the corrected generalised model in the middle of the figure). Points not assigned to any plane of the generalised model are mainly part of the balustrade of the building's roof. The balustrade is represented in the detailed model only. In the top views, zigzag elements of the façade are visible in the tie point cloud and in the detailed model. In the generalised models,

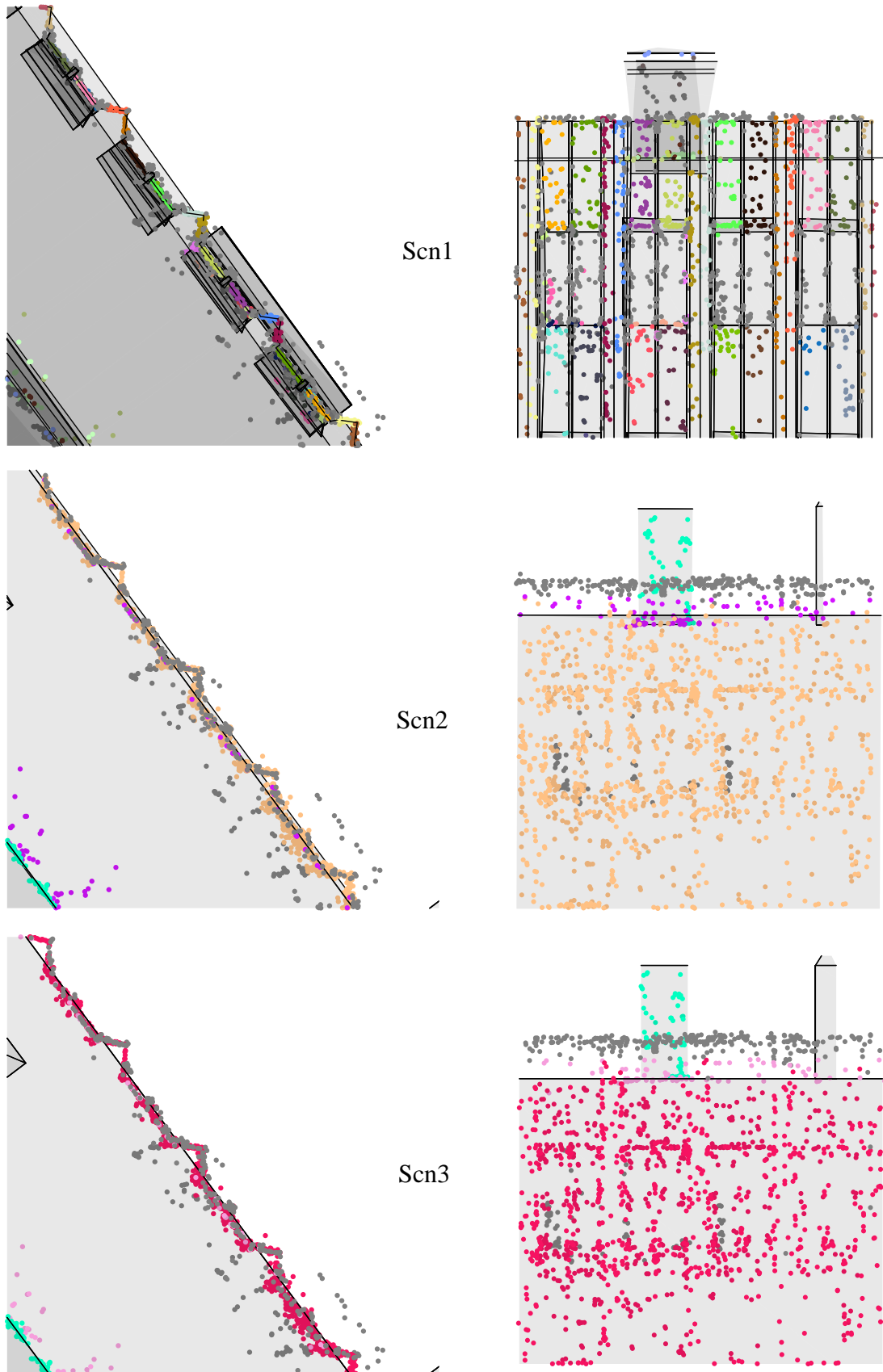


Figure 16: Short sequence: Comparison of the scenarios 1 (top of figure), 2 (mid) and 3 (bottom) at one façade in top (left) and orthogonal (right) views. Planes of the building models are shown in transparent grey. Assigned points are coloured per corresponding plane, unassigned points in grey.

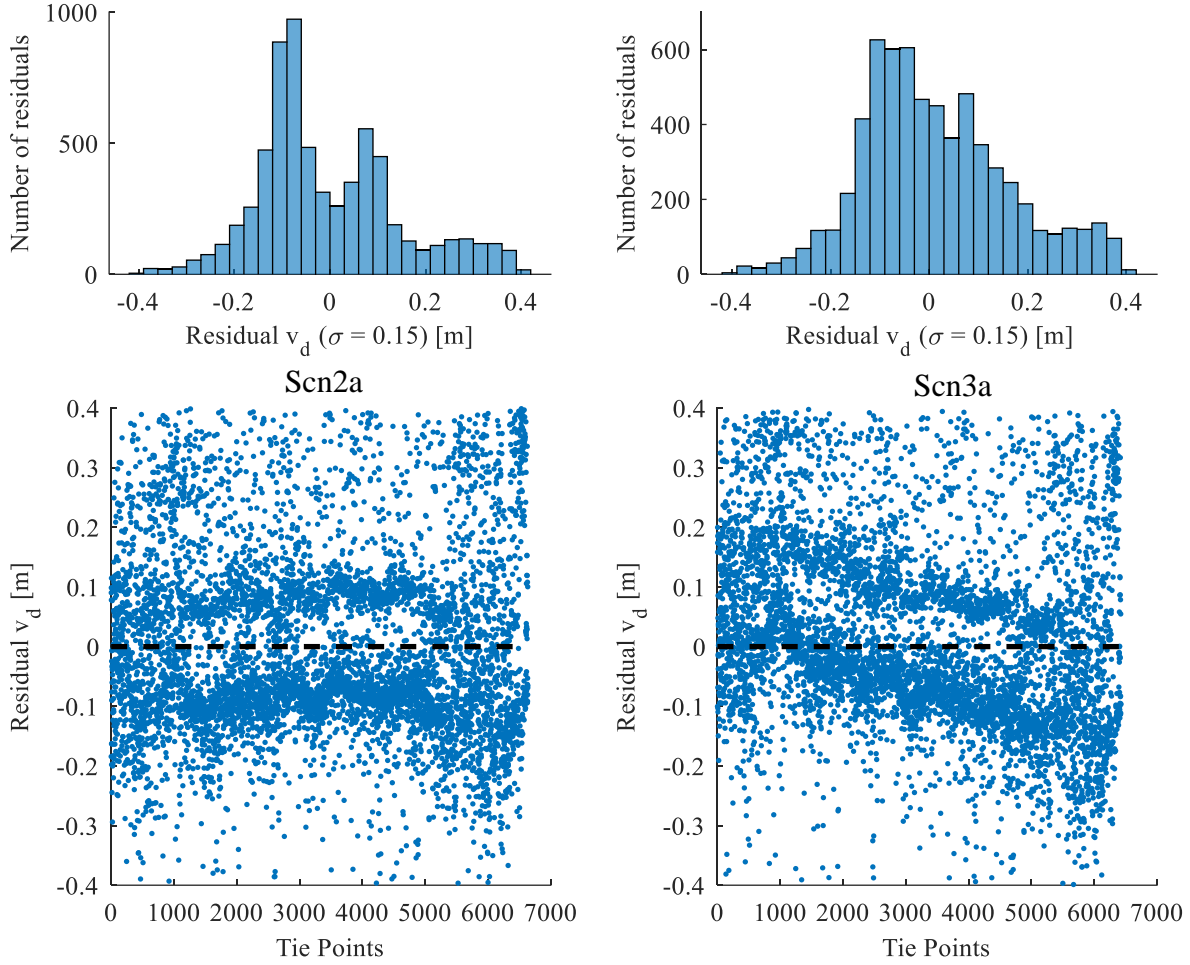


Figure 17: Top of figure: Histogram of the residuals of the fictitious observations linking tie points to the façade plane shown in Figure 16 as a result of Scn2a (left) and Scn3a (right) with the standard deviation σ of the residuals of this plane. Bottom: Corresponding residuals of the plane per assigned tie point ordered roughly along the image sequence.

they are not represented, but tie points on the prongs are assigned to the façade plane because of the distance threshold used in the assignment.

The example of the generalisation of the zigzag elements is further analysed for the residuals of the fictitious observations of this façade plane. Figure 17 shows histograms of the residuals (top) and the residuals per assigned tie point along the image sequence for the scenario with the corrected generalised building model (left) and the original generalised building model (right). For the histograms, standard deviations σ of the residuals are listed. With 15 cm, σ is close to the a priori standard deviation for the fictitious observations of tie points to generalised model planes of $\sigma_{d_{TP}} = 20$ cm. The residuals are limited to the maximum distance threshold of the direct assignment method, which was set to 40 cm. The histograms show a small number of residuals slightly larger than 40 cm, which is due to the change in estimated tie point coordinates during adjustment.

The image sequence was captured by flying sideways along the façade. Consequently, tie points at the main wall and at the prongs appear along the sequence (starting at tie point index 0) and lead to two groups of residuals around -0.1 and 0.1 m along the normal of the mean façade. The mean façade lies approximately at residual zero, which is depicted as a dashed black line in Figure 17.

In Scn2a, the tie points at the main wall and at the prongs lead to two peaks in the histogram. In Scn3a, the zigzag generalisation effects are mixed with a systematic effect (mainly a rotation) of the plane present in the original model. This systematic effect is corrected in Scn2a.

Note that the detailed building model only covers a small area with a very limited number of different façade and roof types. To further investigate the loss of accuracy due to generalisation effects, more scenes covered by a detailed building model would be needed.

Systematic effects

To investigate the systematic effects of the whole image block, a Helmert transformation between the known coordinates of check points and their estimated coordinates for each scenario is calculated (section 5.1.3). Figure 18 shows the check point errors after the adjustment (left column) and the remaining errors after transforming the estimated check point coordinates with the Helmert transformation per scenario for the short sequence.

Table 10 contains the corresponding transformation parameters per scenario and the RMS check point errors before and after transformation. For all four scenarios, an improvement in check point errors after the Helmert transformations is observed. For scenarios 2, 3 and 5, the RMS check point errors after the transformation are reduced to a comparable size of around 5-7 cm in X and Y and 1-2 cm in Z direction. The check point errors after transformation shown in Figure 18 show a similar distribution for these three scenarios.

Only for Scn1 that uses the detailed building model, even smaller errors around 2 cm in all three coordinate directions remain. Also the transformation parameters are smallest for Scn1; the transformation mainly reduces the RMS $\Delta_{Z_{CP}}$ by an offset in Z-direction. For all scenarios, the scale s_t is very close to 1. Check point errors after transformation are similar for scenarios 2, 3 and 5. This similarity means that with respect to block deformations there is no difference between using a generalised building model (Scn2, Scn3) and only using GNSS observations for projection centre positions (Scn5). Block deformations potentially present in Scn5 are not notably corrected using a generalised building model in the short sequence. The largest transformation parameters for Scn5 show nevertheless that this scenario, as expected, performs worse with respect to the datum of the image block than the scenarios that include a building model as absolute information.

Short Seq.	X_t [m]	Y_t [m]	Z_t [m]	eX_t [°]	eY_t [°]	eZ_t [°]	s_t	RMS $\Delta_{X_{CP}}$ [m]	RMS $\Delta_{Y_{CP}}$ [m]	RMS $\Delta_{Z_{CP}}$ [m]	RMS $\Delta_{X_{CP,t}}$ [m]	RMS $\Delta_{Y_{CP,t}}$ [m]	RMS $\Delta_{Z_{CP,t}}$ [m]
Scn1a	-0.01	-0.01	0.09	-0.05	0.11	0.08	1.00	0.03	0.04	0.09	0.02	0.02	0.01
Scn2a	0.03	-0.14	0.00	0.26	-0.19	0.03	1.00	0.08	0.16	0.04	0.06	0.07	0.02
Scn3a	-0.04	0.16	0.07	0.15	-0.04	-0.23	0.99	0.09	0.22	0.08	0.05	0.05	0.01
Scn5	-0.15	0.06	0.49	-0.62	0.34	0.17	0.99	0.27	0.20	0.50	0.05	0.06	0.01

Table 10: Parameters ($X_t, Y_t, Z_t, eX_t, eY_t, eZ_t, s_t$) of the Helmert transformations between known and estimated check point coordinates for some of the scenarios of the short sequence and RMS of coordinate differences of the known and the estimated check point coordinates ($\Delta_{X_{CP}}, \Delta_{Y_{CP}}, \Delta_{Z_{CP}}$) as well as of the known and the transformed estimated check point coordinates ($\Delta_{X_{CP,t}}, \Delta_{Y_{CP,t}}, \Delta_{Z_{CP,t}}$).

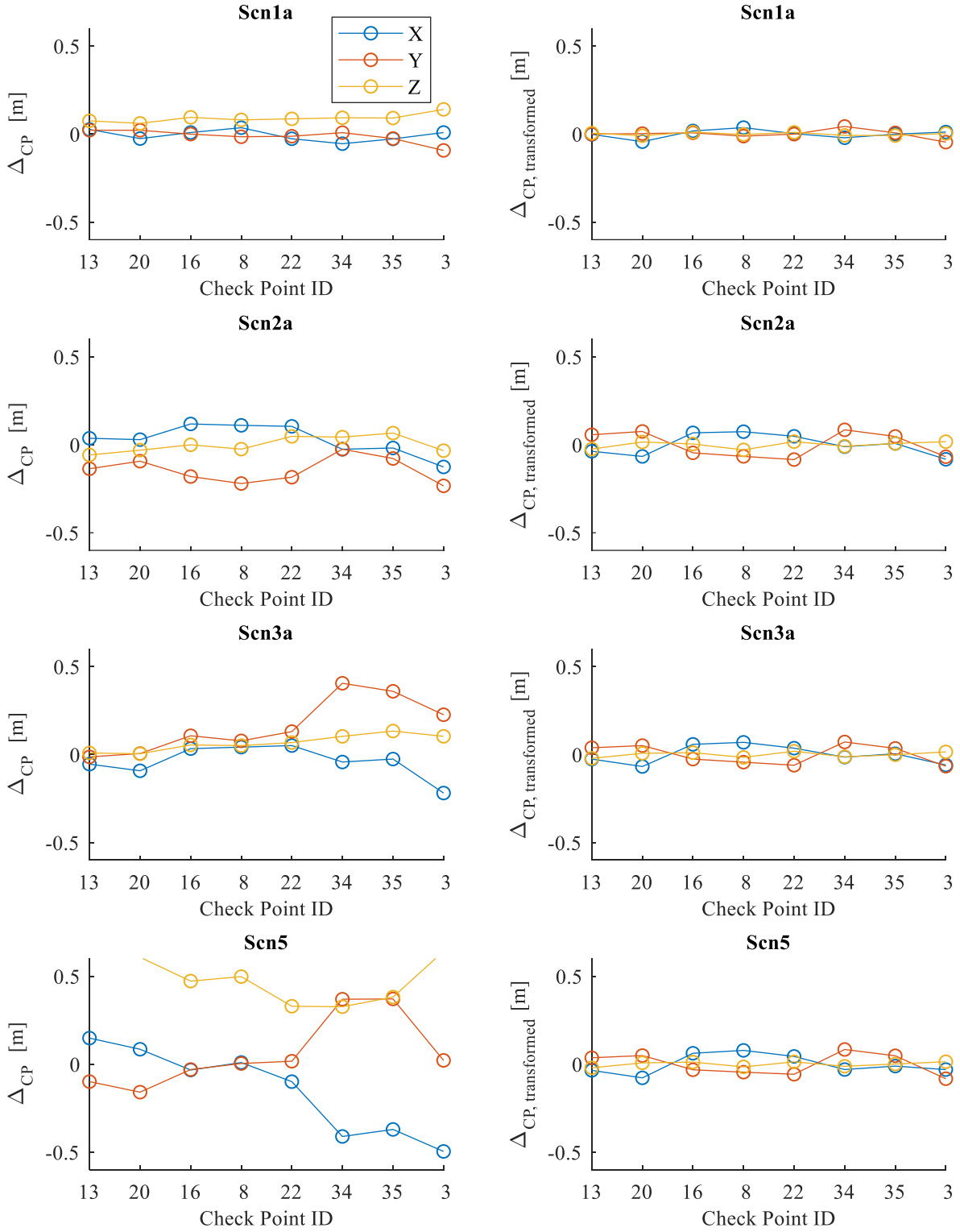


Figure 18: Short sequence: Check point errors per check point in the order of their appearance in the sequence before (left) and after transformation (right).

Block deformations are so small in the short sequence that an improvement is observed only with a detailed building model (Scn1). The smaller remaining check point errors for Scn1 after transformation indicate corrections of block deformations which were not corrected in scenarios 2, 3 and 5.

The remaining errors after transformation $\Delta_{CP,t}$ for Scn1 are in the range of errors to be expected for the photogrammetric block. This was verified by using all check points except one as ground control points in bundle adjustments without any building model or GNSS observations. The adjustment was repeated for every check point with all others as ground control points. The RMS check point error over the adjustments with < 3 cm reflects the accuracy of the transformed result $\Delta_{CP,t}$ of Scn1.

6.2 The Long Sequence: Generalisation & Systematic Effects, Block Deformations

With 350 key frames the long sequence covers a higher variety of building parts and due to the longer trajectory allows for a more detailed analysis of generalisation and systematic effects. Another effect to be analysed with the long sequence are block deformations and drift effects along the image sequence that depend on the availability of GNSS measurements for projection centre positions and the usage of a building model. Block deformations and drift effects are assumed to be limited when integrating the generalised building model. Largest block deformations are expected if neither a model nor GNSS measurements are used to support the pose estimation. To investigate the influence of the GNSS measurements on drift effects and block deformations, Scn4 and Scn6 are defined where only the GNSS measurements of the 50 first key frames are used to initialise the datum. These scenarios allow for the analysis of block deformations in image orientation with respect to the availability of a generalised building model (Scn3 & Scn4 versus Scn5 & Scn6) and GNSS measurements (Scn3 versus Scn4 and Scn5 versus Scn6).

