

## A GRAPH BASED BUNDLE ADJUSTMENT FOR INS-CAMERA CALIBRATION

D. Bender<sup>a,b</sup>, M. Schikora<sup>a,b</sup>, J. Sturm<sup>b</sup>, D. Cremers<sup>b</sup>

<sup>a</sup> Department Sensor Data and Information Fusion, Fraunhofer FKIE, Wachtberg, Germany  
(daniel.bender, marek.schikora)@fkie.fraunhofer.de

<sup>b</sup> Computer Vision Group, Technical University of Munich, Garching, Germany  
(sturmju, cremers)@in.tum.de

**KEY WORDS:** GPS/INS, Camera, Non-Metric, Calibration, Georeferencing, Aerial, Bundle Adjustment, Graph Optimization

### ABSTRACT:

In this paper, we present a graph based approach for performing the system calibration of a sensor suite containing a fixed mounted camera and an inertial navigation system. The aim of the presented work is to obtain accurate direct georeferencing of camera images collected with small unmanned aerial systems. Prerequisite for using the pose measurements from the inertial navigation system as exterior orientation for the camera is the knowledge of the static offsets between these devices. Furthermore, the intrinsic parameters of the camera obtained in a laboratory tend to deviate slightly from the values during flights. This induces an in-flight calibration of the intrinsic camera parameters in addition to the mounting offsets between the two devices. The optimization of these values can be done by introducing them as parameters into a bundle adjustment process. We show how to solve this by exploiting a graph optimization framework, which is designed for the least square optimization of general error functions.

### 1 INTRODUCTION

There is ongoing research regarding the photogrammetric usage of small unmanned aerial systems (UAS) with payload capabilities of a few kilograms. These platforms are usually equipped with a GPS corrected inertial navigation system (INS) and a camera (Eisenbeiss, 2008). The exterior orientations of the aerial images are determinable by the usage of a bundle adjustment (BA) software. Prerequisites are a sufficient number of ground control points (GCP) and high image overlaps between adjacent flight strips. The observations of the GCPs have to be identified in the images as input for the BA procedure. This makes clear that the BA is a time consuming post processing step, which has to be performed after the measurement flight is completed.

In contrast to this development, time-critical surveillance and rescue tasks have a high demand for a more flexible flight planning and the direct determination of object positions (Schikora et al., 2010). Despite the great development in visual navigation (Engel et al., 2012), the usage of a high-precision INS leads to the most accurate realtime pose information in outdoor scenarios. The utilization of the measured positions and attitudes for the exterior camera orientations eliminates the need for the time consuming post processing in form of a BA to obtain the camera poses. Required is the knowledge of the static coordinate system transformation between the rigid mounted sensors. More precisely the position offset (lever-arm) and the angle misalignments (boresight) have to be estimated (Figure 1). This is known as INS-camera calibration. The most accurate calibration procedures estimate these parameters with an extended Kalman filter (Weiss and Achtelik, 2012) or integrate them as unknowns in a BA (Pinto and Forlani, 2002).

Presently, no freely available BA package is able to perform an INS-camera calibration and the adaptation of the BA implementations is an error prone and time consuming task. In contrast to this, the general graph optimization framework ( $g^2o$ ) is directly designed for the least square optimization of general error functions (Kuemmerle and Grisetti, 2011). The problem has to be embedded in a graph by representing the parameters to be optimized as vertices and the observations between them as edges. Further requirements are the definition of error functions for the



Figure 1: The calibration of the static coordinate system offsets between a camera and an INS enables the direct georeferencing of images collected with small UAS.

observations and good initial values for the state variables. The numerical solution of the problem can be computed with an implementation of the Levenberg-Marquardt algorithm.

In this paper, we present a graph based BA approach for the INS-camera calibration. The paper starts with an overview on related research areas. It follows a description of the system setup which is focused on in this work. After the general definition of the problem, it is restated into a graph structure and the design of the error functions is described. Finally the approach is analyzed in numerical studies and the achieved results are discussed.

### 2 RELATED WORK

A related process of the INS-camera calibration is the so called hand-eye calibration. Given a camera mounted on a robot arm, the rigid-body transformation between the coordinate systems of these devices is estimated. As a result, measurements from the acquired images can be transformed in the robot arm coordinate system. This is necessary to interact with objects recognized and located in the images. The calibration is done by estimating the rigid body transformation from corresponding poses. A direct solution can be computed by firstly optimizing the rotational part and solving the equations for the translation afterwards (Tsai and Lenz, 1989). In contrast, it was shown that the nonlinear

optimization for rotation and translation at the same time leads to more robust results in case of noise and measurement errors (Horn and Dornika, 1995). The motion of the robot arm is typically obtained from encoders, whereas nearly all approaches determine the camera movement by observing a calibration pattern. In contrast to that, the camera poses can be determined by a structure-from-motion approach. As a drawback, the camera movement can only be estimated up to a similarity transformation. This leads to an unknown scale, which as well has to be estimated during the calibration process. The obtained results are not as accurate as the calibrations from methods using camera calibration patterns (Andreff et al., 2001). However the approach has the advantage of being feasible without any additional equipment and therefore allows a recalibration during operation.

