Fast and Effective Online Pose Estimation and Mapping for UAVs

Johannes Schneider Christian Eling Lasse Klingbeil Heiner Kuhlmann Wolfgang Förstner Cyrill Stachniss

***Abstract*— Online pose estimation and mapping in unknown environments is essential for most mobile robots. Especially autonomous unmanned aerial vehicles require good pose estimates at comparably high frequencies. In this paper, we propose an effective system for online pose and simultaneous map estimation designed for light-weight UAVs. Our system consists of two components: (1) real-time pose estimation combining RTK-GPS and IMU at 100 Hz and (2) an effective SLAM solution running at 10 Hz using image data from an omnidirectional multi-fisheye-camera system. The SLAM procedure combines spatial resection computed based on the map that is incrementally refined through bundle adjustment and combines the image data with raw GPS observations and IMU data on keyframes. The overall system yields a real-time, georeferenced pose at 100 Hz in GPS-friendly situations. Additionally, we obtain a precise pose and feature map at 10 Hz even in cases where the GPS is not observable or underconstrained. Our system has been implemented and thoroughly tested on a 5 kg copter and yields accurate and reliable pose estimation at high frequencies. We compare the point cloud obtained by our method with a model generated from georeferenced terrestrial laser scanner.**

I. Introduction

Maps are needed for a wide range of applications and most robotic navigation systems rely on maps. Building such maps is often referred to as the SLAM or simultaneous localization and mapping problem and a large number of different techniques to tackle this problem have been proposed in the robotics community. Popular filtering approaches rely on Kalman filters or particle filters and to emphasize their incremental nature, such filtering approaches are usually referred to as online SLAM methods. In contrast to that, most optimization approaches estimate the full trajectory and not only the current pose. They address the full SLAM problem and typically rely on least-squares or related optimization techniques.

SLAM for light-weight UAVs such as the platform shown in Figure 1 is challenging for several reasons: First, the sensors and computers have to be light-weight and are often not comparable to high-quality sensors and powerful computers used on ground robots. Second, UAVs lead to full 6 DoF SLAM and several simplifying assumptions that are reasonable for wheeled robots cannot be made. Third, autonomous UAVs require good pose estimates at high frequencies and in near real-time to allow for a stable control of the platform. In addition to that, one is – at least for surveying applications – interested in building a model that accurately reflects the real geometry of the scene.

All authors are with Institute for Geodesy and Geoinformation, University of Bonn, Germany. This work has partly been supported by the DFG-Project FOR 1505 Mapping on Demand and by the EC under contract number H2020-ICT-644227-FLOURISH.

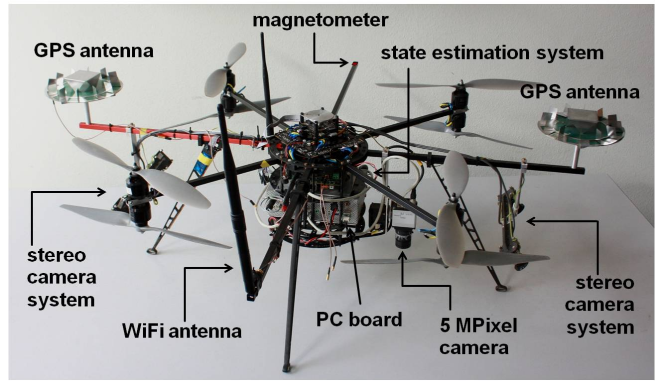

Fig. 1: Our 5 kg UAV platform equipped with several sensors used for effective online pose estimation and mapping on the platform.

The contribution of this paper is a highly integrated system for fast and effective pose estimation and mapping on light-weight UAVs. Our system provides pose estimates at 100 Hz combining real-time kinematic GPS (RTK-GPS) and an inertial measurement unit (IMU). In contrast to most existing systems, our pose estimate is computed using GPS carrier phase ranges and runs fully onboard on a light-weight UAV. This 100 Hz solution can be used to control the copter. In addition to that, we provide an incremental SLAM solution at 10 Hz that is also computed fully on the copter, fusing multiple fisheye stereo cameras, IMU, and raw GPS measurements. The exploitation of GPS carrier phase ranges allows us to even exploit measurements in underconstrained situations, i.e., if only one or two satellites are visible. The estimation is done in a statistically sound manner and provides accurate 6 DoF pose estimates of the platform as well as accurate 3D locations of the feature points. We have implemented and tested our system partially as ROS modules and partially in hardware to meet the real-time constraints. As our evaluation shows, we provide accurate pose estimates at the given frequencies, we can handle underconstrained GPS situations well, and the resulting 3D models are accurate and georeferenced.