Scn3 and Scn4 lead to similar results: As soon as a building model is used in the investigated image sequence, GNSS measurements are not relevant anymore – they are nevertheless needed for the initialisation of the datum. Note that GNSS measurements of the image positions could become relevant in windows in which problems with the relation of tie points to model planes occur (e.g. unfavourable configuration of the planes or a low number of assigned planes). In such cases, GNSS measurements could contribute to bridge gaps in the sequence, where no relations to a building model were found. As the results of Scn3 and Scn4 are similar, only the results for Scn4 are shown in the following.

Figure 19 depicts the RMS check point errors after the global hybrid adjustment (step 7 of the workflow of the sliding window pose estimation in Table 5). Values of the horizontal axis at bars that exceed the figure are written at the end of the bars.

While for the short sequence, the difference between using the corrected generalised building model (Scn2) and the original version (Scn4) was small (20 versus 25 cm), with the long sequence this changes: Including the corrected generalised building model (Scn2) leads to RMS check point errors of 12 cm after the global adjustment. With the original generalised building model (Scn4), RMS $\Delta_{XYZ_{CP}}$ of 31 cm result. Compared to the 25 cm achieved using the original generalised

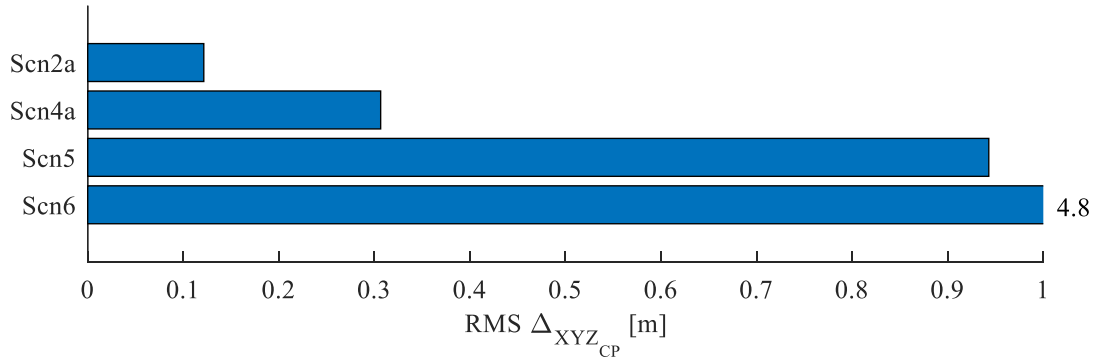


Figure 19: Long sequence: RMS check point errors of all check points for the global adjustment.

building model in the short sequence, the 31 cm for the long sequence indicate slightly larger check point errors over all captured buildings than for the building captured in the short sequence.

If no building model is used, check point errors are in the metre range (Scn5) and reach an RMS of almost 5 m if GNSS is used only at the beginning of the sequence (Scn6).

Discussion

The RMS check point error achieved using the corrected generalised building model (Scn2a) with the long sequence is smaller than the error for Scn2a in the short sequence. This indicates that after correcting for systematic model errors, smaller effects remain in the building model compared to the building part captured by the short sequence. At the building captured by the short sequence, stronger effects (e.g. at the prongs) due to generalisation of the building model occur than on most other buildings captured by the long sequence.

As expected according to GNSS accuracy, the RMS check point errors of Scn5 and Scn6 are in the metre range and larger errors occur when only using GNSS at the beginning of the sequence. Compared to the short sequence, the error is larger for Scn5 in the long sequence. This shows that the quality of GNSS measurements regarding factors like signal reception, geometry and multi-path effects is not as good along the long sequence as it is at the short sequence. The long sequence features higher buildings and lower flying heights than the part covered by the short sequence, where most images are captured flying at roof height.

Using the generalised building model leads to the same RMS $\Delta_{XYZ_{CP}}$ regardless if GNSS measurements of projection centre positions are used for all key frames or only at the beginning of the image sequence. In the scenarios 5 and 6 that do not use a building model, the difference with respect to these GNSS measurements is clearly visible. Therefore, it is concluded, that the hybrid bundle adjustment allows for the replacement of missing GNSS observations with the generalised building model.

The estimated tie points for Scn4a are shown in Figure 20 coloured by the planes they are assigned to. Some large façade planes have many points assigned to them, but also roof planes covered by much fewer points are related to the tie point cloud.

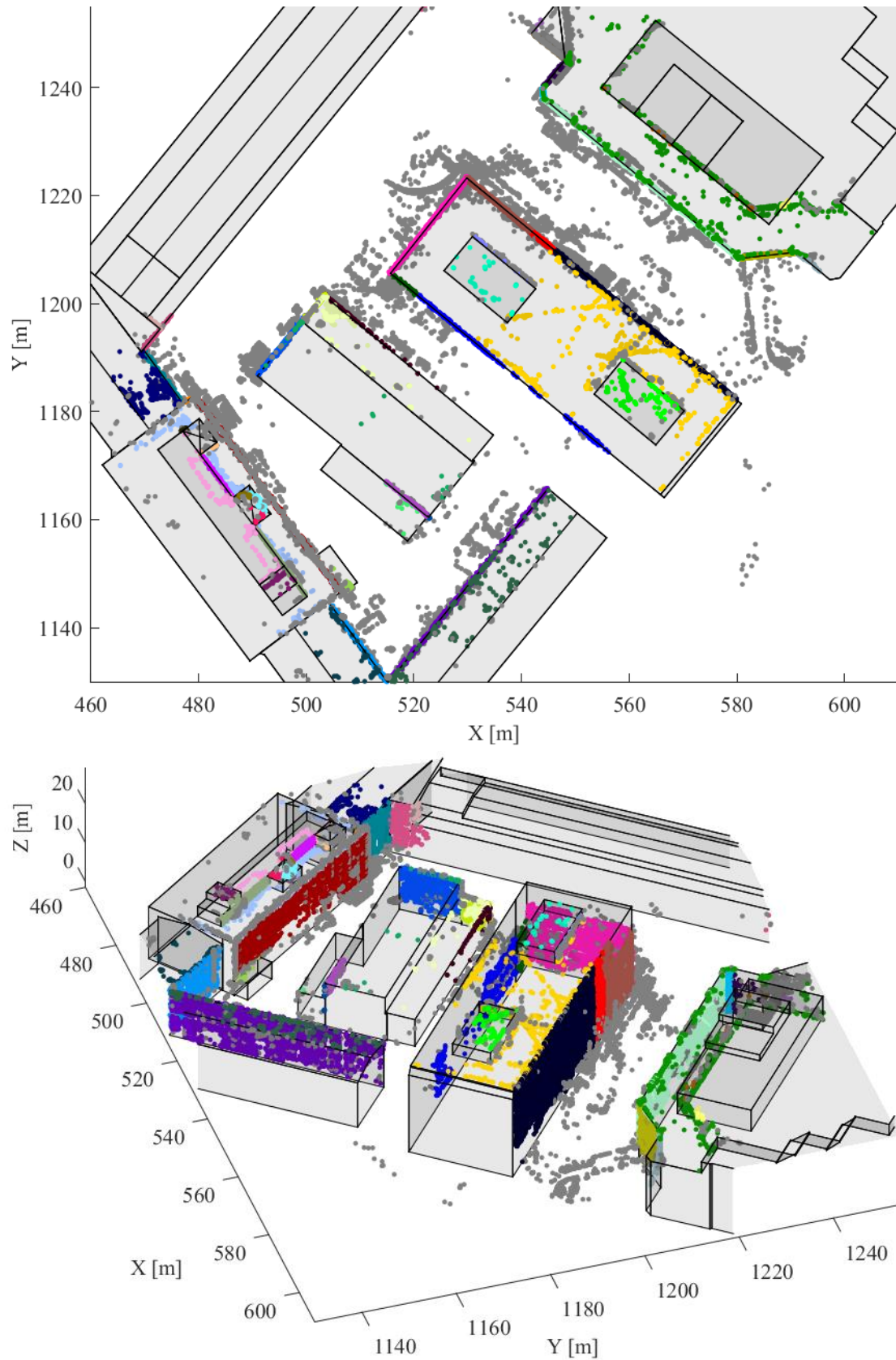


Figure 20: Generalised building model with tie points of Scn4a of the long sequence in top and side views. Tie points not assigned to a plane are shown in grey, colours indicate assignments to specific planes.

Systematic effects

As described for the short sequence, pairs of known and estimated check point coordinates of the global adjustment of the long sequence are used to calculate parameters of Helmert transformations for each scenario. The check point errors before and after transformation are shown in Figure 21 for Scn2 and Scn4. Table 11 shows the transformation parameters and the RMS check point errors.

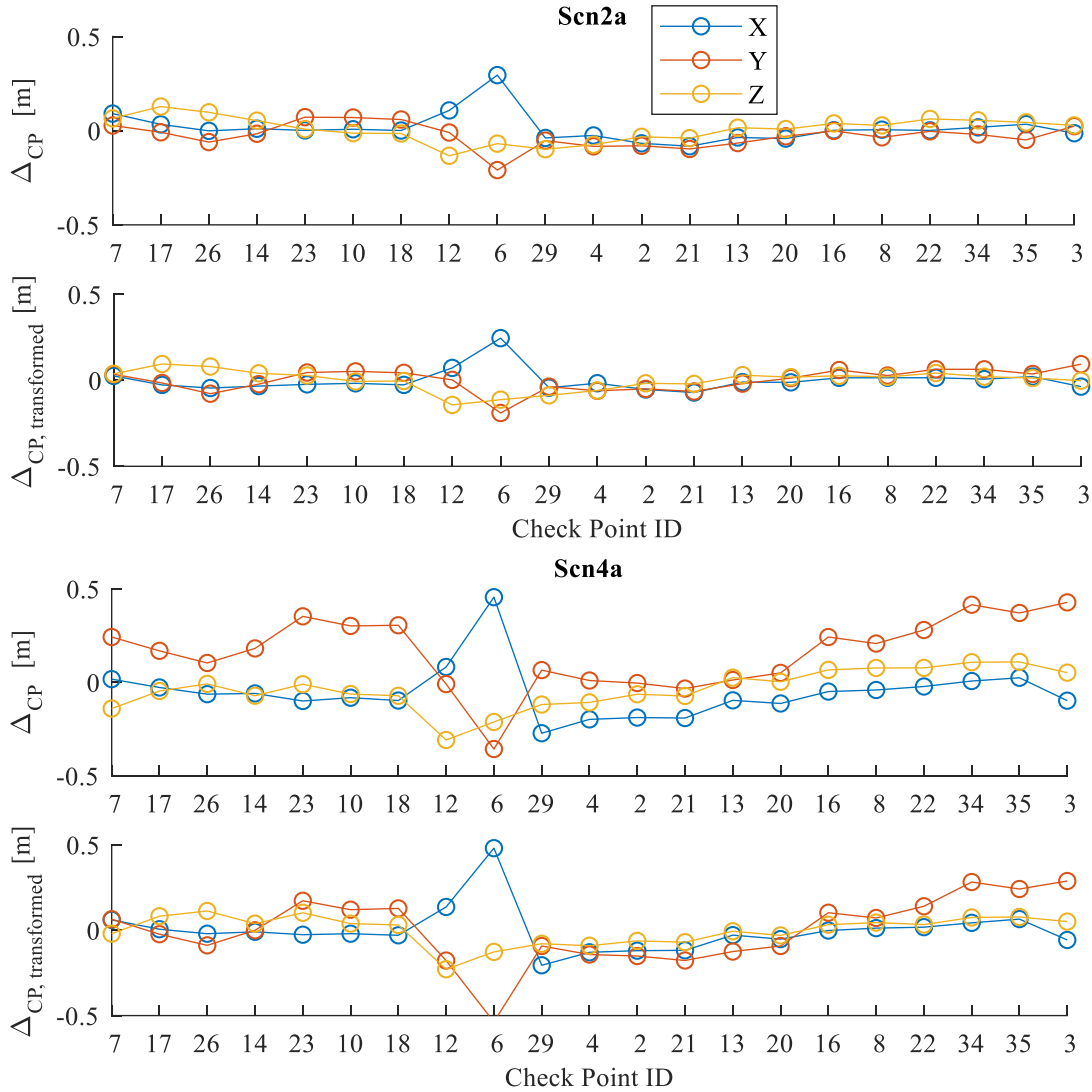


Figure 21: Long sequence: Check point errors per check point in the order of their appearance in the sequence before and after transformation for Scn2a and Scn4a.

Using the corrected generalised building model in Scn2, the check point errors and the transformation parameters are smaller than with the original generalised model in Scn4. The transformation parameters for Scn2 show a maximum of -3 cm for the Y-shift y_t and 0.04° for the Y-rotation eY_t . The largest transformation parameter for Scn4 is the shift of 16 cm in Y-direction.

Relatively large errors are revealed for check point 6, which lies at the top left of the large façade captured in a loop (see Figure 12). Figure 22 shows the check point errors for Scn4a plotted as arrows in top view coloured according to the Z-error. Check point 6 is shifted along the model plane of the façade shown in Figure 20 with the assigned tie points in black. The direction of the

Long Seq.	X_t [m]	Y_t [m]	Z_t [m]	eX_t [°]	eY_t [°]	eZ_t [°]	s_t	RMS $\Delta_{X_{CP}}$ [m]	RMS $\Delta_{Y_{CP}}$ [m]	RMS $\Delta_{Z_{CP}}$ [m]	RMS $\Delta_{X_{CP,t}}$ [m]	RMS $\Delta_{Y_{CP,t}}$ [m]	RMS $\Delta_{Z_{CP,t}}$ [m]
Scn2a	0.01	-0.03	0.01	0.03	0.04	-0.01	1.00	0.08	0.07	0.06	0.06	0.06	0.06
Scn4a	-0.05	0.16	-0.04	0.05	-0.07	-0.02	1.00	0.15	0.24	0.11	0.13	0.19	0.08
Scn5	0.24	-0.11	0.31	0.88	-0.13	0.53	1.00	0.50	0.44	0.67	0.27	0.34	0.13
Scn6	0.44	0.25	2.96	4.34	-2.51	0.74	0.99	0.63	0.76	4.66	0.29	0.34	0.13

Table 11: Parameters ($X_t, Y_t, Z_t, eX_t, eY_t, eZ_t, s_t$) of the Helmert transformations between known and estimated check point coordinates for some of the scenarios of the long sequence and RMS of coordinate differences of the known and the estimated check point coordinates ($\Delta_{X_{CP}}, \Delta_{Y_{CP}}, \Delta_{Z_{CP}}$) as well as of the known and the transformed estimated check point coordinates ($\Delta_{X_{CP,t}}, \Delta_{Y_{CP,t}}, \Delta_{Z_{CP,t}}$).

check point errors indicates that for the images captured by flying a loop along the façade mainly the façade plane contributed as absolute information along the plane's normal direction. Tie points could move orthogonally to the normal within the plane as there are almost no planes that add absolute information in the directions along the plane. Only some images of the loop capture roof planes of the building adding some fictitious observations that contribute to the Z-direction. As the area around check point 6 is the longest part of the sequence going along just one façade, the large error at this check point indicates that block deformations are dispersed to this area.

The check point errors before and after Helmert transformation are shown in Figure 23 for Scn5 and Scn6 that do not integrate a building model and only use GNSS measurements of projection centre positions as absolute information.

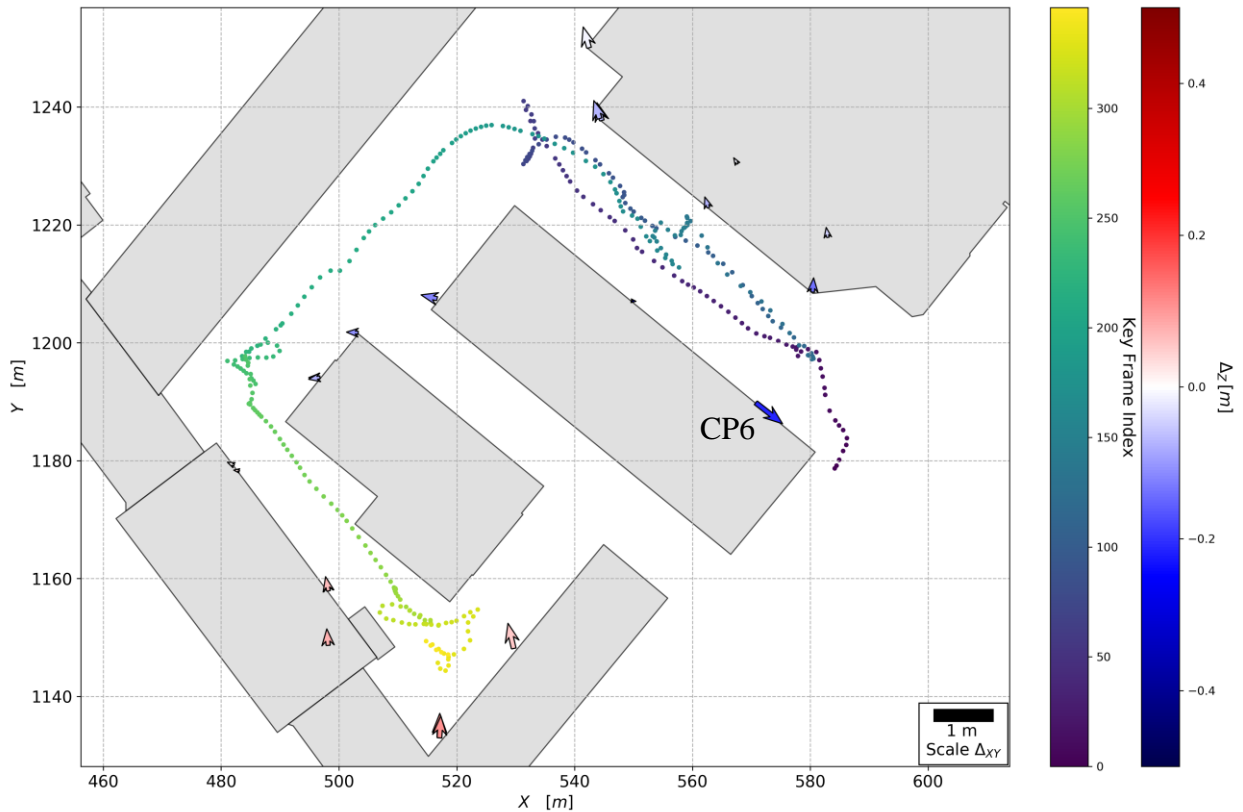


Figure 22: Check point errors and estimated image positions for Scn4a of the long sequence. Δ_{xy} [dm] shown as arrows coloured by Δ_z [dm].