The algorithms developed for the hand-eye calibration had a big influence on another related problem: The calibration between an inertial measurement unit (IMU) and a camera. These devices can be combined to a vision-aided inertial navigation system. Measurements of the IMU in form of rotational velocities and linear accelerations can be integrated to determine the positions, velocities and attitudes of the device. During this process small estimation errors are summed up over time, which should be corrected by an aided sensor. Typically this is achieved by exploiting GPS measurements, which is not possible in space, underwater or indoor applications. An alternative is the usage of a camera in addition to the IMU. Prerequisite for exploiting camera based corrections is the knowledge of the transformation between the two devices. These offsets can be computed with modified hand-eye calibration algorithms (Lobo and Dias, 2007). The fact that consecutive measurements from shaft encoders are uncorrelated in contrast to data from an IMU results in different noise characteristics. Considering these time correlations leads to a higher accuracy of the estimations (Mirzaei and Roumeliotis, 2008). The authors utilize an extended Kalman filter (EKF) for estimating the pose transformation, which facilitates the computation of the covariance for the estimated parameters as an indicator of the achieved accuracy. In (Weiss and Achtelik, 2012) the authors describe an approach for navigating an UAS based on measurements from an IMU and a camera. By not using a calibration pattern, they realize the online estimation of the mounting parameters between these sensors with an EKF formulation.

To calibrate the transformation between a GPS-aided INS and a camera, the algorithms for the IMU-camera calibration can be used. This is based on the fact that an IMU is a fundamental component of an INS. The only prerequisite for applying an IMU-camera calibration is that the raw measurements of the IMU are accessible. However this is not the predominantly used method to perform a system calibration between an INS and a camera. The commercially available INS provide an integrated filtering process, exploiting the GPS measurements to correct the IMU estimations. In conjunction with GPS correction signals from ground control stations an INS accuracy in the range of a few centimeters for the position and a few hundredths of a degree for the attitude is achievable. Thus the INS provides a reliable stand alone source describing its own movement. This leads to the estimation of the rigid-body transformation between the INS and the camera with methods similar to the hand-eye calibration. In a first step the camera movement is calculated with a structure from motion (SFM) approach and refined in a BA procedure. The observations of ground control points are used to scale the 3D-model to real world coordinates. In a second step, the transformation between the two devices is estimated by relating these absolute camera poses to time synchronized measurements from the INS. This widely used approach is known as two step procedure (Cramer et al., 2000). The advantage is that each bundle

adjustment package can be used without modifications. On the other hand, the integration of the mounting parameters as variables to optimize in the BA is possible. This approach is known as single-step calibration and induces a simpler calibration of the mounting offsets due to more flexible flight courses (Pinto and Forlani, 2002). The simultaneous optimization of the rigid-body transformation between the devices and the intrinsic camera parameters should consider correlations between these values. An analysis of the flight pattern influence on the calibration parameters is discussed in (Kersting et al., 2011). The authors also state that at least one GCP is needed for the estimation of the vertical lever-arm and that the addition of multiple GCP does not improve the estimation results identifiable. The integrated sensor orientation has been examined and discussed widely within the OEEPE test (Heipke et al., 2002). They conclude that it is a serious alternative for many applications, even though it does not achieve the accuracies of the classical bundle adjustment.

The investigations for the INS-camera calibration performed in the last decades are targeted at manned aircrafts equipped with high-precision INS and aerial cameras at high altitude. In contrast to this our goal is the direct georeferencing with small UAS. In this work, we investigate how the calibration process can be performed with an open source toolkit for general graph optimization to bypass the need for a commercial BA package.

### 3 SYSTEM SETUP

The objective of this research is the estimation of the rigid body transformation between a camera and a high-precision INS. The latter is based on fiber optic gyroscopes, which have a stability up to some hundredths of a degree per hour. In combination with real time kinematic enhanced GPS measurements very accurate pose information are generated. An absolute accuracy of  $\pm 2\text{cm}$  in the position and  $0.01^\circ$  for the attitude angles are achieved in our system setup. This pose information shall be exploited for the determination of the exterior orientation of an optical camera. Precondition for this are a rigid mounting and a hardware synchronization between the two devices. If these conditions are fulfilled, the mounting offsets (Figure 2) can be estimated by processing data from a measurement flight.