II. Related Work

Simultaneous pose estimation and mapping has always been a central research focus in mobile robots, independently of the type of robot. This includes wheeled robots, underwater system, or unmanned aerial vehicles. Thus, a large number of SLAM system have been proposed.

Whenever UAVs operate outdoors, they typically make use of GPS observations for global positioning without a direct need of performing the mapping task simultaneously. Usually, GPS-based state estimation on light-weight UAVs is based on a L1 C/A-code GPS receiver, MEMS inertial sensors and a magnetometer [1], [2], [3]. Such a sensor combination only leads to global position accuracies of approx. 2-10 m and attitude accuracies of approx. 1-5 deg. This is often good enough to autonomously follow waypoints, but it is typically insufficient for UAV control or for geodetic-grade surveying and mapping applications.

The first systems that realized cm-accurate real-time kinematic GPS (RTK-GPS) solutions on UAVs, were presented in [4], [5], [6], [7]. In none of these developments, however, the position and attitude estimation is performed in real-time *onboard* of the UAV platform. Especially for the UAV control and precise autonomous flight, a real-time solution is key to robust operation. Since early 2015, comparably precise commercial state estimation systems for light-weight UAVs can be purchased. Examples are the Ellipse-D system by SBG [8] or the APX-15 UAV by Applanix [9], both launched this year. In order to build a highly integrated SLAM system – as the one proposed in this paper – it is important to have access to all raw measurements as well as the state estimation algorithms. Therefore, we developed an own hardware system with full control over the measurements, algorithms, and internal states. This enables us to effectively couple the RTK-GPS state estimation with our SLAM system and to incorporate camera information, inertial sensor readings, and GPS carrier phase measurements on the level of raw observations in real-time onboard of the UAV. This furthermore enables us to handle underconstrained RTK-GPS situations effectively, yielding a system that provides accurate and reliable state estimates at 100 Hz and 10 Hz, depending on the exploited sensor modalities.

In the context of the simultaneous localization and mapping problem, with or without the use of GPS, graph-based approaches to SLAM are a popular framework since around 20 years. After the work of Lu and Milios [10], several systems have been proposed. For example, Dellaert and Kaess [11] focus on exploiting sparse matrix factorizations. The incremental variant by Kaess et al. [12] exploits partial reorderings to compute the sparse factorization. Others investigate the use of stochastic gradient descent [13], [14] for SLAM in 2D and 3D.

While most of the SLAM back-ends are independent from the sensing modality, several systems have been tailored to visual SLAM. In this context, dense 3D reconstruction approaches have been proposed such as DTAM by Newcombe et al. [15] or the approach by Stühmer et al. [16], which computes a dense reconstruction using variational methods. Optimizing the dense geometry and camera parameters is possible but a rather computationally intensive task, see [17]. To tackle the computational complexity for real-time operation, semi-dense approaches have been proposed, for example by Engel et al. [18].

Due to the low weight of cameras, visual reconstruction techniques received considerable attention for light-weight UAVs. Pizzoli et al. [19] propose a reconstruction approach for UAVs that combines Bayesian estimation and convex optimization. They execute the reconstruction on a GPU at frame rate. Also combinations of cameras on an indoor UAV and RGB-D cameras on a ground vehicle have been used for simultaneous localization and mapping tasks aligning the camera information with dense ground models [20]. Harmat et al. [21] adapted the parallel tracking and mapping (PTAM) algorithm by Klein and Murray [22] to handle omnidirectional multi-camera systems to estimate the pose of a small UAV. Onboard methods for autonomous navigation of an UAV exploiting Kalman filter to process stereo camera and IMU input are presented in [23] and [24], in the latter additionally with laser input. Ellum [25] investigate the accuracy and reliability of tight coupling of raw GPS code pseudo-ranges into a offline bundle adjustment.