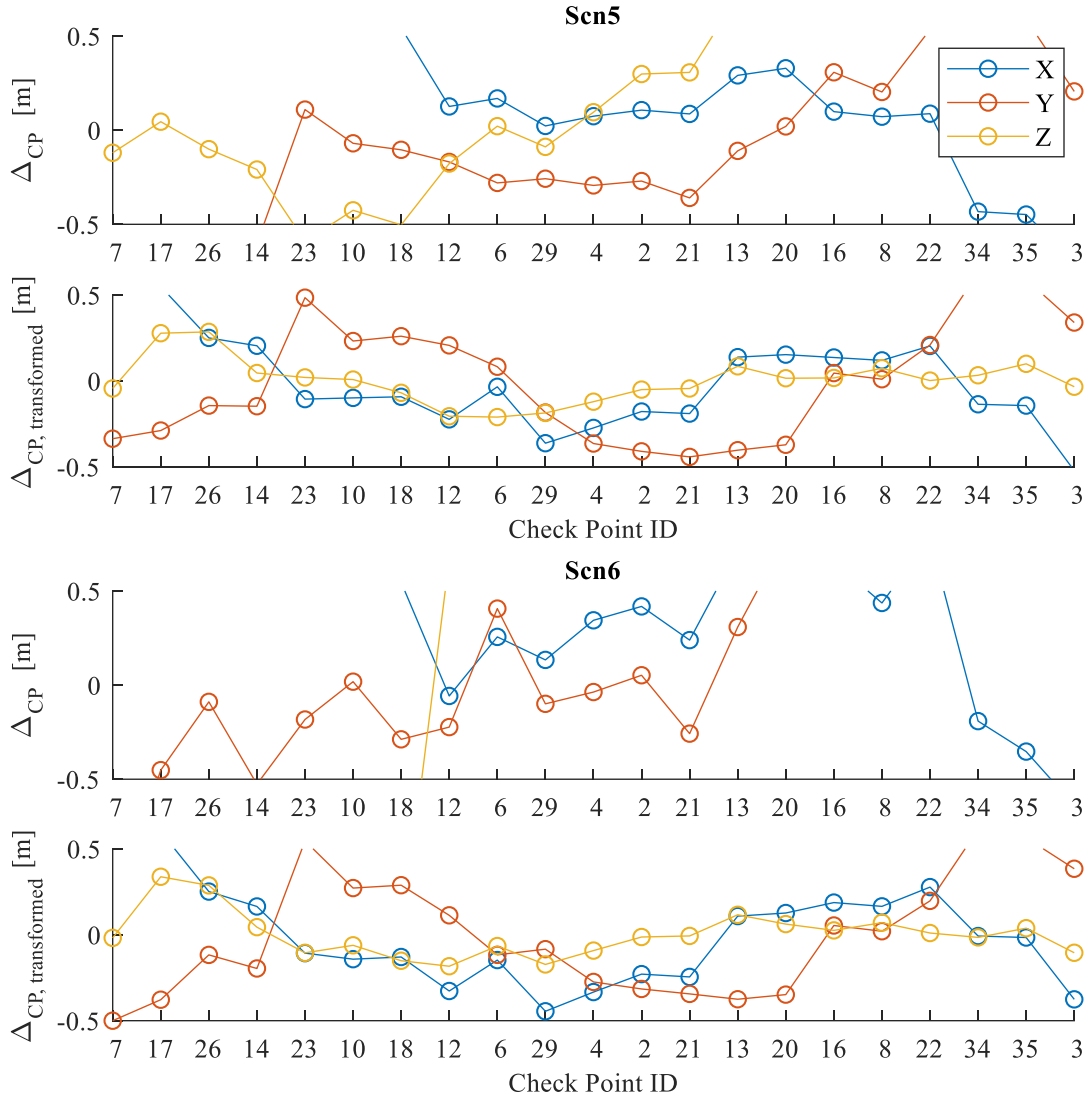


Figure 23: Long sequence without a building model: Check point errors per check point in the order of their appearance in the sequence before and after transformation for Scn5 and Scn6.

The check point errors after global adjustment are larger for Scn6 compared to Scn5. Scn6 only uses GNSS measurements at the beginning of the image sequence. The check point errors after transformation are distributed similarly for Scn5 and Scn6.

After transformation, larger check point errors remain for Scn5 and Scn6 that do not integrate a building model compared to the errors of Scn2 and Scn4, which integrate the generalised building model (Figure 21). Such differences were not observed for the short sequence where scenarios 2, 3 and 5 result in similar check point errors after transformation.

Discussion

Differences in check point errors between scenarios after transformation represent local block deformations. Check point errors are smallest in Scn2 with the corrected generalised building model. Without the correction of systematic effects of model planes in Scn4, they are larger by the amount of block deformations introduced by systematic effects. The fact that even larger errors remain without the usage of a building model (Scn5/Scn6) shows that block deformations are reduced by using a generalised building model for the long sequence. The correction of local block

deformations in the scenarios that integrate a building model also shows that it is helpful to integrate the plane correspondences individually into the optimisation instead of using them just to estimate a datum transformation to correct an image block using a building model. ICP-based approaches that estimate a transformation for the whole image block would not have been able to correct the local block deformations.

If GNSS measurements are used only at the beginning of the sequence (Scn6), the transformation parameters are larger, but the remaining errors are close to those of Scn5, in which GNSS measurements are used along the whole sequence. This shows that the block in these two scenarios is deformed equally but the datum is initialised worse by just using the first 50 GNSS observations. While the GNSS measurements help to improve the datum, they are not accurate enough to notably correct block deformations within the sequence.

6.3 Check Point Errors versus Estimated Standard Deviations

The RMS check point errors as external diagnostics are now compared to the RMS of the estimated standard deviations of the check points and to the check point errors after a Helmert transformation. For clarity, RMS check point errors are denoted as Δ_{CP} and $\Delta_{CP,t}$ and the RMS of estimated standard deviations as $\hat{\sigma}_{CP}$ in this section: All discussed numbers in this section are RMS values. The comparison is shown for the short sequence only, as the findings were found to be valid also for the long and the full sequences. The estimated variance factor $\hat{\sigma}_0$ with all sequences for all scenarios is ~ 0.8 .

Figure 24 extends Figure 14 (page 65) by $\hat{\sigma}_{XYZ_{CP}}$ and the $\Delta_{XYZ_{CP,t}}$ for the short sequence. For Scn5, $\hat{\sigma}_{XYZ_{CP}}$ is almost twice as large as $\Delta_{XYZ_{CP}}$. For the scenarios incorporating a building model, $\hat{\sigma}_{XYZ_{CP}}$ is considerably smaller than $\Delta_{XYZ_{CP}}$: The actual errors are significantly larger than the precision estimated by the adjustment. While $\Delta_{XYZ_{CP}}$ significantly decreases from Scn3 to Scn2 and from Scn2 to Scn1, for these three scenarios $\hat{\sigma}_{XYZ_{CP}}$ remains in the same range of 1 cm. The $\hat{\sigma}_{CP}$ values are in the range of a precision that can be expected for image observations of the captured image sequence with a mean GSD in the sub-cm-range.

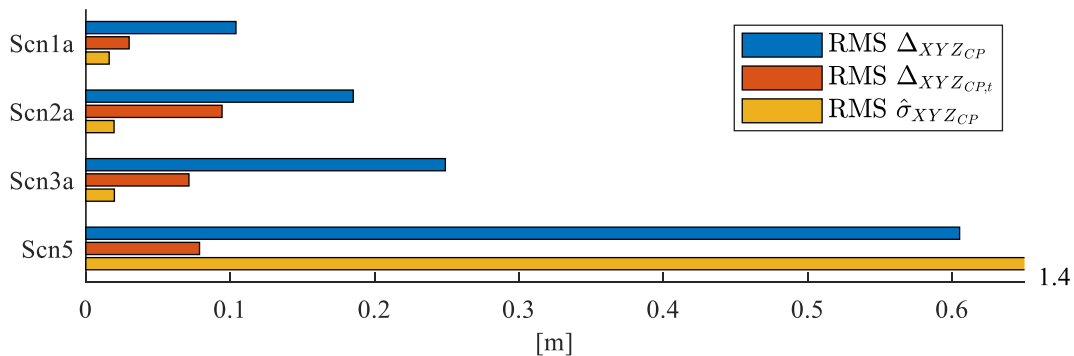


Figure 24: Short sequence: Comparison of the RMS check point errors before and after transformation and the RMS of the estimated standard deviations of all check points per scenario.

The $\Delta_{XYZ_{CP,t}}$ in terms of magnitude between the original check point errors and the estimated standard deviations show that the Helmert transformation reduces the errors. But only in Scn1, $\Delta_{XYZ_{CP,t}}$ and $\hat{\sigma}_{XYZ_{CP}}$ are close to each other.

Figure 25 shows $\hat{\sigma}_{CP}$, $\Delta_{CP,t}$ and Δ_{CP} for all three coordinate axes. As mentioned before, check point errors after transformation are quite similar for all scenarios except Scn1.

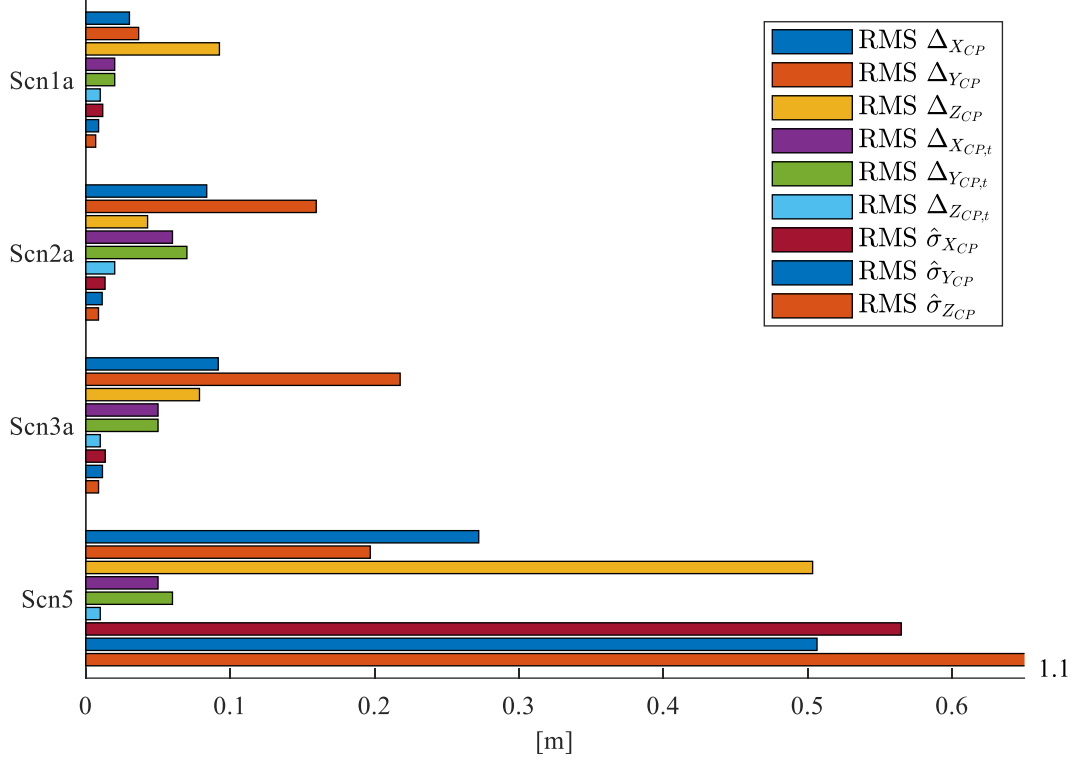


Figure 25: Short sequence: Comparison of the RMS check point errors before and after transformation and the RMS of the estimated standard deviations of all check points per coordinate axis per scenario.

lists the $\hat{\sigma}_{CP}$ and Δ_{CP} values and in the right columns additionally shows the relations of $\Delta_{CP}/\hat{\sigma}_{CP}$. Scn1, as shown in section 6.1, has the largest check point errors in Z-direction. While the errors in X- and Y-direction are 3 and 4 times larger than $\hat{\sigma}_{CP}$, the difference is even larger for the Z-direction. In Scn3, Δ_{CP} is 7 to 19 times larger than $\hat{\sigma}_{CP}$, while in Scn2 this ratio is between 5 and 14.

Short Seq.	$\Delta_{X_{CP}}$ [m]	$\Delta_{Y_{CP}}$ [m]	$\Delta_{Z_{CP}}$ [m]	$\hat{\sigma}_{X_{CP}}$ [m]	$\hat{\sigma}_{Y_{CP}}$ [m]	$\hat{\sigma}_{Z_{CP}}$ [m]	$\frac{\Delta_{X_{CP}}}{\hat{\sigma}_{X_{CP}}}$	$\frac{\Delta_{Y_{CP}}}{\hat{\sigma}_{Y_{CP}}}$	$\frac{\Delta_{Z_{CP}}}{\hat{\sigma}_{Z_{CP}}}$
Scn1a	0.03	0.04	0.09	0.01	0.01	0.01	2.6	4.1	13.6
Scn2a	0.08	0.16	0.04	0.01	0.01	0.01	6.3	14.1	4.9
Scn3a	0.09	0.22	0.08	0.01	0.01	0.01	6.8	18.9	8.9
Scn5	0.27	0.20	0.50	0.56	0.51	1.12	0.5	0.4	0.5

Table 12: $\hat{\sigma}_{CP}$ and Δ_{CP} and their ratio for the short sequence.

Table 13 shows the decreased difference between the estimated standard deviations and the check point errors that remain after the Helmert transformation. The Z-shift of the Helmert transformation strongly reduces the ratio in Z-direction for Scn1. For Scn2 and Scn3, a relatively

large $\Delta_{Y_{CP}}$ is reduced by the transformation. An explanation for the large $\Delta_{Y_{CP}}$ is an unfavourable distribution of planes that support the block as absolute information in Y-direction. Low redundancy in such planes combined with systematic effects of the planes seems to lead to larger errors.

Short Seq.	$\Delta_{X_{CP,t}}$ [m]	$\Delta_{Y_{CP,t}}$ [m]	$\Delta_{Z_{CP,t}}$ [m]	$\hat{\sigma}_{X_{CP}}$ [m]	$\hat{\sigma}_{Y_{CP}}$ [m]	$\hat{\sigma}_{Z_{CP}}$ [m]	$\frac{\Delta_{X_{CP,t}}}{\hat{\sigma}_{X_{CP}}}$	$\frac{\Delta_{Y_{CP,t}}}{\hat{\sigma}_{Y_{CP}}}$	$\frac{\Delta_{Z_{CP,t}}}{\hat{\sigma}_{Z_{CP}}}$
Scn1a	0.02	0.02	0.01	0.01	0.01	0.01	1.7	2.3	1.5
Scn2a	0.06	0.07	0.02	0.01	0.01	0.01	4.5	6.2	2.3
Scn3a	0.05	0.05	0.01	0.01	0.01	0.01	3.7	4.3	1.1
Scn5	0.05	0.06	0.01	0.56	0.51	1.12	0.1	0.1	0.0

Table 13: $\hat{\sigma}_{CP}$ and $\Delta_{CP,t}$ and their ratio for the short sequence.

Discussion

For all scenarios with a building model, the estimated $\hat{\sigma}_{CP}$ values are significantly smaller than the Δ_{CP} values. This discrepancy indicates remaining systematic effects.

The reduced check point errors after the Helmert transformation suggest that large portions of the discrepancy can be related to the datum of the image block: The $\Delta_{CP,t}$ values are closer to $\hat{\sigma}_{CP}$ than Δ_{CP} . Given that the datum of the building models and the check points is the same, other systematic effects have to be analysed. As mentioned above for the large $\Delta_{Y_{CP}}$ values, a low redundancy of model planes combined with systematic errors of the model planes is a potential source of datum errors.

As soon as a building model is used, the influence of the GNSS measurements on the datum of the image block is very small. Therefore, systematic errors of model planes as absolute information are considered by comparing Scn2 and Scn3: The similar $\hat{\sigma}_{CP}$ values for Scn2 and Scn3 are the expected result, as systematic errors of the generalised building model do not influence $\hat{\sigma}_{CP}$. The systematic errors of model planes that lead to larger Δ_{CP} values with the original generalised model are not included in $\hat{\sigma}_{CP}$, which consequently do not change.

The remaining discrepancy between Δ_{CP} and $\hat{\sigma}_{CP}$ after Scn2 shows that systematic errors still exist in the model after the correction and even after the Helmert transformation. The correction of the generalised building model only partly eliminates systematic errors of model planes. The generalisation effects themselves are partly systematic within the planes. As mentioned when setting up the stochastic model in section 3.2.3, Gaussian-distributed errors are assumed by setting an a priori standard deviation for the fictitious observations. This does not reflect the often systematic character of deviations due to generalisation.

As shown in section 6.1, even for the short sequence, block deformations occur that are reduced using the detailed building model (Scn1, see Figure 18). Scn1 results in a further reduced discrepancy between estimated standard deviations and check point errors in X- and Y-direction. Note that while the detailed model is retrieved from the plans of the architect by correction of the individual planes with a reference point cloud, the topology is not changed – no details are added

and structure not present in the plans is still missing in the detailed model. This leads to remaining systematic effects causing the remaining discrepancy between Δ_{CP} and $\hat{\sigma}_{CP}$ even for the detailed building model.

As described, model planes that contribute to the Z-direction are neither covered by tie points sufficiently nor described by the model in the same detail and accuracy as the tie points. This explains the larger discrepancy in Z-direction with Scn1. With a ratio of 13.6, this discrepancy is even larger than with a generalised building model.

Figure 25 shows that in the scenarios that use a generalised building model the check point errors in Z-direction are even smaller than those of Scn1. Whereas the detail and amount of generalisation effects at many roof planes are comparable for the detailed and the generalised model, the used parameters are not: Assignment thresholds and a priori standard deviations are smaller in Scn1 than in scenarios 2 and 3, which leads to fewer assignments at roof planes for Scn1. The quality of these assignments likely is overestimated by the $\hat{\sigma}_{ZCP}$ in Scn1, which in combination with systematic effects at the roof planes leads to the large ratio of $\Delta_{ZCP}/\hat{\sigma}_{ZCP}$.

The a priori standard deviation of the GNSS measurements was set to several metres, which, in Scn5, turns out to be larger than the actual observed check point errors and the $\hat{\sigma}_{CP}$. The larger $\hat{\sigma}_{CP}$ compared to the Δ_{CP} shows that the accuracy of the GNSS measurements is underestimated for the captured sequence.

The results of this section show that as soon as a building model is used, even with a detailed building model, the estimated standard deviations are affected by remaining systematic effects. The more generalisation effects are present in the building model, the larger these effects are.