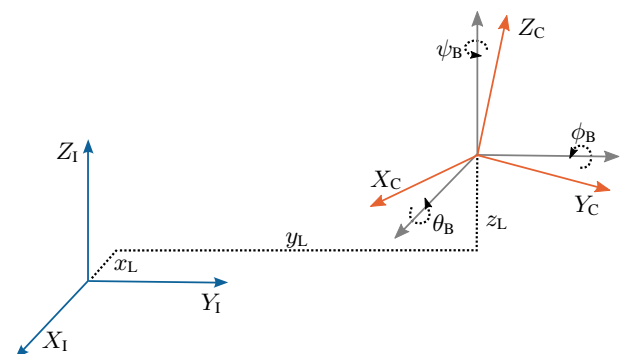


Figure 2: The rigid mounting of a camera and an INS induces static offsets between the underlying coordinate systems denoted by C and I. More precisely position offsets  $x_L$ ,  $y_L$  and  $z_L$  as well as angle misalignments, here visualized in form of Euler angles yaw =  $\psi_B$ , pitch =  $\theta_B$  and roll =  $\phi_B$ , arise. These have to be determined in an INS-camera calibration.

Nearly every outdoor application utilizing a camera mounted on an aerial platform induces that the camera focus has been set to infinity. Furthermore, wide angle lenses are frequently the right choice for a given task. The minimal distance for a focused image and the big opening angle lead to an impractical size for the

calibration pattern and make standard camera calibration procedures laborious. An easier estimation of these parameters can be achieved by performing an in-flight calibration, which uses a structure from motion (SFM) approach on the images acquired during a flight. Another advantage of this is that the intrinsic camera parameters are estimated during real conditions, which leads to slightly different values in comparison to a laboratory calibration due to climate and temperature changes (Jacobsen, 2001).

The INS measures the attitude as Euler angles, namely yaw, pitch and roll, with regard to local navigation systems. The latter are local East, North, Up (ENU) coordinate systems, each with the origin in the current device position described through a GPS measurement in form of latitude, longitude and altitude (Figure 3). This implies tiny differences between the orientations of these coordinate systems and thus also in the measured attitude for successive timestamps. Our INS-camera calibration will be performed in an area of only a few hundred square meters, which allows to neglect these differences. Besides, we perform the INS-camera calibration in an ENU coordinate system with the origin at a ground control point in the middle of the observed area. This prevents the need for earth curvature corrections which occur in mapping frames like the universal transverse mercator (UTM) coordinate system. To obtain the transformation from the position information of the INS given in latitude, longitude and altitude to ENU coordinates the earth-centered, earth-fixed coordinate system (Figure 3) is used as an intermediate step.

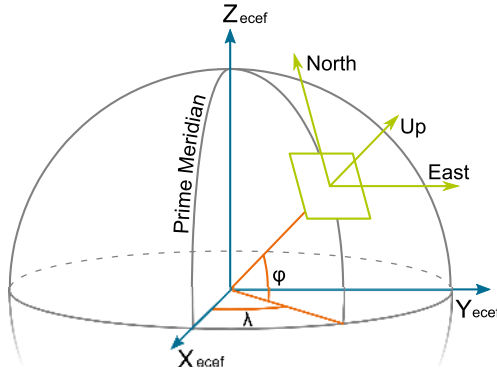


Figure 3: Both, the navigation systems of the INS and the world reference frame are based on East, North, Up (ENU) coordinate systems (green). They are local Cartesian coordinate systems with the origin tangential to the earth ellipsoid. Further we use the global earth-centered, earth fixed (ECEF) coordinate system (blue), to convert between ENU-coordinates and GPS measurements taken in latitude ( $\varphi$ ) and longitude ( $\lambda$ ) as polar coordinates (orange).

#### 4 DEFINITIONS

We define the world reference frame  $W$  as ENU coordinate system with the origin being approximately in the middle of our calibration area. The rigid body motion  $g_{c,w}$  describes the camera pose at the middle exposure time  $t_i, i = 1, 2, \dots, n$  with regard to the world frame  $W$ . Likewise specifies  $g_{i,w}$  the corresponding configuration of the INS as rigid body motion.