Our system described in this paper combines several techniques. Parts of the individual components have been previously published outside the robotics community and are only described briefly here. The incremental visual pose estimation using multiple fisheye stereo cameras is subject of [26], [27], the fast pose estimation at using RTK-GPS and IMU is subject of [28], [29], [30], [31]. The key novelties here are the combination of the pose estimation at 100 Hz using RTK-GPS and IMU as well as the online estimation at 10 Hz fusing all sensor modalities on the same light-weight UAV platform. Furthermore, our 10 Hz SLAM solution can exploit information from underconstrained RTK-GPS situations, i.e, less then four available satellites, due to the integration of raw carrier phase ranges and it provides accurate estimates comparing the resulting models to those of terrestrial laser scans.

III. Fast and Effective Online State Estimation for Light-Weight UAVs

The first subsection describes the high-speed pose estimation using a Kalman filter and RTK-GPS and IMU to provide a 100 Hz pose estimate. The second subsection describes our 10 times slower approach that uses image data from a multi-fisheye-camera system for spatial resection for pose determination and combines the visual information with the RTK-GPS and IMU information in a graph-based SLAM/bundle adjustment manner.

A. Pose Estimation at 100 Hz using RTK-GPS and IMU

The GPS provides *two types of observations* for the positioning of mobile objects. The first type of observations are the *code ranges* (e.g. C/A code) and the second ones are *carrier phase ranges* on multiple frequencies (L1, L2, and L5). Geodetic-grade survey receivers measure the phase of these signals with an accuracy of 1-2% of their wavelength. Since the carrier phase wavelengths are approx. 20 cm and the C/A-code wavelength is approx. 300 m, the former leads to substantially higher accuracies and is therefore key to realize a cm-accurate positioning with GPS.

GPS positioning of mobile objects based on carrier phase ranges in real-time is called RTK-GPS and it is a relative positioning procedure. This means, that the unknown coordinates of a movable station (the UAV) are determined with respect to a stationary master station. The advantage of this relative positioning is an improved accuracy that comes from single- and double-differencing of the observations, see [32]. By using single-differences, which are calculated from a signal of one satellite measured at both receivers (UAV and the master), the satellite clock bias as well as the atmosphere refractions can be reduced significantly. Double-differences are calculated from the single-differences of two satellites and therefore eliminate the receiver clock bias and other receiver dependent effects.

The main difficulty in RTK-GPS processing is, that the receiver is only able to measure the fractional part of a carrier wave cycle, while the remaining integer number of cycles is inherently unknown. Resolving this number is called ambiguity resolution and it is the key to RTK-GPS positioning. Especially in kinematic applications, it has to be very fast and robust at the same time, since the satellite signals may be interrupted for short times quite frequently, and this always requires a re-initialization of the ambiguities. The method most commonly used to resolve the ambiguities is the least-squares ambiguity decorrelation adjustment (LAMBDA) [33]. To allow for a fast and robust ambiguity resolution, we use the inertial sensor readings and the magnetic field observations to aid the ambiguity float solution, which is the process of estimating the ambiguities as real values. This is done within the GPS/IMU integration using the Kalman filter described below.

Our GPS/IMU integration is based on a strapdown algorithm [34]. We determine the high dynamic movement of the system by integrating the angular rates and the accelerations of the MEMS IMU with a rate of 100 Hz. Within a Kalman filter update (10 Hz) the GPS and the magnetic field observations are used to bound the inertial sensor drift.