6.4 Sliding Window versus Global Adjustment

The previous sections analysed the results of the global adjustment. This section compares the results of the sliding window processing to the global adjustment concerning check point errors using the long sequence. The difference between sliding window and global processing has to be interpreted as the potential loss of accuracy that occurs due to only optimising incrementally, reusing some already adjusted images instead of processing all available data at once. Note that the global adjustment uses the estimation results of the windows as initial values and is not run independently on the whole image sequence at once. The reason for not directly running the global approach on the whole sequence is that block deformations along the full sequence prevent a successful assignment and let the global hybrid adjustment fail. The assignment with large block deformations and drift effects fails as the assumption of good initial values for tie point coordinates close to model planes becomes violated. Therefore, the sequential processing in windows of the hybrid adjustment is carried out first to be able to start the global hybrid adjustment from good initial values.

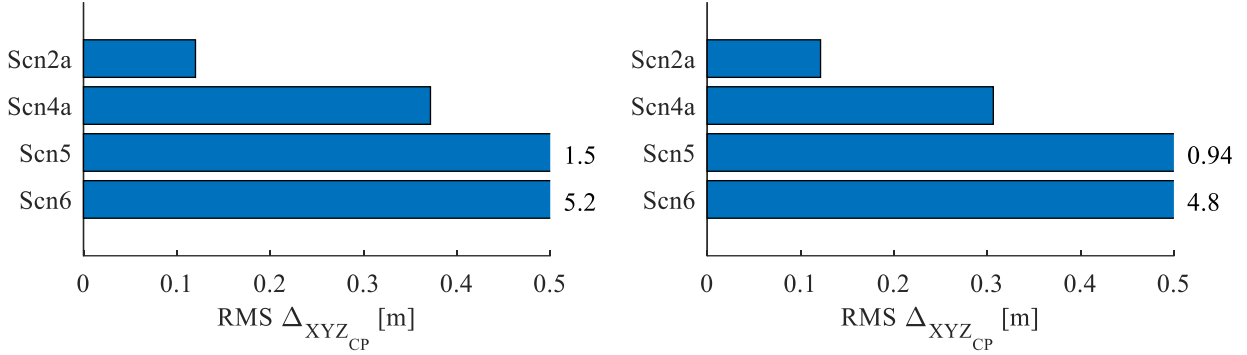


Figure 26: Long sequence: Comparison of the RMS of the RMS check point errors of all windows (left) and for the global adjustment (right).

In Figure 26, the check point errors for the sliding window processing, i.e. RMS over all RMS check point errors of all windows, are presented along with the RMS check point errors of the global hybrid adjustment already shown in Figure 19. In all scenarios except Scn2a the sliding window approach shows higher errors compared to the global adjustment. Errors are approximately equal for Scn2a. The error over all windows after sliding window processing is larger by 5 cm in Scn4a and around half a meter in Scn5 and Scn6.

The relatively small loss of accuracy caused by sliding window processing shows that the windows themselves result in check point errors relatively close to those of the global optimisation. This depends on the parameter settings, especially on the window size. Windows ideally should be large enough to cover the same length of trajectory and number of model planes that influence the results locally within the global adjustment. The better the building model stabilises the windows in themselves, the lower is the potential improvement of a global solution.

For Scn2a and Scn4a, Figure 27 shows the RMS errors at check points per window (8 windows) and for the global adjustment (last bar in the figure). As already seen in the overall RMS errors, the RMS values $\Delta_{XYZ_{CP}}$ are considerably smaller in Scn2a than in Scn4a: They vary between 8 and 22 cm in Scn2a and between 20 and 80 cm in Scn4a.

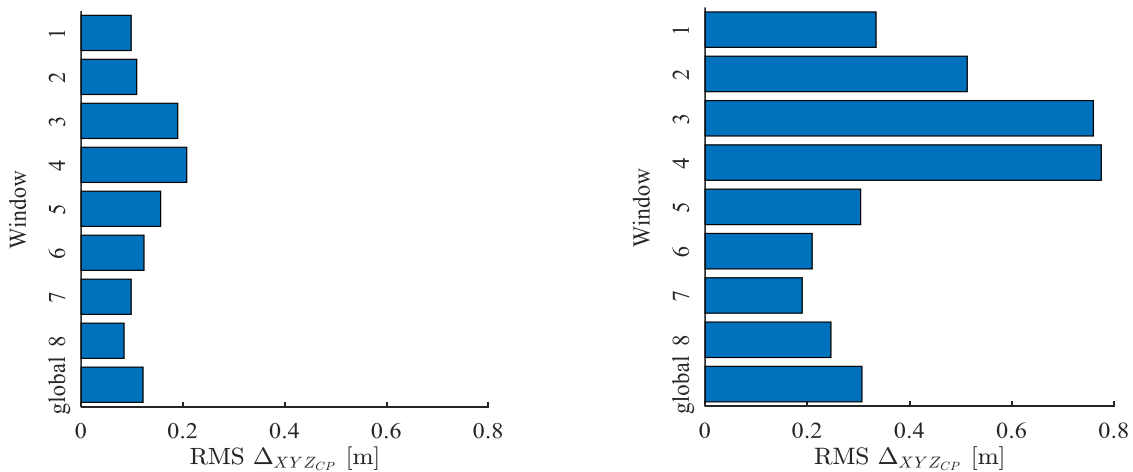


Figure 27: Long sequence, Scenarios 2a (left) and 4a (right): RMS errors over all check points for each window and for the global adjustment.

Windows 3 and 4 show the largest check point errors in both scenarios that include the generalised building model. These windows include the façade of check points 6 and 12 captured in a loop along the façade plane. The large check point errors observed before at check point 6 occur as well in the sliding window processing.

In summary, the global adjustment leads to slightly improved check point errors compared to the sliding windows. The errors vary from window to window and are largest at windows with poor model information. Overall, the results show that window size and overlap are sufficient to cover most parts of the sequence in a way that locally relevant observations and unknowns take part in the local optimisations.

6.5 Assignment Strategies

This section compares the three assignment strategies introduced in chapter 4 to each other based on RMS check point errors. The comparison uses the long sequence. Results of the short sequence are summarised in the following.

For the short sequence, the assignment strategies lead to similar results with the generalised building model. Assignment methods (a) and (c) perform similarly with the detailed building model (Scn1a and Scn1c) while method (b) fails to segment many small planes from the tie point cloud and results in false assignments. Due to the low number of assigned planes in total and a relatively high number of wrong assignments of model planes, the RMS Δ_{XYZCP} of Scn1b with the short sequence are even larger than those of Scn5 that does not integrate any building model. This leads to the conclusion that method (b) is not feasible for a detailed model with the given setup.

For the long sequence, Figure 28 shows the RMS Δ_{CP} per axis for Scn2 and Scn4 after sliding window processing (left) and after the global adjustment (right). The check point errors for

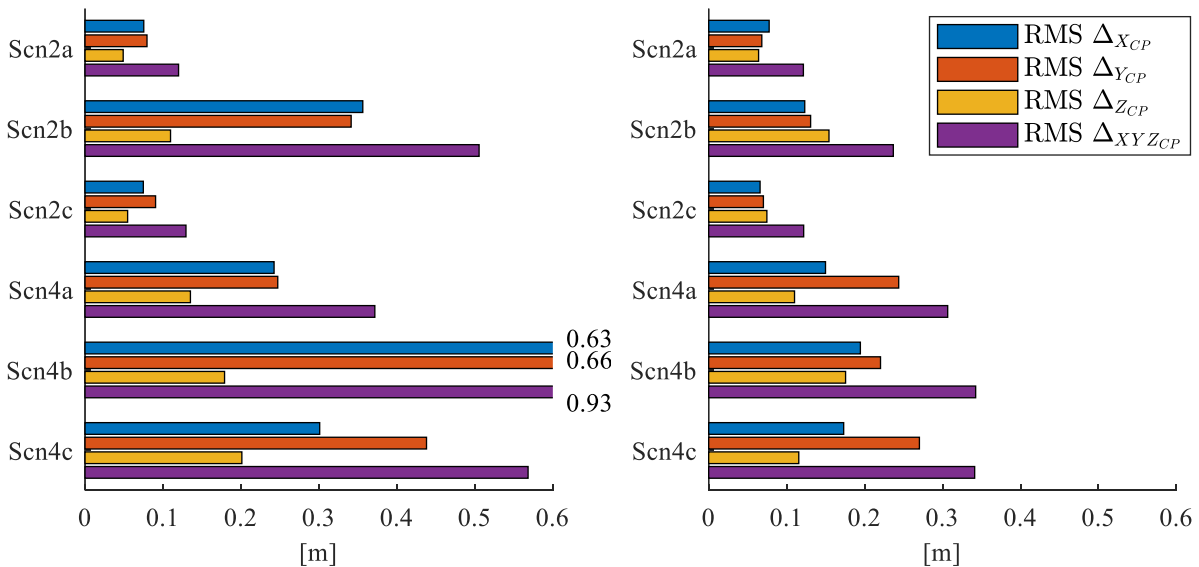


Figure 28: Long sequence: RMS errors of all check points for the windows (left) and the global adjustment (right).

methods (a) and (c) are close by each other with a tendency of lower errors with method (a). For the sliding window processing, assignment method (b) leads to the highest check point errors in Scn2 as well as in Scn4.

As seen with assignment method (a) in the sections before, the global processing overall leads to smaller check point errors compared to the sliding windows. The improvement is largest with assignment method (b): With global adjustment, Scn4b results in errors closer to those of scenarios 4a and 4c, and also the differences of Scn2b to scenarios 2a and 2c are smaller.

To further analyse the differences of the methods, the total number of assigned tie points and the total number of model planes with assignments of tie points are compared. For Scn4, Table 14 shows these statistics and the number of key frames, tie points and the number of observations eliminated by robust estimation.

	Number of Key Frames	Number of Tie Points	Number of Image Observations	Number of Planes with Assignments	Number of Assignments, Fictitious Obs.
Scn4a	346	117951	965412 (6454)	51	96548 (0)
Scn4b	349	120173	980104 (10063)	15	60793 (3)
Scn4c	332	112662	912528 (5739)	54	90504 (0)

Table 14: Statistics for the global adjustment of the long sequence with Scn4. Number of observations eliminated by robust estimation are given in parentheses.

As described before, RANSAC procedures are used in relative image orientation and in the elimination of erroneous image matches when generating initial values for the adjustment. The usage of RANSAC leads to varying numbers of key frames. Scn4c consists of fewer key frames than Scn4a and Scn4b, leading to lower numbers of tie points and assignments of tie points. However, for Scn4a and Scn4c, 80% of the tie points are assigned to model planes, while Scn4b results in significantly fewer planes with assignments and only 50% assigned tie points.

Comparison at an example building

Figure 29 shows the assignment results with the three assignment methods for the building denoted as c) in Figure 10, section 5.2.2. The assignments of methods (a) and (c) turn out to be nearly equal, whereas method (b) only assigned tie points to one façade plane in this area. Tie points at the garage gates are not assigned to the model planes by all three methods. Tie points of the balustrade of the balcony also are not assigned as the balustrade is not represented in the generalised model. The façade between balcony and building roof is an example of a plane where the model plane and the captured tie points deviate that much from each other that no assignment takes place. The good fit of the assigned tie point cloud to the other façades suggests that this unassigned façade is inaccurately modelled. In the first iterations of the hybrid bundle adjustment, this plane gets assignments. But with decreasing distance thresholds, these assignments are not found again. As the a priori standard deviation of the direct observations of vertex coordinates is set to 1 cm for the experiments, the plane does not change notably. The correction of model planes by the hybrid bundle adjustment is investigated in section 6.7.

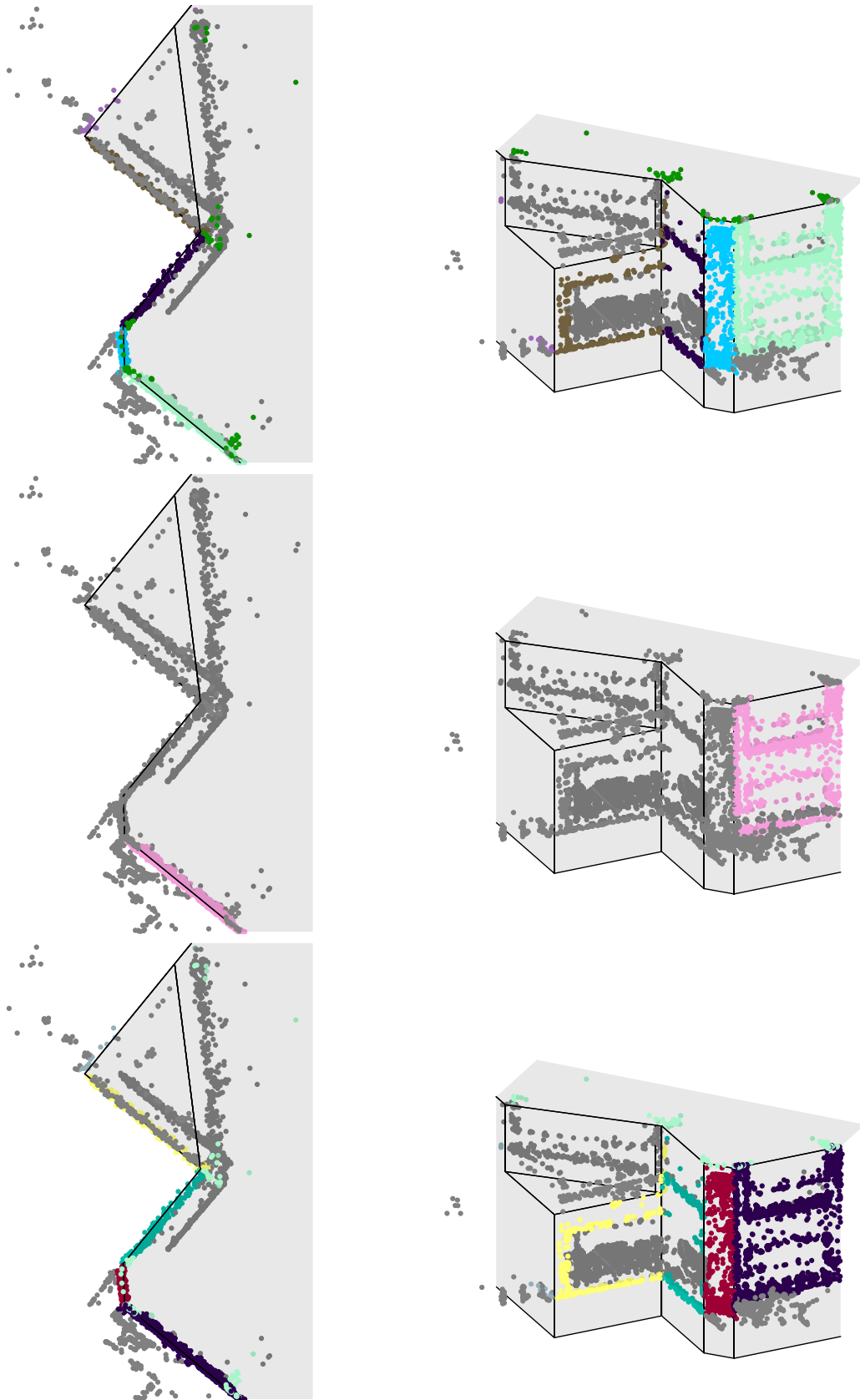


Figure 29: Comparison of the assignments of Scn4a (top), Scn4b (middle) and Scn4c (bottom) at one façade in top and side views.

Comparison at one plane

Figure 30 shows the results of the first five iterations of the global hybrid bundle adjustment for the vertical façade plane shown with yellow assigned tie points in Figure 29 bottom. As was shown in Figure 10, section 5.2.2, where this building is shown as example c), there is a garage door not represented by the generalised building model. With the large distance threshold of 2 m

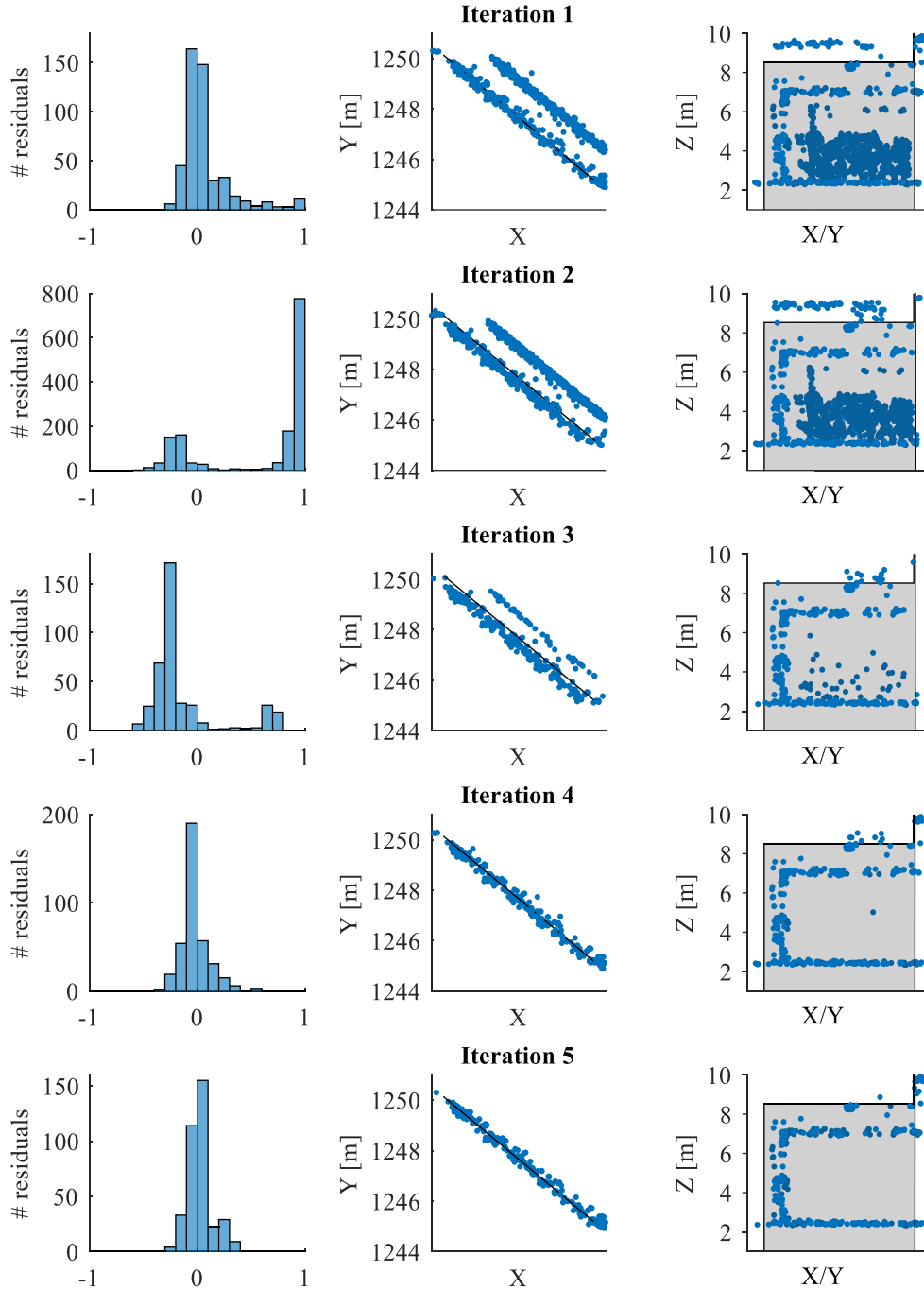


Figure 30: Results of the direct assignment (a) for the first five iterations of the hybrid adjustment for the left model plane with a garage door at building c) in Figure 10, section 5.2.2 and in Figure 29: Histogram of the fictitious distance observations of assigned tie points to the model planes (left), top (middle) and orthogonal view on the model plane (shown in grey) with only the assigned tie points in blue.

in the first iteration of the hybrid adjustment, all points in the surroundings of the façade are assigned to the model plane (Figure 30). The assignment includes points of the garage door which form a plane parallel to the model plane. After one iteration, these points are assigned again. They appear in the shown histogram at the 1 m residuals as the adjustment lead to a shift of the tie point cloud, bringing the points of the garage door closer to the model plane. Other planes of the image block prevent the adjustment to “pull” the points on the garage door even closer to the model plane.