In general, a rigid body motion  $g \in SE(3)$  describes how the points of a rigid object change over time. Instead of considering the continuous path of the movement, we bring into focus the mapping between the initial and the final configuration of the rigid body motion. This movement can be described by a rotation matrix  $\mathbf{R} \in SO(3)$  and a translation vector  $\mathbf{t} \in \mathbb{R}^3$ . Conse-

quently the rigid body displacement  $G$  of a 3D point  $\mathbf{p} \in \mathbb{R}^3$  can be performed by

$$G : SE(3) \times \mathbb{R}^3 \rightarrow \mathbb{R}^3, \quad G(g, \mathbf{p}) = \mathbf{R}\mathbf{p} + \mathbf{t}. \quad (1)$$

The representation of the rotational part in form of the over-terminated rotation matrix  $\mathbf{R}$  is not suitable for the optimization performed in this work. A minimal representation is required. Being aware of the singularities which can occur by using Euler angles, we represent a rigid body motion by twist coordinates  $\xi = (\mathbf{v}, \omega)^T \in \mathbb{R}^6$ . Thereby,  $\mathbf{v} \in \mathbb{R}^3$  describes the translational and the skew-symmetric matrix  $\hat{\omega} \in so(3)$  the rotational part of the full motion. The rotation angle in radians is encoded as  $\|\omega\|_2$ . An element  $\hat{\xi} \in se(3)$  can be written in its homogeneous representation as

$$\hat{\xi} = \begin{pmatrix} \hat{\omega} & \mathbf{v} \\ 0 & 0 \end{pmatrix}. \quad (2)$$

Given  $\hat{\xi} \in se(3)$ , we get a rigid body motion by the matrix exponential, which is defined as the always converging power series:

$$\exp : se(3) \rightarrow SE(3), \quad \exp(\hat{\xi}) = \sum_{k=0}^{\infty} \frac{\hat{\xi}^k}{k!}. \quad (3)$$

This leads to an alternative formulation of Equation (1) using twist coordinates to describe the displacement of a point  $\mathbf{p} \in \mathbb{R}^3$  according to the rigid body motion  $g$  as follows:

$$G : SE(3) \times \mathbb{R}^3 \rightarrow \mathbb{R}^3, \quad G(g, \mathbf{p}) = \exp(\hat{\xi})\mathbf{p}. \quad (4)$$

For more details on the used representation of rigid body motions we refer to (Ma et al., 2003).

Further we define the set of intrinsic camera calibration parameters

$$\mathbf{k} = \{f_x, f_y, o_x, o_y, k_1, k_2\}, \quad (5)$$

whereby  $(f_x, f_y)$  describe the focal length,  $(o_x, o_y)$  the principal point and  $(k_1, k_2)$  the radial distortion of the camera. The projection  $\pi$  performs the mapping from a transformed 3D point  $G(g, \mathbf{p}) = (x, y, z)^T$  to pixel coordinates as

$$\pi(\mathbf{k}, G(g, \mathbf{p})) = \begin{pmatrix} \frac{df_x x}{z} - o_x, \frac{df_y y}{z} - o_y \end{pmatrix}^T, \quad (6)$$

with the radial distortion factor  $d$  being defined by

$$d = 1 + k_1 \left( \frac{x^2 + y^2}{z^2} \right) + k_2 \left( \frac{x^2 + y^2}{z^2} \right)^2. \quad (7)$$

#### 5 PROBLEM FORMULATION

In order to describe the rigid body motion  $g_{c,w}$  of the camera using the measured rigid body motion  $g_{i,w}$  of the INS (Figure 4), the devices have to be rigidly mounted. This induces that the offsets between them are static and especially comprises that the rigid body motions describing the movements from the INS to the camera at various exposure times are equal

$$\forall k, l \in \{1, 2, \dots, n\} : g_{c,l} \stackrel{!}{=} g_{c,l} g_{i,l}, \quad (8)$$

with  $\{1, 2, \dots, n\}$  being the middle exposure times of the images. Therefore we can simplify the notation by omitting the indices for the rigid body motion  $g_{c,l}$  describing the mounting offsets. The composition with the measured INS movement leads to the camera motion

$$g_{c,w} = g_{c,l} g_{i,w}. \quad (9)$$

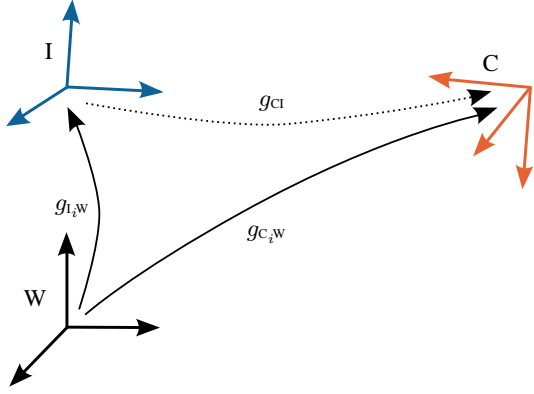


Figure 4: Both, the poses of the INS frame I and of the camera frame C can be described by rigid body motions with regard to the world frame W. The estimation of the static rigid body motion  $g_{CI}$  between the devices (dotted arrow), enables by composition with the INS measurements  $g_{IiW}$  the description of the camera poses  $g_{CiW}$ .