In the filter, the navigation equations of the body-frame, denoted with b, are expressed in an earth-fixed frame, denoted with e. The full state vector $\mathbf{x}$ includes the position $\mathbf{x}^e$ and the velocity $\mathbf{v}^e$, represented in the e-frame. For the attitude determination, we use a quaternion $\mathbf{q}^e$. The accelerometer bias $\mathbf{b}_a^b$ and the gyro bias $\mathbf{b}_\omega^b$ are represented in the b-frame and are estimated as well. Further the real valued ambiguity parameters $\mathbf{N}_{R,M}^{jk}$ are estimated so the state space becomes:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{e\top} & \mathbf{v}^{e\top} & \mathbf{q}^{e\top} & \mathbf{b}_a^{b\top} & \mathbf{b}_\omega^{b\top} & \mathbf{N}_{R,M}^{jk\top} \end{bmatrix}^\top . \quad (1)$$

The observations in the measurement model are:

1) DD carrier phase observations on the L1 and the L2 frequency, measured in the e-frame, with a rate of 10 Hz,
2) DD code ranges on the L1 and the L2 frequency, measured in the e-frame, with a rate of 10 Hz,
3) magnetic field observations, measured in the b-frame, with a rate of 10 Hz,
4) onboard GPS attitude baseline vector, expressed in the e-frame, with a rate of 1 Hz, see [29].

The estimated real valued ambiguities now have to be fixed to integer numbers. This is done by applying the modified LAMBDA method [35], which leads to faster computation times than the original LAMBDA approach. The ambiguity resolution can take a few measurement epochs, but usually the ambiguities can be fixed instantaneously. Due to the GPS/IMU integration, cycle slips in the carrier phases can be detected and repaired reliably, see [36] for further details.

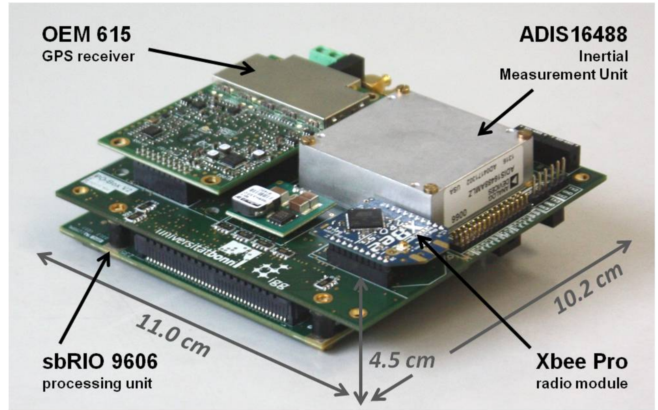

Fig. 2: The 100 Hz RTK-GPS/IMU state estimation board.

As soon as the ambiguities have been resolved for all satellites, the double- differenced carrier phase observations including the resolved ambiguities are provided to the 10 Hz visual SLAM process explained in the subsequent section.

To address the need for a flexible, small and lightweight solution to process the navigation solution in real-time on the UAV, we developed a custom sensor system based on a 400 MHz processor and an FPGA, see Figure 2. Our setup includes a geodetic-grade GPS receiver (Novatel OEM 615), a low-cost single-frequency GPS Chip (Ublox LEA6T), an IMU (Analog devices ADIS16488) and a magnetometer (Honeywell HMC5883L). The dual-frequency GPS receiver is used for the RTK-GPS positioning as described above. At the same time, the antennas of both GPS receivers form a short baseline on the UAV platform, see Figure 1, which improves the heading determination, since the magnetometer is usually distorted by any hard and soft iron effects or the currents of the rotors. The usage of a real-time operating system on the sensor board in combination with the FPGA enables a high-frequency and time deterministic acquisition of all sensor readings. Furthermore, all computations including Kalman filter and ambiguity resolution is computed on this board so that no external computations are needed and the pose estimate can directly be fed into the flight-controller of the UAV. The total weight of the UNIT is 240 g and the dimensions are 11 cm $\times$ 10.2 m $\times$ 4.5 cm. More details can be found in [30].

B. Online Pose Estimation and Mapping at 10 Hz Fusing All Sensor Modalities

Compared to the previous section, we describe here a ten times slower estimation procedure without real-time guarantees. In contrast to the 100 Hz solution, it allows us to incorporate all sensing modalities and to compute a feature map of the environment.