As the distance threshold is lowered in each iteration due to an overall better fit of tie points to model planes, this lower threshold leads to the assignment shown for iteration three: Only some closer points of the garage door are still assigned to the model plane. The majority exceeds the lowered threshold and is not assigned again. The adjustment consequently results in image poses shifting the tie points back, so that the façade points come closer to the model plane and the tie points of the garage gate move away from the model plane again. In combination with a further reduced distance threshold, the assignment in iteration 4 converges to only the points close to the façade and correctly uses them for the hybrid adjustment to relate the image block to the model façade in following iterations. Due to the lowered distance threshold, no tie points of the garage door are assigned to the model plane again.

Figure 31 shows the results of the assignment with the indirect method (b) without using ROIs for the same model plane as in Figure 30. With method (b) the plane is only detected three times. After the three iterations, the distance of the COG of the extracted plane exceeds the lowered distance threshold and results in the situation seen in Figure 29: This plane has no accepted

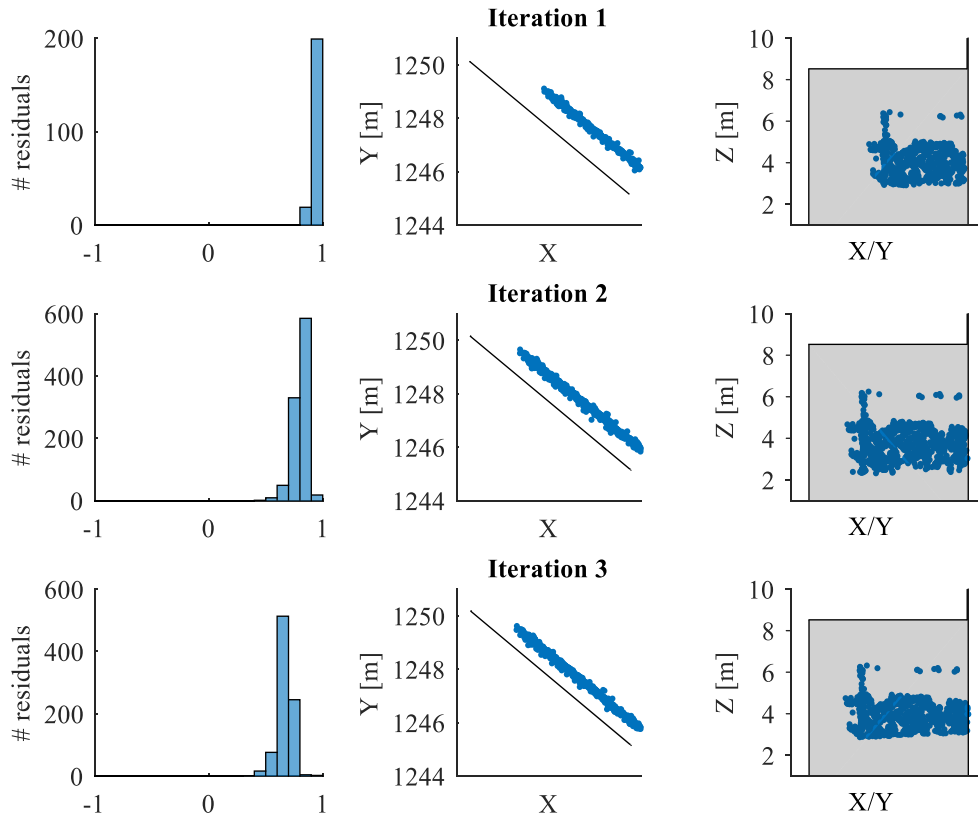


Figure 31: Assignment results for the indirect method (b).

assignments. The reason for this is that only the garage door is found as a plane by MSAC in the tie point cloud.

The usage of an ROI to search for planes leads to the results shown in Figure 32 for assignment method (c). In the beginning, tie points of the gate form the plane with most inliers in the ROI around the model plane. With lowered ROI distance thresholds from iteration 4 on, the tie points at the gate fall outside the ROI. Therefore, the façade points are detected as the extracted plane close enough to the model plane leading to the correct adjustment of assigned façade points to the model plane rejecting assignments of points of the garage gate.

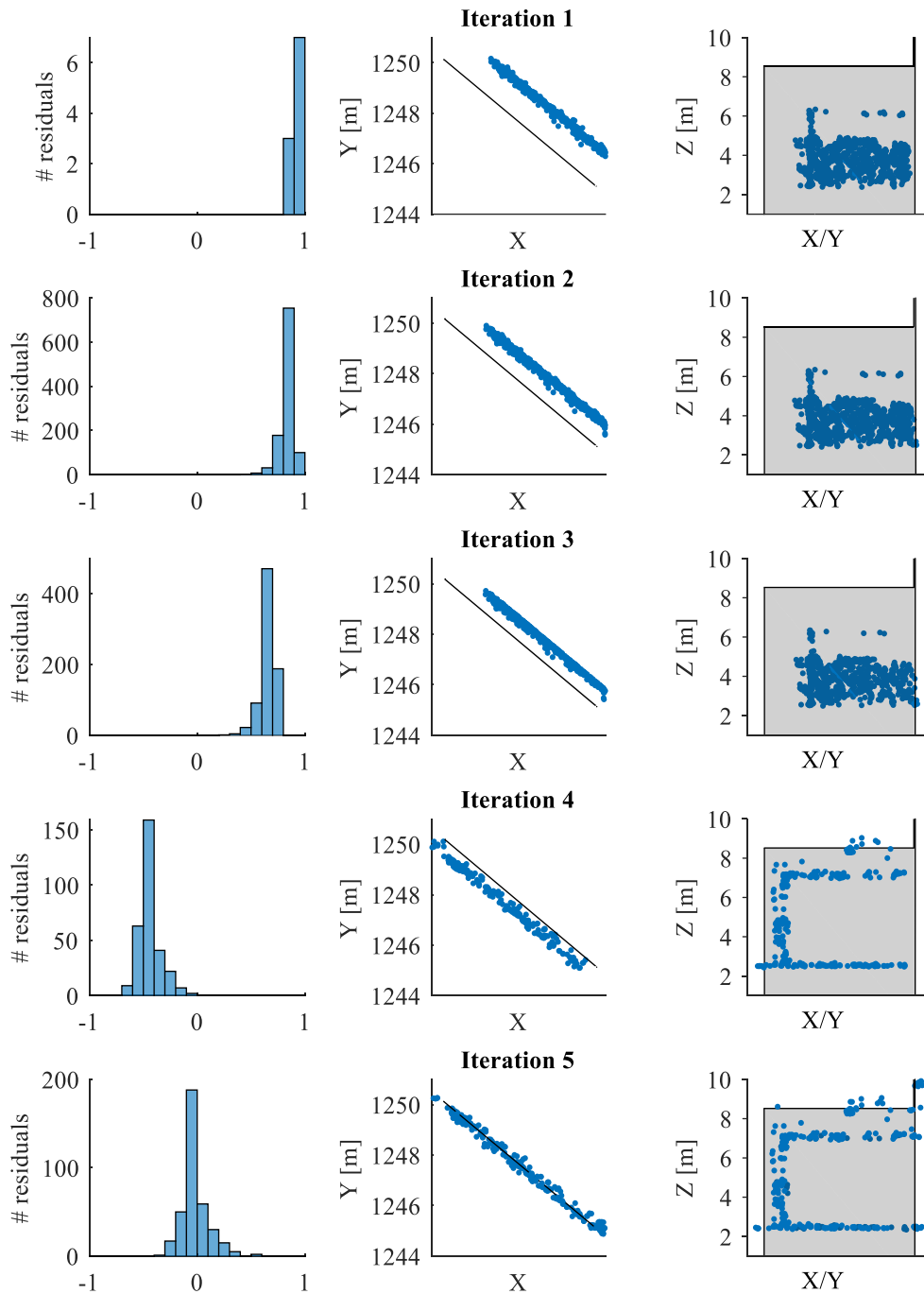


Figure 32: Assignment results for the indirect method (c) with ROIs.

Overall comparison

In general, assignment method (b) finds assignments to fewer planes than the other two methods. However, the planes found by method (b) typically are the dominating large planes. This finding also explains, why 70% fewer planes are found by method (b) compared to the other two methods, but the total number of point-plane assignments is lower by 30% only. Figure 33 shows the assignments of the methods (a) and (b) for the whole scene.

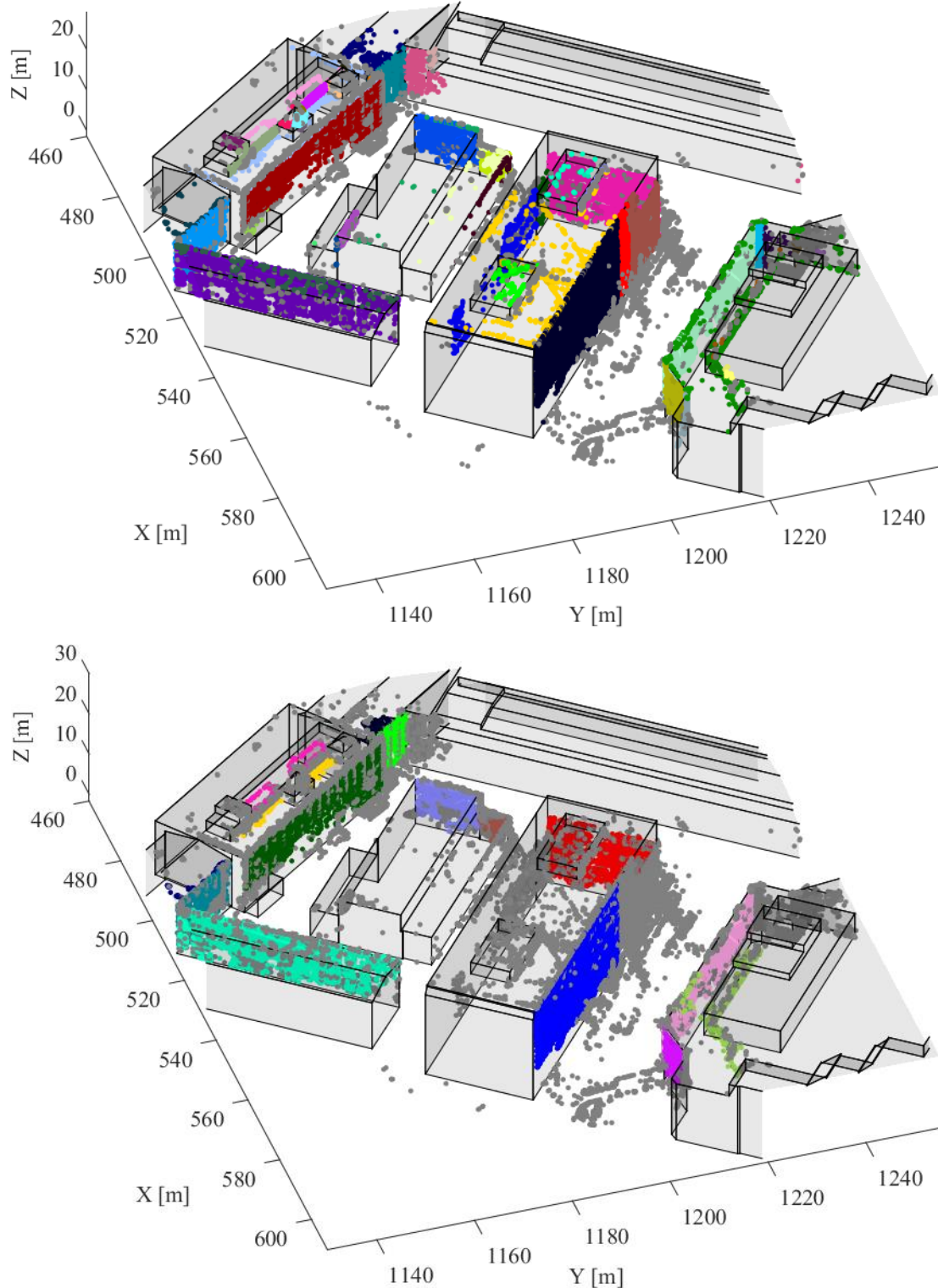


Figure 33: Comparison of the assignments of Scn4a (top) and Scn4b (bottom).

Discussion

Using method (b), the discrepancy between errors after sliding window and global optimisation is larger than with the other methods. This shows that a lower number of found planes degrades the orientation results of windows but has a smaller negative effect on the global adjustment. This finding suggests that method (b) might require a larger window size compared to methods (a) and (c).

Method (b) performs worse than the other two methods, but at the same time is the only method that independently searches for planes in the tie point cloud. Method (b), therefore, shows the largest potential to be expanded for scenarios that violate the assumption of good initial poses. Such scenarios would require to match the whole set of extracted and model planes instead of only assigning neighbouring planes.

Method (c) requires more processing steps than method (a) but does not outperform the simpler direct assignment in the experiments. Therefore, the results suggest to prefer method (a) over (c). Further investigations would be needed to analyse differences of those two methods in more detail.

6.6 The Full Sequence

The full sequence consists of the long sequence and additional images at the beginning and the end of the sequence. These images add more complex configurations that are investigated in the following. In the beginning, the flight starts between buildings with the worst GNSS reception and captures images along two vertical façade planes only. GNSS observations are off by almost 5 m for the first images. At the end of the sequence, images are recorded with small base lengths and large rotations in the yaw direction. Additionally, these images capture building parts with strongly varying distance from the sensor, leading to an unfavourable distribution of matched image features within images due to occlusions.

Figure 34 depicts the trajectory reconstructed for the full sequence in Scn4a with the XY-check point errors as arrows coloured by the Z-error. The additional key frames in the beginning and at the end of the sequence at check points 15, 11, 19 and 25 show the largest deviations. Check points 11 and 15 are clearly shifted along the façade plane. This shift is attributed to missing point-plane relations to other planes with different orientation as there is only this vertical façade plane at the beginning of the sequence. At check points 19 and 25, the errors are attributed to the unfavourable image capturing conditions: At check point 3, the copter rotated around the yaw axis while there was not much translational movement. This movement generates key frames with an insufficient base length that are not skipped because otherwise image overlap would decrease and no image matches would be found.

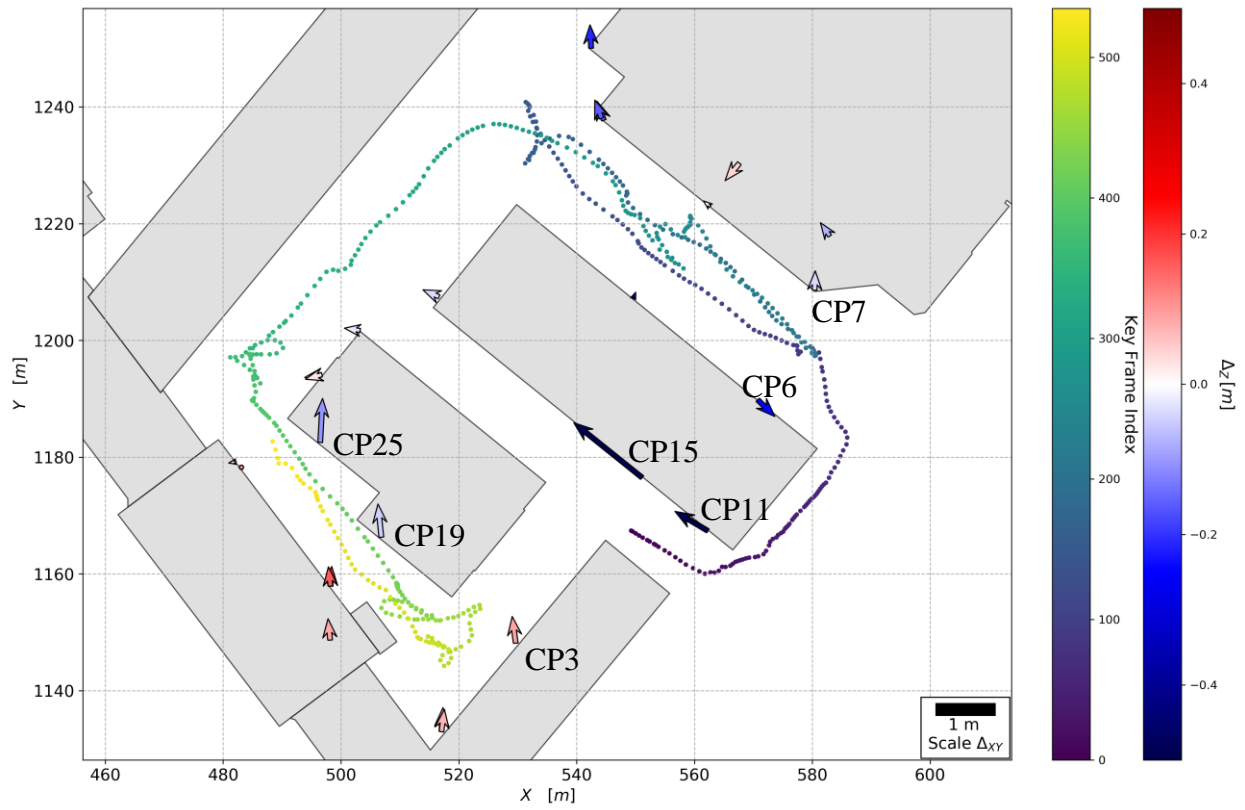


Figure 34: Check point errors and estimated image positions for Scn4a of the full sequence. Δ_{xy} [m] shown as arrows coloured by Δ_z [m].