This results in the problem considered in this work, the estimation of the rigid body motion  $g_{CI}$  out of synchronized data from measurement flights. As a first step a SFM approach, which delivers an initial sparse 3D structure of the observed area and the corresponding pixel observations, has to be performed (Hartley and Zisserman, 2004). The refinement of this is usually done in a BA process, where the initial estimation of the 3D points and camera poses are optimized. Given a number of  $n$  images associated with the camera poses  $C_i$  and  $m$  3D points  $\mathbf{p}_j$  with corresponding pixel observations  $\mathbf{x}_{ij}$ . The classical BA approach minimizes the reprojection error of the 3D points according to:

$$\min_{\mathbf{k}, g_{CiW}, \mathbf{p}_j} \sum_i^n \sum_j^m (\pi(\mathbf{k}, G(g_{CiW}, \mathbf{p}_j)) - \mathbf{x}_{ij})^2. \quad (10)$$

The estimations for the intrinsic camera parameters  $\mathbf{k}$  tend to differ slightly from a laboratory calibration, due to the climate and environmental conditions. Thus it is an advantage to optimize  $\mathbf{k}$  by using images from a measurement flight in the classical BA procedure. Nevertheless, errors can be introduced to the intrinsic parameters, which are compensated by the independent camera positions. Therefore the joint calibration of the intrinsic camera parameters in conjunction with the mounting offsets  $g_{CI}$  promises more accurate results. The latter are introduced by adding a term which measures how well the parameters of the camera poses  $g_{CiW}$  in composition with the mounting offsets  $g_{CI}$  satisfy the INS measurements. This is realized by comparing a synthetic measurement generated out of the actual camera pose and mounting offsets with the measured INS pose by using the inverse and the composition of rigid body motions. By adding this constraint in the form of an additional sum to Equation (10), we specify our objective function:

$$\min_{\mathbf{k}, g_{CiW}, \mathbf{p}_j, g_{CI}} \sum_i^n \sum_j^m (\pi(\mathbf{k}, G(g_{CiW}, \mathbf{p}_j)) - \mathbf{x}_{ij})^2 + \sum_i^n (g_{CI}^{-1} g_{CiW})^{-1} g_{IiW}. \quad (11)$$

## 6 GRAPH OPTIMIZATION

The non linear least square problem defined in Equation (11) can be optimized by using  $g^2o$ , the general graph optimization frame-

work (Kuemmerle and Grisetti, 2011). The problem has to be embedded in a graph by introducing the variables to optimize as nodes and the observations between them as edges. In the following we will present the restatement of our objective function into the graph-based formulation.

The calibration parameters represented as the intrinsic camera parameters  $\mathbf{k}$  and the mounting offsets  $g_{CI}$  are added as nodes to the graph. Furthermore, we add each 3D point  $\mathbf{p}_j$  and each camera pose  $g_{CiW}$  as a node to the graph. The connection between these nodes is given by inserting our observations as edges into the graph. A pixel measurement connects three different nodes, namely: a camera, a 3D-point and the intrinsic calibration parameters. This constraint can be realized with a hyperedge, which is able to connect an arbitrary number of nodes. The edge of an INS measurement connects the corresponding rigid body motion of the camera with the mounting offsets. A visualization of the graph is given in Figure 5.

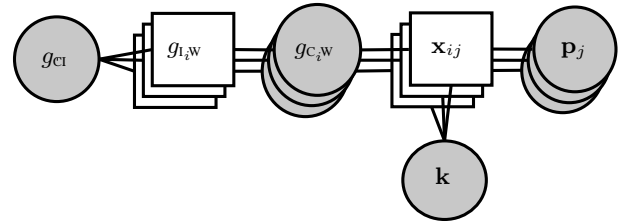


Figure 5: The objective function of the stated problem can be illustrated by a hyper-graph. The measurements (boxes) are presented as links between the nodes concerning each multiple sets of variables (circles). For improved overview multiple state variables and measurements of the same type are visualized in a stacked view unrelated to their number of occurrence.