Our current setup utilizes four fisheye cameras that are triggered simultaneously at 10 Hz. We refer to images taken at the same time of exposure as a *frame set*. The pose determination of each frame set relies on image feature points with known association to scene points in an incrementally refined and extended map. The estimation and refinement of the map is performed in an parallel running bundle adjustment on selected keyframes that also integrates the double differences from GPS as well as IMU data. All processing is done on an onboard PC. The overall process consists of the following steps:

1) The data acquisition and association detects feature points and performs the matching to provide corresponding image points to the previous frame set and the other cameras.
2) The orientation of each frame set with fast resection provides a fast pose estimate and allows to select keyframes.
3) An incremental bundle adjustment merges the new information at a keyframe with the previous information in a statistically optimally way.

We aim at efficient methods for reliable data association, for fast pose determination, and to target on an outlier-free information for the bundle adjustment step. Our optimization also considers the GPS and IMU information. This optimization step is the most costly one as it uses all available data on the selected keyframes. To avoid long computation times, the optimization is performed with the incremental optimization by Kaess et al. [12]. The remainder of this section describes the three steps in detail.

Our approach requires calibrated cameras. We calibrate each fisheye camera in advance according to [37]. For calibration, we model the fisheye objective with the equidistant-model described in [38] allowing for ray directions with an angular distance larger than 90° to the viewing direction. The mutual orientation of the fisheye cameras in the multi-camera system is determined in advance as described in our previous work [39]. Further, we observe GPS control points in the images to derive the offset of the camera-system to the phase center of the GPS antenna.

1) Visual Data Acquisition and Association: Our visual pose estimation and mapping procedure exploits point features extracted from the images. To allow for handling four cameras onboard the copter, an efficient feature extractor is essential. To this end, we select KLT features that are tracked in the individual cameras. We detect interest points that are corners in the gradient image with a large smallest eigenvalue of the structure tensor, cf. [40] and track them with the iterative Lucas-Kanade method with pyramids according to [41]. Figure 3 shows an example of tracked interest points in the four fisheye images of a frame set.

Having calibrated cameras each tracked feature point can be converted into a ray direction $\mathbf{x}'$ that points in the individual camera system to the observed scene point. The uncertainty of the image coordinates can be transformed to the uncertainty of $\mathbf{x}'$ via variance propagation yielding $\Sigma_{\mathbf{x}'\mathbf{x}'}$. In all cases, the covariance matrix of the camera rays is

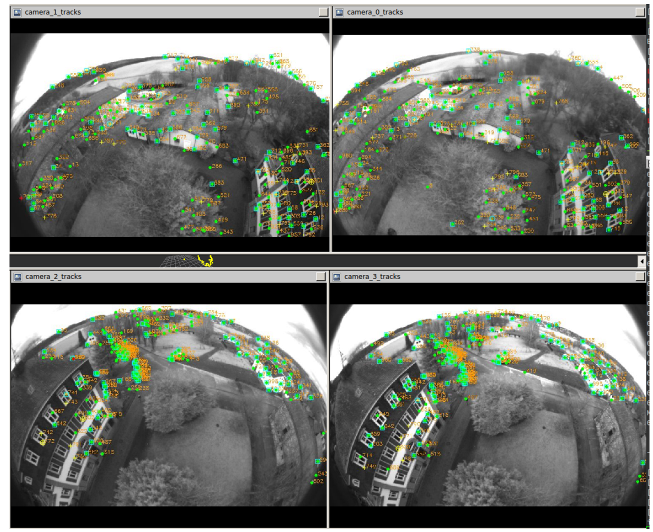

Fig. 3: Synchronized triggered frame set of the four fisheye cameras. Each image contains around 200 feature points that are tracked using a KLT tracker.

singular, as the normalized 3-vector only depends on two observed image coordinates. To match interest points in the overlapping images of each stereo camera pair, we use correlation coefficients and exploit epipolar geometry to reduce candidates on the corresponding epipolar lines within their propagated error bounds.

We use the camera rays with its covariance information for the spatial resection at frame rate (see Sec. III-B.2) and for incremental bundle adjustment on keyframes (see Sec. III-B.3).

2) Fast Pose Estimation: In our approach, we use feature maps, which are defined as a set of scene points $\mathcal{X} = \{\mathcal{X}_i\}$. In theory, the location of these scene points and the pose of the camera system can be estimated through bundle adjustment (BA) directly. Given the computational demands, it is impossible to compute a BA solution at 10 Hz on the copter. Therefore, we execute the BA only on selected keyframes at around 1 Hz. To compute the camera poses and between the keyframes, we compute the UAV poses by spatial resection on each frame set.