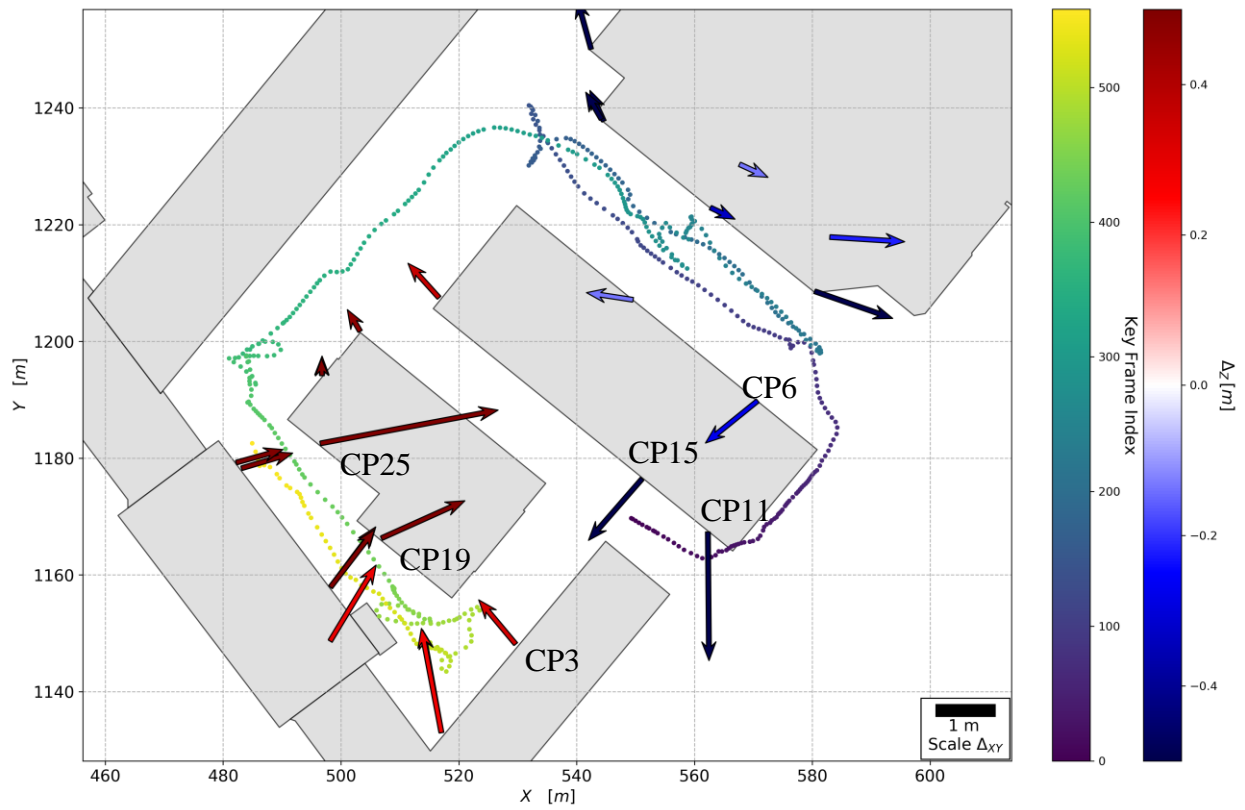


Figure 35: Check point errors and estimated image positions for Scn5 of the full sequence. Δ_{xy} [m] shown as arrows coloured by Δ_z [m].

In Scn5, without the building model, the largest check point errors result at the beginning and the end of the sequence (Figure 35). The check point errors at the end of the sequence at check points 19 and 25 are largely reduced after Scn4a, but the block deformations were not corrected completely by the integration of the generalised building model.

Systematic effects (scenarios 2a and 4a)

Figure 36 shows the check point errors for Scn2a and Scn4a before and after Helmert transformation. As already seen in the check point error plots, the largest errors occur at the beginning and the end of the full sequence, added in comparison to the long sequence. The block deformation around check points 12 and 6 already observed with the long sequence is also visible and is lower with the corrected generalised building model of Scn2a. The Helmert transformation does not lead to largely reduced check point errors. This shows that the errors are primarily due to local block deformations and not due to overall datum defects. The large errors at check points at the end of the sequence only affect the last window of sliding window processing, while they are potentially distributed to larger parts of the full sequence in the global adjustment.

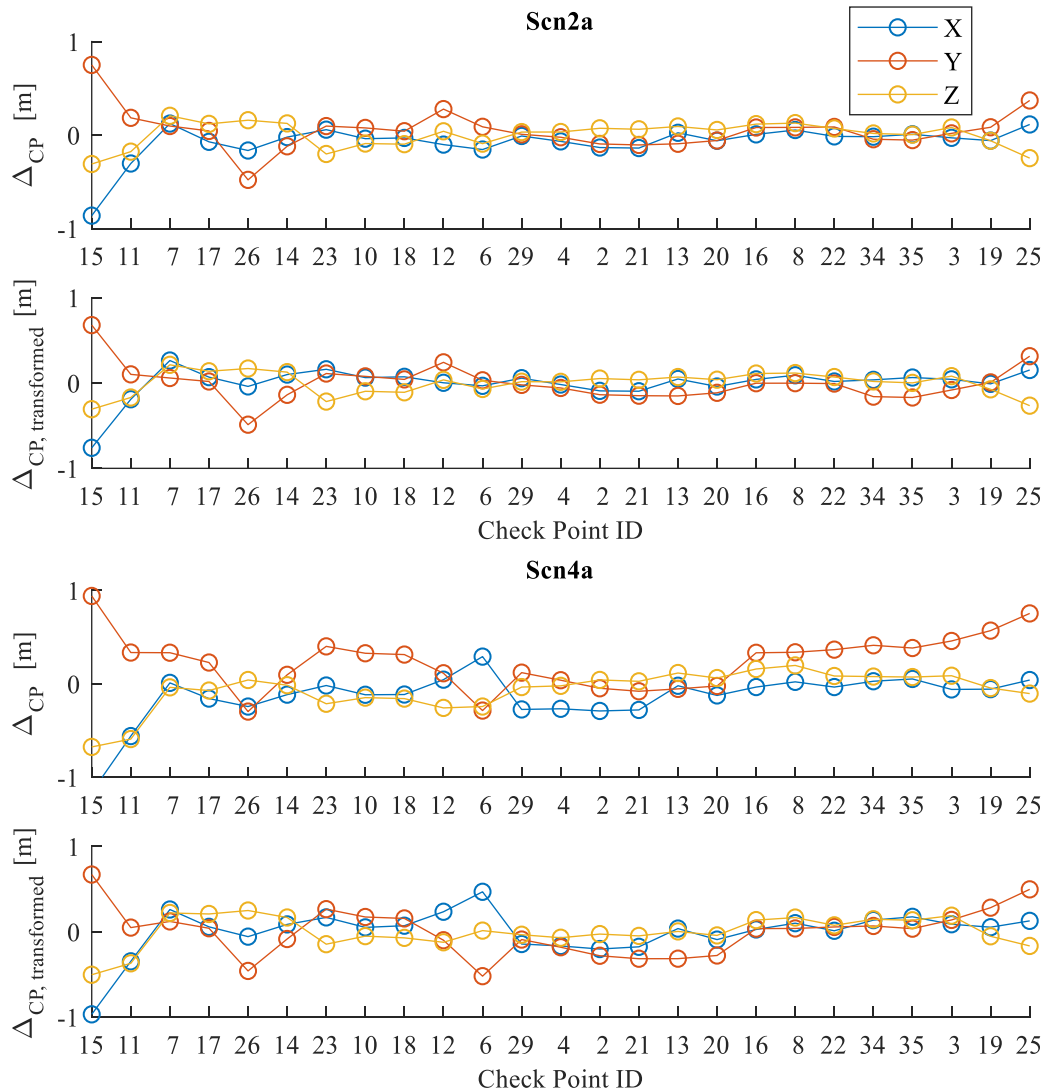


Figure 36: Full sequence: Check point errors per check point in the order of their appearance in the sequence before and after transformation for Scn4a and Scn2a.

Systematic effects without the integration of a building model (scenarios 5 and 6)

Figure 37 shows the check point errors for Scn5 and Scn6. Note the scale factor 10 of the y-axis to show the check point errors without transformation for Scn6. The errors at CP6 in Scn6 are relatively small because the size of the errors correlates with their distance from the start of the sequence where GNSS observations for the projection centre positions were used. The image block is relatively stable compared to the large errors that occur due to low GNSS accuracy. At CP6 the trajectory comes back closer to the first images.

As was observed already with the long sequence, the errors after transformation for scenarios 5 and 6 show larger remaining block deformations than the errors of Scn2 and Scn4 in Figure 36. This underlines the positive effect of integrating a generalised building model into the estimation of tie point coordinates and image poses.

In addition, the remaining check point errors for Scn5 and Scn6 show large block deformations in the beginning and at the end of the image sequence, which means, that the block deformations

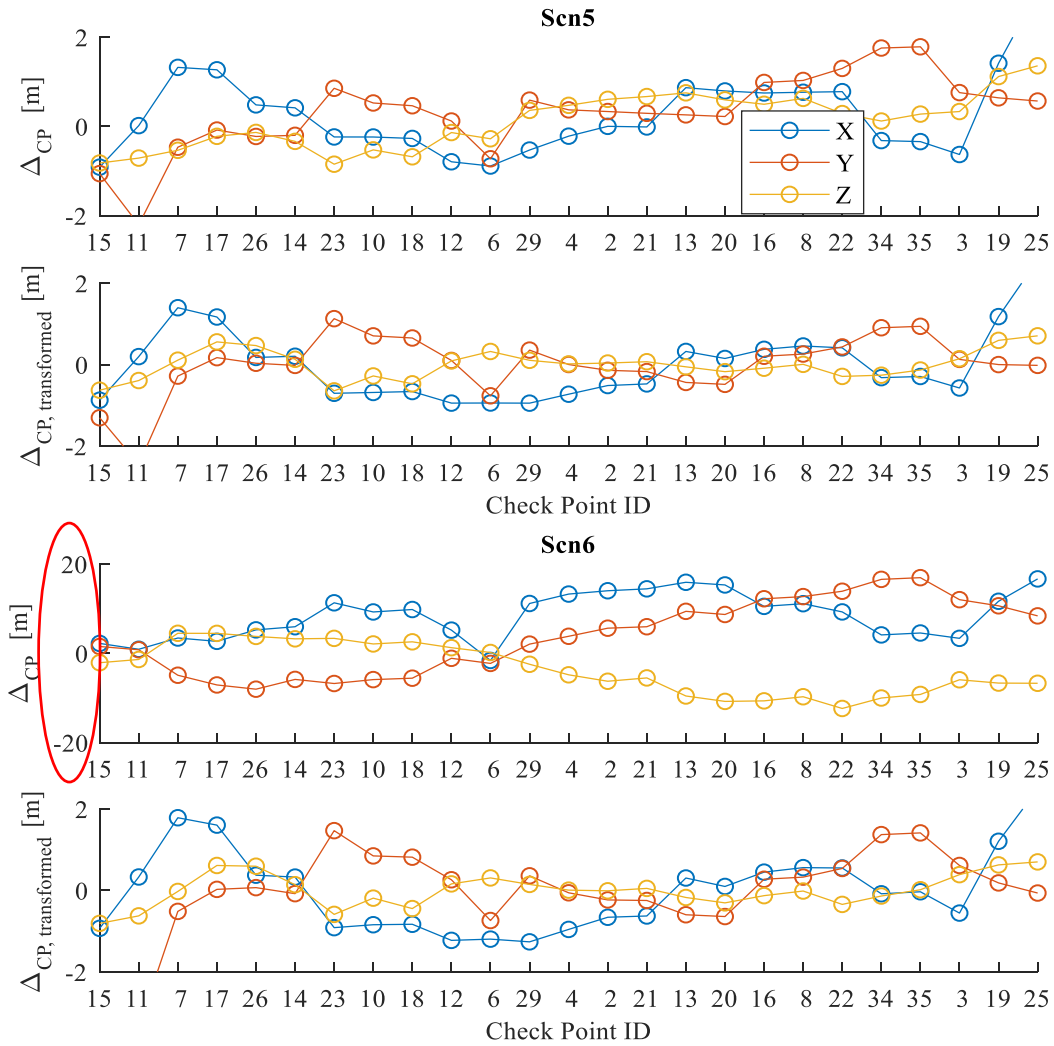


Figure 37: Full sequence: Check point errors per check point in the order of their appearance in the sequence before and after transformation for Scn5 and Scn6. Note the different scaling of the Y-axis compared to Figure 36 and for Scn6 without the transformation.

Long Seq.	X_t [m]	Y_t [m]	Z_t [m]	eX_t [°]	eY_t [°]	eZ_t [°]	s_t	RMS $\Delta_{X_{CP}}$ [m]	RMS $\Delta_{Y_{CP}}$ [m]	RMS $\Delta_{Z_{CP}}$ [m]	RMS $\Delta_{X_{CP,t}}$ [m]	RMS $\Delta_{Y_{CP,t}}$ [m]	RMS $\Delta_{Z_{CP,t}}$ [m]
Scn2a	-0.08	0.05	0.01	-0.01	-0.02	-0.01	1.00	0.20	0.22	0.13	0.18	0.21	0.13
Scn4a	-0.14	0.24	-0.07	-0.05	-0.24	0.00	1.00	0.30	0.37	0.22	0.26	0.27	0.18
Scn5	0.26	0.32	0.12	0.13	-0.83	0.37	0.99	0.93	0.90	0.61	0.86	0.73	0.35
Scn6	8.40	3.76	-3.54	-4.43	5.06	5.01	0.84	9.82	8.81	6.55	0.99	1.02	0.39

Table 15: Parameters ($X_t, Y_t, Z_t, eX_t, eY_t, eZ_t, s_t$) of the Helmert transformations between known and estimated check point coordinates for some of the scenarios of the full sequence and RMS of coordinate differences of the known and the estimated check point coordinates ($\Delta_{X_{CP}}, \Delta_{Y_{CP}}, \Delta_{Z_{CP}}$) as well as of the known and the transformed estimated check point coordinates ($\Delta_{X_{CP,t}}, \Delta_{Y_{CP,t}}, \Delta_{Z_{CP,t}}$).

observed in scenarios 2 and 4 at least partly are already present before integrating any building model into the adjustment.

Table 15 lists the parameters of the Helmert transformation and the RMS check point errors per scenario. In Scn6, without any building model and with GNSS observations for the projection centre positions of only the first 50 images of the sequence, check point errors are largely reduced after a Helmert transformation. The scale parameter shows that the scale of this scenario is off by 16%. The shift and rotation parameters also underline the low accuracy of the solution with Scn6. Similar check point errors for Scn5 and Scn6 after Helmert transformation show that the usage of inaccurate GNSS measurements for all projection centre positions of the sequence as absolute information improves the overall datum but does not correct block deformations. Like already observed with the long sequence, the block deformations are smaller than the GNSS accuracy and are only significantly reduced by the integration of the building model in Scn2 and Scn4.

Conclusion

The beginning of the full sequence is an example of insufficient ground control that shows how important the distribution of model planes over the captured scene is. The results of the full sequence show that the proposed method still improves image poses given larger GNSS observation errors, insufficient model plane configuration and unfavourable image capturing conditions. The results also point at factors limiting the method: If the plane configuration is insufficient, errors in relative image orientation can get dispersed to areas of the sequence not optimally covered by model information.

6.7 Parameter Variation

In the following, the most relevant parameters of the presented approach are investigated. For each parameter, the check point errors with the standard parameter setting (see section 5.3) are compared to smaller and/or larger settings and are discussed with respect to deviations from the expected outcome.

The long sequence is used for all parameter variations to be able to compare the influence of the parameters on sliding window and global adjustment and to cover a variety of buildings. For

every parameter, first, the assumptions on adjustment results with larger and smaller settings are described, second, the resulting check point errors are discussed and compared to those with the standard parameter setting.

6.7.1 Fictitious Distance Observations of Tie Points

Two important parameters are the a priori standard deviation of fictitious observations of tie points $\sigma_{d_{TP}}$ and the distance threshold chosen as two times $\sigma_{d_{TP}}$ (see next section 6.7.2) as they reflect the assumptions made on the size of the generalisation effects. To further analyse the influence of these parameters on the estimation concerning check point errors, smaller and larger settings are investigated. This way, the choice of the distance threshold as two times $\sigma_{d_{TP}}$ is also analysed.

Standard setting: $\sigma_{d_{TP}} = 20$ cm

Larger: The less restrictive setting can result in less correction of local block deformations as the influence of the observations of fictitious distances of tie points is reduced.

Smaller: The more restrictive setting might result in lower check point errors at planes with fewer generalisation effects (depending on systematic effects of the planes); possibly larger errors at planes with generalisation effects that exceed the selected value of $\sigma_{d_{TP}}$ if these planes systematically “pull” tie points in a certain direction, e.g. zigzag façade where the zigzag is always in front of the façade plane.