Further, we have to define error functions, which measure how well our measurements are described by the state variables they are connecting. Our first constraint measures the error occurring by the reprojection of a 3D point into the image, in the same form as in Equation (10). The error function for this constraint can be expressed as:

$$e_{ij}^k(\mathbf{k}, g_{CiW}, \mathbf{p}_j, \mathbf{x}_{ij}) = \pi(\mathbf{k}, G(g_{CiW}, \mathbf{p}_j)) - \mathbf{x}_{ij}. \quad (12)$$

The resulting error vector has dimension two and is  $\mathbf{0}$  if the pixel observation is perfectly described by the state variables. Our second error function states how well the INS measurements can be described by the composition of the camera poses  $g_{CiW}$  and the mounting offsets  $g_{CI}$  as follows:

$$e_i^g(g_{CI}, g_{CiW}, g_{IiW}) = (g_{CI}^{-1} g_{CiW})^{-1} g_{IiW}. \quad (13)$$

Using the twist representation for the rigid body motions, we receive a 6-dimensional error vector, which is  $\mathbf{0}$  if the parameters perfectly satisfy the measurement.

Without limiting the generality, we refer to the whole state vector  $[\mathbf{k}, g_{CiW}, \mathbf{p}_j, g_{CI}]$  as  $\mathbf{y}$  and reformulate our objective function stated in Equation (11) as follows:

$$\min_{\mathbf{y}} \sum_i^n \sum_j^m e_{ij}^k(\mathbf{y})^\top \mathbf{I} e_{ij}^k(\mathbf{y}) + \sum_i^n e_i^g(\mathbf{y})^\top \mathbf{I} e_i^g(\mathbf{y}). \quad (14)$$

The  $g^2o$  framework uses the Levenberg-Marquardt (LM) algorithm to compute a numerical solution of Equation (14) and therefore needs a good initial guess  $\tilde{\mathbf{y}}$  of the state vector. Iteratively, the first order Taylor expansion around the current guess  $\tilde{\mathbf{y}}$  is used to approximate Equation (14) and optimize the local increments



$\Delta y$  by solving the resulting sparse linear system. The center for the next iteration is obtained by adding the optimized increments to the current guess. This is done by using the motion composition for the state variables represented by rigid body motions and a simple addition for the 3D points and intrinsic camera parameters. For a detailed description of the LM algorithm we refer the reader to (Lourakis and Argyros, 2009).

## 7 EVALUATION

In this section, we present results from numerical studies to validate the proposed approach. The main purpose of this is to analyze the influence of various flight configuration and settings on the calibration parameters by comparing the estimations with the known ground truth.

In the simulations we set the mounting parameters consisting of the lever-arm components  $x_L$ ,  $y_L$  and  $z_L$  as well as the angle misalignments in form of the Euler angles yaw =  $\psi_B$ , pitch =  $\theta_B$  and roll =  $\phi_B$  (Figure 2) according to the ground truth defined in Table 1. The initialization with  $\theta_B = 180^\circ$  is a result of the definition of the INS frame as ENU coordinate system, while the downward directed camera is modeled with the z-coordinate pointing in viewing direction. The intrinsic camera parameters stated in Table 1 describe the modeled camera with an image size of 3296 by 2472 pixels using a wide angle lens. The simulated flight courses consist of INS poses sampled along straight lines according to the camera exposure times, determined by the velocity of the UAS and the cameras frames per second. For a more realistic flight path, the ideal poses are modified by a small random factor and the resulting poses are considered as ground truth. One GCP is defined in the middle of the observed area. A fixed number of additional 3D points are sampled in the observed area. The pixel observations are determined by projecting the 3D points into the images according to Equation (6) using both, the generated camera poses and the ground truth of the intrinsic camera parameters defined in Table 1. A realistic number of observations is produced by defining a probability of detection of 50% instead of using all projections of the 3D-points which are in the field of view of a camera. Concerning noise, we simulate the image observations with an accuracy of 0.5 pixels. Further we add white Gaussian noise to the INS poses using a standard deviation of  $\sigma_L = 0.02$  for the position components and  $\sigma_B = 0.01$  for the Euler angles.

The principle of the flight configuration stated as optimal in (Kersting et al., 2011) was used to define our first simulated flight course. At a flying height of 20 m we defined two lines with a

distance of 20 m (Figure 6a). We sampled our INS poses along these lines twice, once in each direction. This results in an image side overlap of 50% for poses on the adjacent lines. The same maneuver was simulated at a flying height of 30 m resulting in a side overlap of about 66%. Two different altitudes in combination with a GCP were used to decouple the high correlation between the lever-arm and the focal length as well as the principal point. The definition of 20 m strip length in conjunction with a speed of 36 km/h and 5 images per second leads to a total number of 80 images with more than 93% image overlap in flying direction. This course was used in Simulation 1, 2 and 3 by sampling 1000, 3000 and 6000 points. As assumed the results in Table 2 show that increasing the point number leads to smaller errors for the calibration parameters. Remarkable is that the angle misalignment is nearly the same for Simulation 2 and 3, which shows that increasing this value further will have little influence on these parameters. The error in the lever arm component  $z_L$  is for all three Simulations about four times bigger than in the other directions, which is also reflected in the estimation of the correlated focal length and principal point (Table 2).