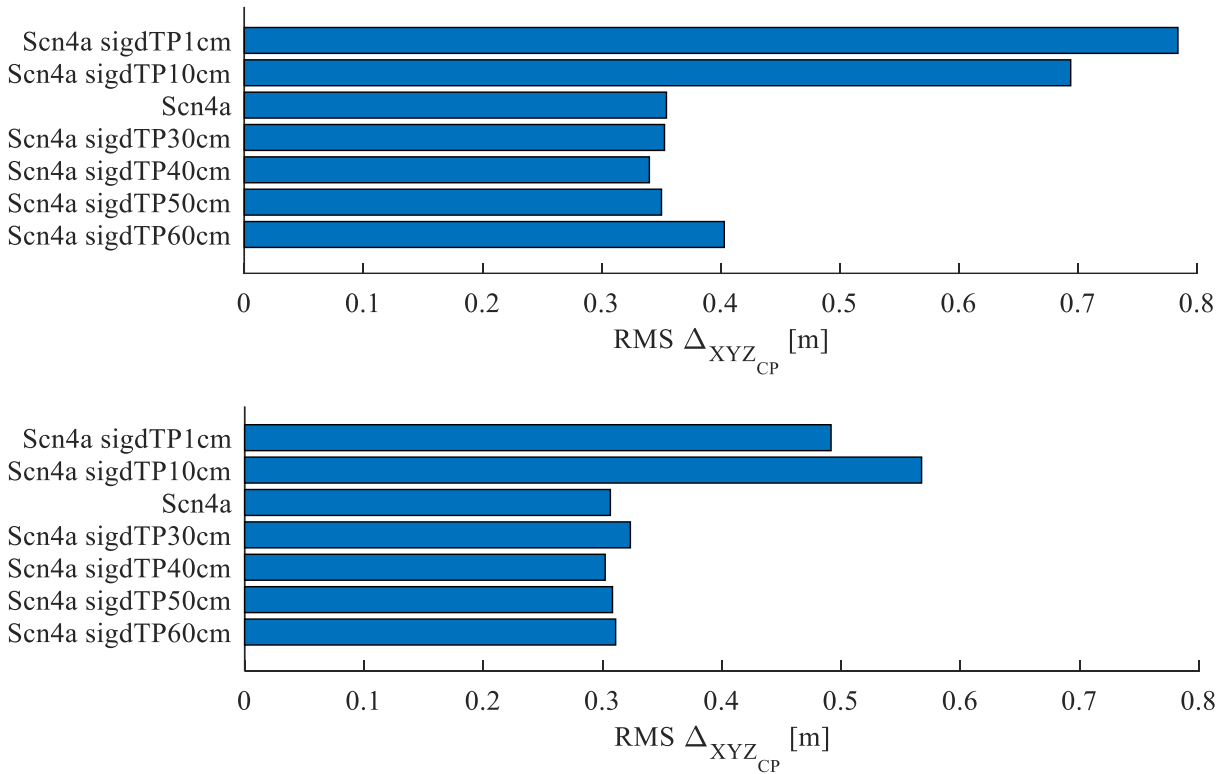


Figure 38: Variation of $\sigma_{d_{TP}}$: RMS check point errors of the long sequence in sliding window (top) and global processing (bottom).

Figure 38 shows the original Scn4 RMS check point errors with those of Scn4 calculated with smaller and larger $\sigma_{d_{TP}}$. As expected, smaller values of $\sigma_{d_{TP}}$ increase check point errors. Larger values of $\sigma_{d_{TP}}$ lead to similar results compared to the original $\sigma_{d_{TP}} = 20$ cm, not showing a notable influence of uncorrected local block deformations. Therefore, the setting of $\sigma_{d_{TP}}$ is concluded to be less critical for larger values than to smaller ones.

6.7.2 Maximum Distance of Tie Points to Model Planes

Standard setting: $2 \cdot \sigma_{d_{TP}} = 40$ cm

Larger: A larger maximum distance potentially leads to more assignments at the price of a higher probability of false assignments leading to increased check point errors.

Smaller: A reduced maximum distance leads to fewer assignments and may result in larger check point errors as correct assignments are more likely to be missed.

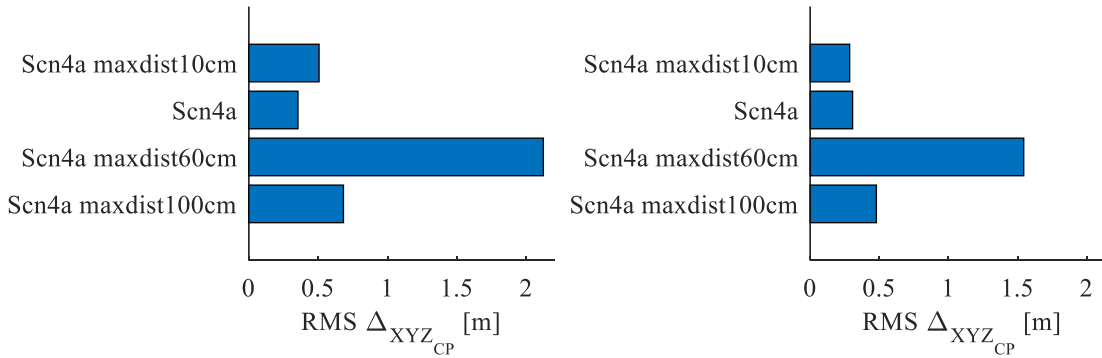


Figure 39: Variation of the maximum distance for the assignment of tie points to model planes: RMS check point errors with the long sequence in sliding window (left) and global processing (right).

According to the results shown in Figure 39, the decreased maximum distance threshold of 10 cm leads to larger check point errors in sliding window processing but does not affect the global adjustment notably. With a maximum distance of 60 cm, the check point errors are larger and even exceed 1 m. Not only is the garage door falsely assigned to the model plane, but also the neighbouring façade orthogonal to the garage gate is not related to any tie points as it is further away than the 60 cm. By chance, this changes with a maximum distance of 100 cm: Check point errors around 50 cm are the result because the garage door still is assigned falsely but the neighbouring plane is assigned correctly. The correct assignment of the neighbouring plane reduces the effect of the falsely assigned garage door. Both larger maximum distance settings compared to the standard 40 cm show that false assignments increase the overall RMS check point errors as expected.

With a maximum distance of 100 cm, fictitious distances can exceed $3 \cdot \sigma_{d_{TP}} = 3 \cdot 20$ cm = 60 cm, which means that several fictitious observations with point-to-plane distances > 60 cm are identified as outliers in the adjustment and are consequently automatically excluded from the adjustment. Nevertheless the check point errors with 100 cm are smaller than when setting the maximum distance to 60 cm. This is due to the robust adjustment iterations including the

assignments with > 60 cm point-to-plane distance before they are removed. After removal of outliers, the assignments of tie points to model planes are kept fix for the adjustment with the original weights, which prevents the adjustment to converge to the same solution than with the setting of 60 cm for the maximum distance of tie points to model planes.

The results show that check point errors are smaller if point-to-plane distances larger than $2 \cdot \sigma_{d_{TP}} = 40$ cm are rejected already before adjustment instead of introducing fictitious observations of such distances to the adjustment and let the robust adjustment reject fictitious observations with distances $> 3 \cdot \sigma_{d_{TP}}$ as outliers.

6.7.3 Estimation of Vertex Coordinates

Until here the a priori variance of direct observations of vertex coordinates σ_{VT} was set to 1 cm, almost fixing the building model in order not to mix errors of the building model and deformations of the image block. The production of the used LOD2 building model involves measurements from cadastre and from aerial laser scanning as the data source, which means that errors of several decimetres are possible.

σ_{VT} is increased to check its influence and, in addition, it is tested if the model should be fixed in the sliding window but can be refined in a global adjustment rather than already adjusting it during window-based processing.

Standard setting: $\sigma_{VT} = 1$ cm

Larger: The model planes will be estimated as the vertices are allowed to move: instead of correcting drift and local block deformations, the model planes might be moved because model errors, drift effects and block deformations together influence the adjustment. If the redundancy in the number of model planes is low, larger check point errors are to be expected as model planes can be falsely moved instead of moving the image block.

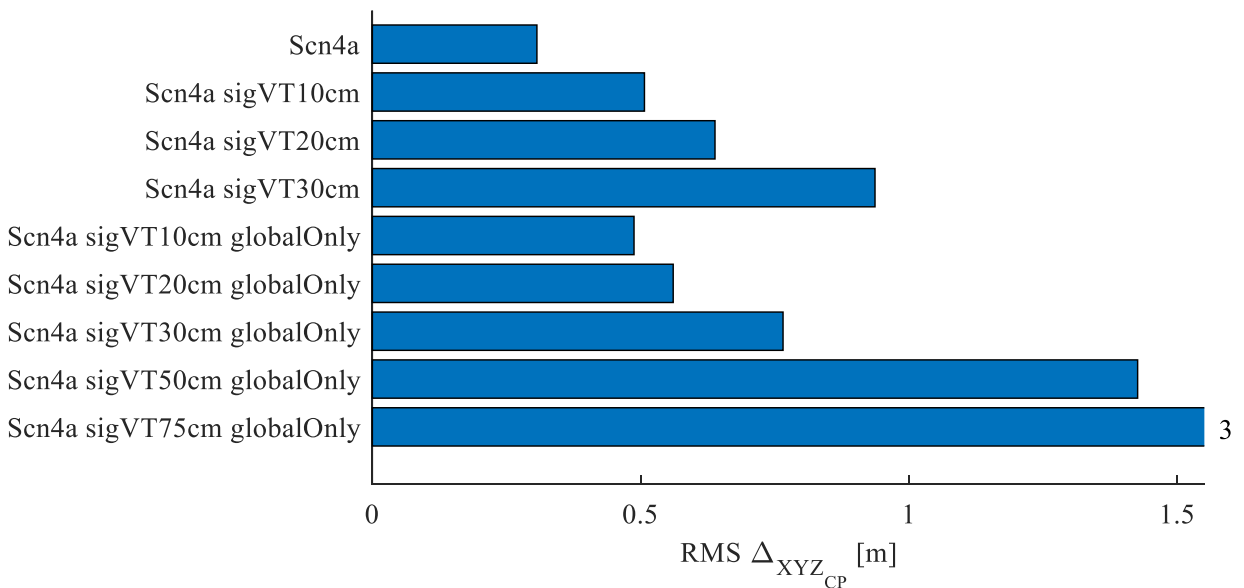


Figure 40: Estimation of the building model vertices by increasing their a priori standard deviation σ_{VT} : RMS check point errors of the long sequence after global processing.

On the other hand, the impact of erroneous model planes can be reduced if sufficient observations and enough redundancy are given.

Note that with the given sequence only one obviously erroneous plane at the façade above the garage door is covered. Assignments of tie points to this erroneous plane already are excluded in the assignment with standard parameters due to the distance threshold. Therefore, this plane does not degrade the results of the adjustment with standard parameters.

Figure 40 shows, that RMS check point errors increase with increased σ_{VT} . Only the results with the global adjustment are shown, as the sliding window processing shows the same trend.

If $\sigma_{VT} = 1$ cm is used in the windows and σ_{VT} is increased for the global adjustment only (“globalOnly”), RMS check point errors with the same setting for σ_{VT} are slightly lower. With $\sigma_{VT} > 30$ cm, the sliding window adjustment fails, due to too large distortions of the model and the image block that lead to the exclusion of all observations for image pairs in a window of the robust sliding window adjustment. Larger settings for σ_{VT} can be processed in the global adjustment and show a continuation of the trend with respect to RMS check point errors.

6.7.4 Window Size N_{ws} and Overlap N_{new}

The window size and overlap in the sliding window workflow are varied to analyse how small the incremental optimisation steps can be made and how often the adjustment has to be carried out during a flight.

Standard setting: $N_{ws} = 100$ key frames

Larger: Larger windows are not tested, as the window size is already large enough to almost lead to the same results as the global adjustment.

Smaller: A smaller window size could potentially increase check point errors if less point-to-plane correspondences were found per window and relevant planes were missing that would be needed to correct block deformations. The impact of a smaller window size on the resulting check point errors will be different depending on the number of observed objects per window.

Standard setting: $N_{new} = 33$ key frames

Larger: A larger value of N_{new} leads to a smaller window overlap, which results in a speed-up in processing time as the optimisation has to be carried out less frequently. However, the smaller the overlap is, the larger accumulated errors can become before the adjustment is carried out, risking convergence to a wrong solution due to erroneous point-to-plane assignments.

Smaller: A larger window overlap as a consequence of a lower value for N_{new} potentially reduces drift effects and block deformations of the incremental image orientation and decreases differences at key frames between windows. The pose parameters of the last images left out at the previous window are not optimised again while the following images are optimised in a next window.

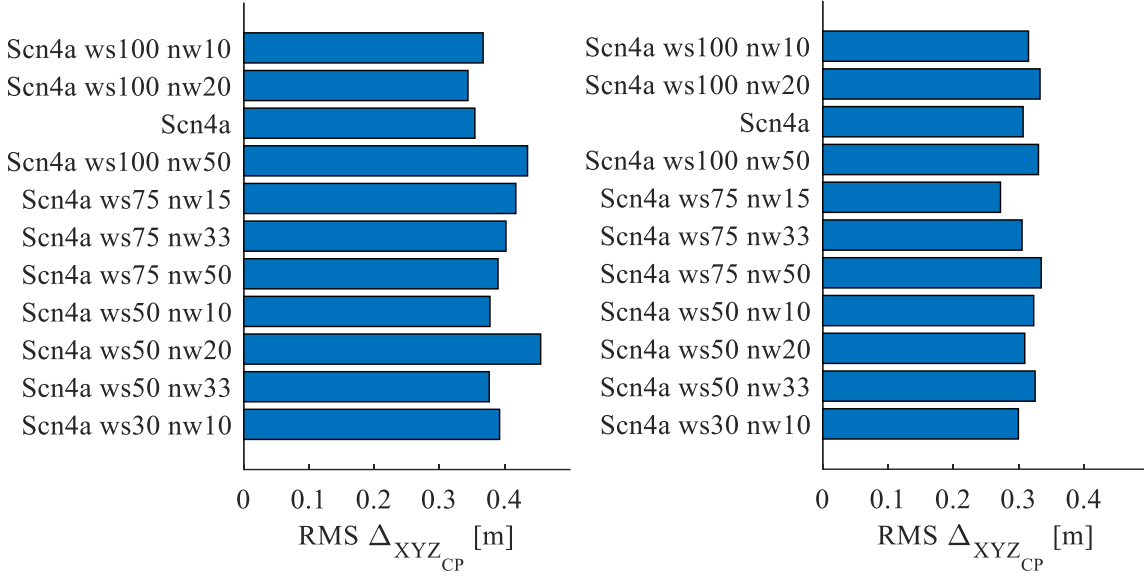


Figure 41: Variation of window size and overlap: RMS check point errors with the long sequence in sliding window (left) and global processing (right).

Figure 41 shows the results with different settings of window size and overlap. No clear trend can be identified for the settings on the results. One reason for the mixed size of the check point errors in sliding window adjustment is that in some windows of the different configurations, the tie points at the garage door are falsely assigned to the model plane. The variation of RMS check point errors is lower for the results of the global adjustment but still does not show a clear dependence on the parameter settings.

Processing with a window size smaller than 75 key frames sometimes did converge to a wrong solution in a window, which ends the processing because too many observations are detected as outliers. The processing was then restarted. For example, $N_{ws} = 30 / N_{new} = 10$ was successfully processed after a restart. The following combinations failed to be processed and therefore are not shown in Figure 41: $N_{ws} = 75 / N_{new} = 70$; $N_{ws} = 25 / N_{new} = 15$; $N_{ws} = 100 / N_{new} = 66$.

Window sizes smaller than 50 key frames and the initiation of a window adjustment less frequently than every 50 key frames are concluded to be critical with the given image sequence. The results suggest that the needed window size and overlap depend on the captured objects and potentially could be varied in future work depending on parameters like the number of found point-to-plane correspondences, number of planes, plane configuration, image overlap and quality of relative orientations, etc.

7 Conclusion and Outlook

The proposed approach for the integration of a generalised building model into photogrammetric pose estimation is shown to significantly improve UAV pose parameters even though the building model differs from the captured real scene due to generalisation effects. The two identified problems addressed by this thesis are the relation of photogrammetric measurements to a given building model and the usage of these relations for the optimisation of the UAVs pose parameters.

The optimisation is realised by a hybrid bundle adjustment (chapter 3) that simultaneously uses image observations and object information in the form of a generalised building model to retrieve optimal image poses and a sparse tie point cloud. Relations of tie points to generalised model planes are formulated in object space using three assignment methods (chapter 4). All three methods lead to improved estimation results and therefore are concluded to integrate the generalised building model successfully. Coordinate differences at check points show that in comparison to a pose estimation without a building model, the hybrid adjustment improves check point errors from meter range to decimetre level. This improvement indicates that a generalised building model is a potential data source to use in UAV applications where other measurements like GNSS or GCPs are inaccurate or unavailable.

Compared to a detailed building model with reduced generalisation effects, larger check point errors occur with the generalised building model. This loss of accuracy due to the usage of only a generalised building model is further analysed by correcting the generalised model for systematic effects in the model planes. The correction is done based on a reference point cloud using the hybrid bundle adjustment's ability to estimate model plane parameters while preserving the model topology. Check point errors improve with the resulting corrected generalised model instead of the original generalised building model. This improvement shows that systematic effects of the model planes as a second issue next to the generalisation partly cause differences in check point errors between the original generalised and the detailed building model. The remaining difference between the corrected generalised model and the detailed one represents the loss in accuracy due to generalisation effects.

Based on pairs of known and estimated check point coordinates, the seven parameters of Helmert transformations were calculated for each scenario to analyse systematic effects that affect the datum of the whole image block. The transformations applied to estimated check point coordinates lead to check point errors that are reduced by overall datum effects. The check point errors after transformation for the short sequence are smallest with the detailed building model.

Larger check point errors after transformation with the generalised building model are a sign of remaining block deformations. These deformations are likely caused by systematic effects of model planes and insufficient control information due to the generalisation of the generalised building model compared to the detailed one. For the long sequence, remaining check point errors are smaller after transformation when using the generalised building model compared to not using any building model. This decrease in errors shows that block deformations that occur without any model are reduced by integrating the generalised building model.

The flight along a large façade shows a situation where block deformations are dispersed along a façade plane, leading to relatively large remaining check point errors even after Helmert transformation. The façade plane as the only ground control in this area is insufficient to locally fix the block, allowing block deformations to be distributed along this plane. The effect is reduced if the corrected generalised building model is used, as deformations of the block are smaller with this model. The beginning of the full sequence is another example of insufficient ground control that shows how important the distribution of model planes over the captured scene is. The results of the full sequence show that the proposed method still improves image poses given larger GNSS observation errors, insufficient model plane coverage and unfavourable image capturing conditions.

The comparison of the three assignment methods leads to the conclusion that the simple distance criterion leads to results comparable to those of the dedicated plane search in ROIs around model planes. In comparison to these two assignment methods (a) and (c), the model-independent detection of planes in the whole tie point cloud that are then matched with model planes (method (b)), leads to inferior results. Nevertheless, method (b) shows the best potential to be used in future work if the assumption of good initial values for tie point coordinates in the coordinate system of the building model is violated. The experiments show no indication to prefer the more complex method (c) to the most straightforward method (a).