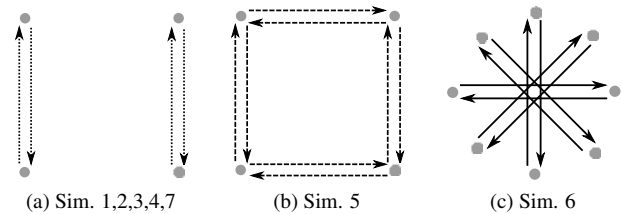


Figure 6: The three flight configurations investigated in this study are visualized in a top view. Each line is sampled in both directions, resulting in an image overlap of approximately 100%.

The first three simulations show that the accuracy of the lever-arm optimization is worse than usual measurements out of constructions drawing or terrestrial measurement. Due to this observation, we fix the lever-arm components to the initialization in the following experiments. For Simulation 4, the same flight and parameter configuration as in Simulation 2 was used. The optimization leads to better results in all estimated parameters (Table 3). The fixed lever-arm also removes the requirement of two flight altitudes. We investigated the influence of the flight course by performing simulations of the patterns visualized in (Figure 6), whereby for each simulation a total number of 3000 points

	Unit	Initialization	Ground truth	Difference
$x_L$	[m]	0.13	0.132	+0.002
$y_L$	[m]	0.1	0.096	-0.004
$z_L$	[m]	0.1	0.104	+0.004
$\psi_B$	[°]	0.0	2.344	+2.344
$\theta_B$	[°]	180.0	183.291	+3.291
$\phi_B$	[°]	0.0	-1.937	-1.937
$f_x$	[pel]	1650.0	1663.31	+13.31
$f_y$	[pel]	1650.0	1662.84	+12.84
$o_x$	[pel]	1648.0	1651.52	+3.52
$o_y$	[pel]	1236.0	1234.67	-1.33
$k_1$	[pel <sup>2</sup> ]	0.0004	0.00076	+0.00036
$k_2$	[pel <sup>4</sup> ]	0.008	0.00908	+0.00108

Table 1: This table states the initialization and ground truth values used in our experiments. For an easier interpretation the difference between the values is also stated.

	RMSE		
	Sim. 1	Sim. 2	Sim. 3
$x_L$	0.00824	0.00849	0.008
$y_L$	0.00728	0.00582	0.00588
$z_L$	0.03169	0.02167	0.01443
$\psi_B$	0.01237	0.01072	0.01052
$\theta_B$	0.0003	0.0009	0.00039
$\phi_B$	0.01925	0.01719	0.01762
$f_x$	2.39323	1.57664	1.0141
$f_y$	2.35976	1.57361	1.01957
$o_x$	0.18362	0.12286	0.07928
$o_y$	0.24893	0.1674	0.11551
$k_1$	$4.1539e^{-5}$	$2.3002e^{-5}$	$1.7486e^{-5}$
$k_2$	$5.1066e^{-5}$	$3.4519e^{-5}$	$2.2469e^{-5}$

Table 2: This table shows the RMSE of the calibration parameters achieved by performing 100 Monte Carlo trials for the first flight course with an increasing number of 3D-points.

was sampled in the observed area. By performing the quadratic course in Simulation 5 and the star pattern in Simulation 6 only at an altitude of 20 m, we get the same number of cameras as in all other simulations. In comparison with Simulation 4, the errors of Simulation 5 are slightly smaller for the intrinsic camera parameters, but higher for the roll angle. The star pattern performed in Simulation 6 leads to a bigger error in the yaw angle (Table 3). Overall the influence of the investigated flight courses on the accuracies of the estimated calibration parameters is small. Another observation was made by performing the flight course depicted in Figure 6a at altitudes of 200 m and 300 m in Simulation 7, which is ten times higher than in Simulation 4. As expected, the errors in the angle misalignment are smaller due to there bigger influence on the object points at higher altitudes (Table 3).

	RMSE			
	Sim. 4	Sim. 5	Sim. 6	Sim. 7
$\psi_B$	0.0111	0.01065	0.02525	0.00357
$\theta_B$	0.00037	0.00123	0.00182	0.00009
$\phi_B$	0.00986	0.01128	0.01028	0.00475
$f_x$	1.10761	0.78936	1.02642	0.15332
$f_y$	1.11969	0.78776	1.02301	0.23633
$o_x$	0.0895	0.05326	0.05658	0.1043
$o_y$	0.1314	0.07603	0.10738	0.13719
$k_1$	$2.2965e^{-5}$	$2.8315e^{-5}$	$3.7536e^{-5}$	$2.5814e^{-5}$
$k_2$	$2.5092e^{-5}$	$1.8952e^{-5}$	$2.3544e^{-5}$	$1.0906e^{-5}$

Table 3: This table shows the RMSE of the calibration parameters achieved by performing 100 Monte Carlo trials for the optimization with an fixed lever-arm for various flight maneuvers.

## 8 CONCLUSION AND FUTURE WORK

In this paper we presented a graph-based approach for the INS-camera calibration. We showed how to integrate the calibration parameters in the bundle adjustment equation and presented a reformulation in a graph structure. The latter was used to estimate both, the mounting offsets between the devices and the intrinsic camera parameters.

The high correlation between the lever-arm of the mounting offsets and the intrinsic camera parameters should be decoupled by performing the flight maneuvers at two different altitudes. Even though, the produced results showed that the accuracy of the lever-arm optimization is worse than usual measurements out of construction drawings or terrestrial measurements. Our simulations confirmed that the best results are achieved if only the intrinsic camera parameters and the misalignment angles between the devices are optimized in the system calibration. Moreover, this eliminates the constraint of using two different altitudes. In conjunction with the small influence of the flight patterns a recalibration during each operation seems possible. Regarding the usage of small UAS, which in general are used at low altitude, we conclude that the calibration of the angle misalignment should be performed as high as possible to achieve the most accurate calibration results.

Future work will investigate the proposed procedure in more detail. Based on the gathered insight, we will perform real flight experiments and evaluate the approach further.

## REFERENCES

Andreff, N., Horaud, R. and Espiau, B., 2001. Robot hand-eye calibration using structure-from-motion. *International Journal of Robotics Research*, 20(3), pp. 228–248.

Cramer, M., Stallmann, D. and Haala, N., 2000. Direct georeferencing using GPS/inertial exterior orientations for photogrammetric applications. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 33 Part B3, pp. 198–205.

Eisenbeiss, H., 2008. The autonomous mini helicopter: a powerful platform for mobile mapping. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37 Part B1, pp. 977–983.

Engel, J., Sturm, J. and Cremers, D., 2012. Camera-based navigation of a low-cost quadcopter. In: *Proc. of the International Conference on Intelligent Robot Systems (IROS)*, pp. 2815–2821.

Hartley, R. and Zisserman, A., 2004. *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Heipke, C., Jacobsen, K. and Wegmann, H., 2002. Analysis of the results of the OEEPE test 'Integrated sensor orientation'. *Integrated sensor orientation - Test report and workshop proceedings*, OEEPE Official Publications No. 43, pp. 31–49.

Horaud, R. and Dornaika, F., 1995. Hand-Eye Calibration. *International Journal of Robotics Research*, 14(3), pp. 195–210.

Jacobsen, K., 2001. Aspects of handling image orientation by direct sensor orientation. In: *Proc. of the ASPRS Annual Convention*.

Kersting, A., Habib, A. and Bang, K., 2011. Mounting Parameters Calibration of GPS/INS-Assisted Photogrammetric Systems. In: *International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping (M2RSM)*, pp. 1–6.

Kuemmerle, R. and Grisetti, G., 2011. g2o: A General Framework for Graph Optimization. In: *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 3607–3613.

Lobo, J. and Dias, J., 2007. Relative Pose Calibration Between Visual and Inertial Sensors. *International Journal of Robotics Research*, 26(6), pp. 561–575.

Lourakis, M. and Argyros, A., 2009. SBA: A software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software* 36(1), pp. 1–30.

Ma, Y., Soatto, S., Kosecka, J. and Sastry, S., 2003. *An Invitation to 3-D Vision - From Images to Geometric Models*. Springer.

Mirzaei, F. and Roumeliotis, S., 2008. A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation. *IEEE Transactions on Robotics*, 24(5), pp. 1143–1156.

Pinto, L. and Forlani, G., 2002. A single step calibration procedure for IMU/GPS in aerial photogrammetry. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34 Part B3, pp. 210–219.

Schikora, M., Bender, D., Koch, W. and Cremers, D., 2010. Multi-target multi-sensor localization and tracking using passive antennas and optical sensors on UAVs. *Proc. SPIE Security + Defence*, 7833, pp. 1–9.

Tsai, R. and Lenz, R., 1989. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. In: *Proc. of the 4th International Symposium on Robotics Research*, pp. 287–297.

Weiss, S. and Achtelik, M., 2012. Versatile distributed pose estimation and sensor self-calibration for an autonomous mav. In: *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 31–38.