The results of the global adjustment are compared to those of the sliding window workflow. Using the standard parameters, the improvement of global optimisation over the sliding windows concerning check point errors is relatively small. Window size and overlap are sufficient to cover most parts of the sequence in a way that locally relevant observations and unknowns take part in the local optimisations. Long sequences can be processed in windows if the global solution would fail due to too large accumulating drift errors or unfeasible processing time. Furthermore, the windows potentially enable the application of the algorithm during data capture to incrementally retrieve corrected pose parameters and support or even replace loop-closure in SLAM.

Related work in comparison to this thesis did not explicitly address the problem of using relatively highly detailed and accurate measurements with object knowledge of relatively low detail, low accuracy and essential generalisation effects. The developed method to relate such measurements to object information and integrate them into a common optimisation framework is shown to successfully address the two problems identified in the introduction. Although the generalised building model is less accurate and of lower detail, it is shown to be helpful to improve image orientation.

Several follow up questions remain for future work on this topic. The assignment of tie points to model planes potentially can be refined in several aspects. The mentioned combination of assignment methods can be taken further to consider other assignment methods from related work: For the initialisation and also for additional correspondences, other features and matching in image space or between image and object space should be considered. Hierarchical approaches could address the assignment problem, e.g., to first solve the problem in lower detail before the generalisation effects come into play at the full resolution. In the assignment, thresholds common to all assignments of tie points to planes are used.

Thresholds, e.g. the distance threshold, could be set individually per plane based on individual statistics like histograms of distances per plane or the local point distribution. Planes with overall close by points could be assumed to have fewer generalisation effects. Consequently, a reduced distance threshold and an increased a priori accuracy of the fictitious observations that relate tie points to this plane could be set.

Distance histograms could also be investigated to identify unreliable planes with ambiguous assignments. The example of a garage door shows that the plane correspondences could be eliminated if two peaks in the distance histogram are detected. The peaks are likely to represent two planar point clusters, making it ambiguous to assign one or the other to the model plane.

Observing assignments over several iterations is another option: Points assigned continuously to a plane might be more reliable than those where the assignment switches between iterations. Another idea would be to vary assignment configurations and compare different assignment variants to pursue several solutions and identify the best one.

The identified difference between tie point errors and estimated standard deviations due to systematic effects is another topic to address in future work: The assumed Gaussian distribution for the fictitious observations simplifies the real error distribution. The distance histograms per plane could be a hint to possible refinements, e.g. by using more individual or overlapping multiple error distributions. Estimated standard deviations that represent all occurring errors would allow for the refinement of the assignment strategies: The estimated standard deviation could be taken into account to refine thresholds in the assignment of tie points to model planes.

The proposed method for the point-to-plane assignment does not use semantic information. A classification in image or object space could be used to eliminate false assignments. Points assigned to classes that are not represented in the building model, e.g. vegetation, street lights and signs, cars, windows, doors, etc., could be excluded from the assignment.

To further investigate the influence of generalisation effects, a detailed model of more and different buildings, without generalisation and systematic effects is needed. Additionally, lower and higher levels of model generalisation should be tested to find out which amount of accuracy of estimated pose parameters is lost or gained depending on those levels. The difference between measurement resolution and model detail could also be varied by changing the distance of the camera to the captured buildings or by changing the image resolution. In the first case, more or

fewer model features would become visible; in the second case, the discrepancy between model details and measured details would be varied.

The described sliding window workflow could be extended and further analysed in future work. Window size and overlap were constant in the experiments but could be set automatically depending on different criteria such as travelled distance, number of found point-to-plane assignments, plane configuration or a posteriori standard deviation, e.g., of estimated tie point coordinates or pose parameters. The transition between windows is another step that leaves room for optimisation. Sliding window processing and filtering approaches are covered in literature for various tasks in more detail. For this thesis, the focus was not on investigating real-time approaches. The set up sliding window processing shows that such approaches potentially can be applied to the hybrid adjustment of images and generalised object knowledge.

Several aspects of the implemented workflow need to be addressed to reduce processing time and memory consumption to apply the method in real-time onboard of a UAV. The experiments show that without dense matching, sufficient correspondences between tie points and model information are found. However, the number of tie points still might be higher than needed for pose refinement and assignment of points to planes. Future work, therefore, could further reduce the number of points and planes handled, skip point-to-plane assignment in some iterations or even pursue image-only odometry and sliding windows until drifts become larger and only then integrate the building model for a window to reset drifts.

Only one camera was used here. The combination of several cameras and the influence of parameters like the opening angle and viewing directions are to be tested. In addition, the assignment in object space could be used to relate points of any sensor that delivers object points to a building model leading to further experiments and development of the method. Simulations with a filtering approach using a laser scanner with the assignment method (a) applied to the laser scanning object points and a non-generalised building model already show promising results (Bureick et al., 2019).

In summary, the two objectives of this thesis have been achieved. Relations of a photogrammetric block to a generalised building model, despite generalisation effects, were successfully formulated, and the experiments show an improvement of the pose estimation from metre range without using a building model to decimetre range with a generalised building model. The presented method is a promising addition or alternative to classical sensor orientation approaches to improve UAV pose estimation for various applications by using available object information in form of a generalised building model.

References

- Aider, O.A., Hoppenot, P., Colle, E., 2005. A model-based method for indoor mobile robot localization using monocular vision and straight-line correspondences. *Robotics and Autonomous Systems* 52 (2-3), pp. 229–246.
- Arth, C., Pirschheim, C., Ventura, J., Schmalstieg, D., Lepetit, V., 2015. Instant outdoor localization and SLAM initialization from 2.5D maps. *IEEE transactions on visualization and computer graphics* 21 (11), pp. 1309–1318.
- 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.
- Besl, P.J., McKay, N.D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (2), pp. 239–256.
- Briskin, G., Geva, A., Rivlin, E., Rotstein, H., 2017. Estimating Pose and Motion Using Bundle Adjustment and Digital Elevation Model Constraints. *IEEE Transactions on Aerospace and Electronic Systems* 53 (4), pp. 1614–1624.
- Bureick, J., Vogel, S., Neumann, I., Unger, J., Alkhatib, H., 2019. Georeferencing of an Unmanned Aerial System by Means of an Iterated Extended Kalman Filter Using a 3D City Model. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 87 (5), pp. 229–247.
- Canny, J., 1986. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8* (6), pp. 679–698.
- 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.
- 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., 2012. Echtzeit-Georegistrierung von Videodaten mit Hilfe von Navigationssensoren geringer Qualität und digitalen 3D-Landschaftsmodellen. PhD Thesis. Institut für Geodäsie und Photogrammetrie, Zürich.
- Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24 (6), pp. 381–395.

- Förstner, W., Wrobel, B.P., 2016. *Photogrammetric Computer Vision*. Springer International Publishing, Cham, Switzerland.
- Frueh, 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.
- Glira, P., 2018. Hybrid orientation of LiDAR point clouds and aerial images. PhD Thesis, Wien.
- Hebel, M., Stilla, U., 2012. Simultaneous Calibration of ALS Systems and Alignment of Multiview LiDAR Scans of Urban Areas. *IEEE Transactions on Geoscience and Remote Sensing* 50 (6), pp. 2364–2379.
- 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.
- Hsu, S., Samarasekera, S., Kumar, R., Sawhney, H.S., 2000. Pose estimation, model refinement, and enhanced visualization using video. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000* (Cat. No.PR00662). IEEE Comput. Soc, pp. 488–495.
- Huber, P.J., 1981. *Robust Statistics*. Wiley, New York.
- Iwaszczuk, D., Hoegner, L., Schmitt, M., Stilla, U., 2012. Line based Matching of Uncertain 3D Building Models with IR Image Sequences for Precise Texture Extraction. *Photogrammetrie - Fernerkundung - Geoinformation* (5), pp. 511–521.
- Iwaszczuk, D., Stilla, U., 2017. Camera pose refinement by matching uncertain 3D building models with thermal infrared image sequences for high quality texture extraction. *ISPRS Journal of Photogrammetry and Remote Sensing* 132, pp. 33–47.
- Jung, J., Sohn, G., Bang, K., Wichmann, A., Armenakis, C., Kada, M., 2016. Matching Aerial Images to 3D Building Models Using Context-Based Geometric Hashing. *Sensors (Basel, Switzerland)* 16 (6, 932).
- 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.
- Kschischang, F.R., Frey, B.J., Loeliger, H.-A., 2001. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory* 47 (2), pp. 498–519.
- 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.
- Li-Chee-Ming, J., Armenakis, C., 2018. UAV navigation system using line-based sensor pose estimation. *Geo-spatial Information Science*, pp. 1–10.
- Lothe, P., 2011. *Localisation et cartographie simultanées par vision monoculaire contraintes par un SIG: application à la géolocalisation d'un véhicule*. PhD Thesis, Clermont-Ferrand.

- 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.
- Monnier, F., Vallet, B., Paparoditis, N., Papelard, J.-P., David, N., 2013. Registration of terrestrial mobile laser data on 2D or 3D geographic database by use of a non-rigid ICP approach. ISPRS - International Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-5/W2, pp. 193–198.
- Nex, F., Remondino, F., 2014. UAV for 3D mapping applications: a review. Applied Geomatics 6 (1), pp. 1–15.
- Open Geospatial Consortium, 2012. OGC city geography markup language (CityGML) encoding standard, Version 2.0.
- 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.
- Rodriguez, J.J., Aggarwal, J.K., 1990. Matching aerial images to 3-D terrain maps. IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (12), pp. 1138–1149.
- 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.
- Stamos, I., Allen, P.K., 2001. Automatic registration of 2-D with 3-D imagery in urban environments. Eighth IEEE International Conference on Computer Vision. Proceedings: July 7-14, 2001, Vancouver, British Columbia, Canada. IEEE Computer Society, Los Alamitos, Calif., pp. 731–736.
- Strasdat, H., Montiel, J.M.M., Davison, A.J., 2012. Visual SLAM: Why filter? Image and Vision Computing 30 (2), pp. 65–77.
- 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.
- Talluri, R., Aggarwal, J.K., 1996. Mobile robot self-location using model-image feature correspondence. IEEE Transactions on Robotics and Automation 12 (1), pp. 63–77.
- Tamaazousti, M., Gay-Bellile, V., Collette, S.N., Bourgeois, S., Dhome, M., 2011. NonLinear refinement of structure from motion reconstruction by taking advantage of a partial knowledge of the environment. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 20 - 25 June 2011, Colorado Springs, CO, USA. IEEE, Piscataway, NJ, pp. 3073–3080.
- Torr, P.H.S., Zisserman, A., 2000. MLESAC. A New Robust Estimator with Application to Estimating Image Geometry. Computer Vision and Image Understanding 78 (1), pp. 138–156.
- Urban, S., 2016. Real-Time Multi-Fisheye Camera Self-Localization and Egomotion Estimation in Complex Indoor Environments. PhD Thesis, Karlsruhe.

- 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.
- Wang, T., Wang, C., Liang, J., Chen, Y., Zhang, Y., 2013. Vision-Aided Inertial Navigation for Small Unmanned Aerial Vehicles in GPS-Denied Environments. *International Journal of Advanced Robotic Systems*, pp. 1–12.
- West, K.F., Webb, B.N., Lersch, J.R., Pothier, S., Triscari, J.M., Iverson, A.E., 2004. Context-driven automated target detection in 3D data. In: Sadjadi, F.A. (Ed.), *Automatic Target Recognition XIV*. SPIE, pp. 133–143.
- Zhao, W., Nister, D., Hsu, S., 2005. Alignment of continuous video onto 3D point clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (8), pp. 1305–1318.
- Zhou, G., Fang, L., Tang, K., Zhang, H., Wang, K., Yang, K., 2015. Guidance: A Visual Sensing Platform For Robotic Applications. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 9–14.

Curriculum Vitae

Personal Information

Jakob Unger

04/1986 at Hannover, Germany

Work Experience

11/2012 – 12/2019	Institute of Photogrammetry and GeoInformation (IPI) Leibniz University Hannover <i>Research Associate</i>
08/2014 – 11/2014	Pontifical Catholic University (PUC) Rio de Janeiro, Brazil <i>Secondment</i>

Education

10/2009 – 05/2012	Geodesy and Geoinformatics Leibniz University Hannover <i>Master of Science</i>
10/2006 – 02/2010	Geodesy and Geoinformatics Leibniz University Hannover <i>Bachelor of Science</i>
1993 – 2006	Freie Waldorfschule Hannover-Maschsee <i>Abitur</i>

Acknowledgements

My deepest thanks go to all the companions and supporters who have been with me during the creation of this work. First of all, Christian Heipke and Franz Rottensteiner are to be honoured who gave me their trust and enabled me to go my way to realise such a work. The detailed discussions and constructive feedback have helped me to focus on the goal and to strengthen my view of the methodology. I very much appreciate the openness to new challenges, e.g. also in dealing with numerous drone flights. I thank Ingo Neumann for acting as a co-examiner and for his cooperation in the UAV project, which has enriched my work and provided valuable feedback. I also thank Markus Gerke for acting as a co-examiner.

I am grateful for the excellent, educational and inspiring time I had at IPI. I thank the colleagues who were part of the team over the years and made this a great time for me. I enjoyed being part of this awesome group that has become more and more international. Especially Martin and Artuom I would like to thank for the cooperation and many enjoyable hours in the fresh air with UAV flights. Thanks to Martin, Lukas and Jonas for proofreading this work and to Uwe Br. for his master thesis, which supported my evaluation. My thanks don't go only to the staff of IPI, but also to those of the other institutes of Geodesy and Geoinformatics at the Leibniz University Hannover. The joint public relations work of the discipline was a valuable addition and diversion to my research activities.

Finally, I thank all those who have contributed in private to make this work possible. My family has laid the foundation to face challenges with interest and to enjoy solving them. Linda has contributed a substantial amount to the necessary breaks and moments of relaxation, which have been beneficial to my work as a balance. Thank you very much for all good words of encouragement, trust and refreshing distractions.

Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover

(Eine vollständige Liste der Wiss. Arb. ist beim Geodätischen Institut, Nienburger Str. 1, 30167 Hannover erhältlich.)

Nr. 334	HOFMANN, Sabine:	Potential von LiDAR Mobile Mapping für hochgenaue Karten (Diss. 2017)
Nr. 335	ALBERT, Lena:	Simultane Klassifikation der Bodenbedeckung und Landnutzung unter Verwendung von Conditional Random Fields (Diss. 2017)
Nr. 336	NIEMEYER, Joachim:	Verwendung von Kontext zur Klassifikation luftgestützter Laserdaten urbaner Gebiete (Diss. 2017)
Nr. 337	EGGERT, Daniel:	Effiziente Verarbeitung und Visualisierung von Mobile Mapping Daten (Diss. 2017)
Nr. 338	FITZNER, Daniel:	Estimation of Spatio-Temporal Moving Fields at High Resolution (Diss. 2017)
Nr. 339	CZIOSKA, Paul:	Meeting point locations for shared rides (Diss. 2018)
Nr. 340	FEUERHAKE, Udo:	Erfassung von Trajektorien und Erkennung von Bewegungsmustern (Diss. 2018)
Nr. 341	HUANG, Hai :	Bayesian Models for Pattern Recognition in Spatial Data (Habil. 2018). (Nicht im Layout unserer Schriftenreihe erschienen, sondern als Heft DGK Reihe C, Heft 751)
Nr. 342	BOCK, Urs Fabian:	Dynamic Parking Maps from Vehicular Crowdsensing (Diss. 2018)
Nr. 343	KRAWINKEL, Thomas:	Improved GNSS Navigation with Chip-scale Atomic Clocks (Diss. 2018)
Nr. 344	GOSWAMI, Sujata:	Understanding the sensor noise in the GRACE range-rate observations by analyzing their residuals (Diss. 2018)
Nr. 345	ZHANG, Yin:	Expected Utilities for Decision Making and Steering of Measurement Processes (Diss. 2018)
Nr. 346	SCHLICHTING, Alexander:	Fahrzeuglokalisierung durch Automotive Laserscanner unter Verwendung statischer Merkmale (Diss. 2018)
Nr. 347	RÖTH, Oliver	Extraktion von hochgenauer Fahrspurgeometrie und -topologie auf der Basis von Fahrzeugtrajektorien und Umgebungsinformationen (Diss. 2018)
Nr. 348	NEELMEIJER, Julia:	Observing Inter- and Intra-Annual Glacier Changes and Lake Loading Effects from Synthetic Aperture Radar Remote Sensing (Diss. 2018)
Nr. 349	HOBERG, Thorsten:	Conditional Random Fields zur Klassifikation multitemporaler Fernerkundungsdaten unterschiedlicher Auflösung (Diss. 2018)
Nr. 350	SCHILLING, Manuel:	Kombination von klassischen Gravimetern mit Quantensensoren (Diss. 2019)
Nr. 351	MILLER, Dominik:	Seismic noise analysis and isolation concepts for the ALPS II experiment at DESY (Diss. 2019)
Nr. 352	ALI, Bashar:	Optimierte Verteilung von Standorten der Schulen unter dem Einfluss des demografischen Wandels am Beispiel Grundschulen (Diss. 2019)
Nr. 353	ZHAO, Xin:	Terrestrial Laser Scanning Data Analysis for Deformation Monitoring (Diss. 2019)
Nr. 354	HAGHIGHI, Mahmud Haghshenas:	Local and Large Scale InSAR Measurement of Ground Surface Deformation (Diss. 2019)
Nr. 355	BUREICK, Johannes:	Robuste Approximation von Laserscan-Profilen mit B-Spline-Kurven (Diss. 2020)
Nr. 356	BLOTT, Gregor:	Multi-View Person Re-Identification (Diss. 2020)
Nr. 357	MAAS, Alina Elisabeth:	Klassifikation multitemporaler Fernerkundungsdaten unter Verwendung fehlerbehafteter topographischer Daten (Diss. 2020)
Nr. 358	NGUYEN, Uyen:	3D Pedestrian Tracking Using Neighbourhood Constraints (Diss. 2020)
Nr. 359	KIELER, Birgit:	Schema-Matching in räumlichen Datensätzen durch Zuordnung von Objektinstanzen (Diss. 2020)
Nr. 360	PAUL, Andreas:	Domänenadaption zur Klassifikation von Luftbildern (Diss. 2020)
Nr. 361	UNGER, Jakob:	Integrated Estimation of UAV Image Orientation with a Generalised Building Model (Diss. 2020)

Die Arbeiten werden im Rahmen des wissenschaftlichen Schriftenaustausches verteilt und sind nicht im Buchhandel erhältlich. Der Erwerb ist zu einem Stückpreis von € 25,00 bei den herausgebenden Instituten möglich.