The location of the points are initialized at the first acquired frame set by forward intersecting the matched ray directions in the stereo pairs. The frame set that initializes $\mathcal{X}$ is chosen as first keyframe $\mathcal{K}_1$.

After initialization of the map, the motion M_t of the camera system in relation to the map is computed at frame rate using resection. For resection we use scene points $\mathbf{X}_i$ that are observed in cameras $c = 1, ..., 4$ at time t and exploit the known system calibration M_c to consider the multiple projection centers. M_c describes the known transformation of each single camera c to the reference frame of the UAV and M_t describes the unknown transformation of the UAV reference frame into the reference coordinate system of the map at time t, thus M_t contains the pose parameters of the

UAV. An estimated pose M_t induces the residual

$$\boldsymbol{v}_{itc} = \mathrm{null}(\mathbf{x}'^{\mathsf{T}}_{itc})^{\mathsf{T}} \mathsf{N}\left(\mathsf{P}\,\mathsf{M}_c^{-1}\mathsf{M}_t^{-1}\mathbf{X}_i\right) \tag{2}$$

of an observed ray direction $\mathbf{x}'_{itc}$ pointing to homogeneous scene point $\mathbf{X}_i$ in camera c at time t. The homogeneous scene point $\mathbf{X}_i$ is transformed with the inverse motion matrices M_t and M_c into a single camera and is projected with $\mathsf{P} = [I_3|\mathbf{0}_3]$ into the predicted direction, which is spherically normalized to unit length with $\mathsf{N}(\mathbf{x}) = \mathbf{x}/|\mathbf{x}|$. To reduce the number of error equations per observed direction from three to two, making the two degrees of freedom of the observed direction explicit, we project the predicted direction into the tangent space of the observed ray direction $\mathbf{x}'_{itc}$, i.e. the null space of $\mathbf{x}'^{\mathsf{T}}_{itc}$, see [42]. Note that motion matrices M_c are determined in advance with a rigorous bundle adjustment following [39].

We optimize the six pose parameters of M_t with an iterative maximum likelihood-type estimation with the robust Huber cost function [43] that down weights observations with large residuals. The estimation of the pose parameters for M_t converges in 2-3 fast iterations using M_{t-1} as initial value. This allows a robust pose estimation at a high frame rate.

To obtain a near outlier-free input for BA, we exploit the estimated weight in the Huber cost function. Observations with low weights are considered as outliers and are not used in BA and excluded from tracking.

3) Keyframe-Based Incremental Bundle Adjustment: The last step in the SLAM pipeline is keyframe-based visual SLAM or bundle adjustment. This optimization step considers the camera images but also incorporates the GPS DD observations as well as the IMU measurements. For the optimization, we rely on iSAM2 [12], which models the problem as a factor graph. Each node on the factor graph corresponds to a keyframe pose ($\mathcal{M}_t$) or a 3D scene point ($\mathcal{X}_i$). The nodes are connected through factors that result from the different observations,

We add a new keyframe $\mathcal{K}_t$ in case a certain geometric distance to the last keyframe $\mathcal{K}_{t-1}$ is exceeded. Each new keyframe contains two kinds of observations, χ_1 and χ_2, where χ_1 are the observations of scene points that are already in the map and χ_2 denotes those observing new scene points. With each new keyframe the map is expanded by forward intersetion with observations χ_2. Note that only χ_1 has been revised from outliers in the robust pose estimation described previously. In order to identify outliers in χ_2 based on their residuals we require a track to consist of at least three keyframes for mapping.

The map $\mathcal{X}$ and keyframe poses in $\mathcal{K}$ are simultaneously refined using bundle adjustment with integrated GPS and IMU information.

a) Integration of camera rays: For bundle adjustment we use the measurement equations in Eq. (2) that is not linear in the scene point and pose parameters of $\mathbf{X}_i$ and M_t. The linearization is shown in detail in [42]. In terms of factor-graphs each observed camera ray $\mathbf{x}'_{itc}$ produces a factor $f(\mathcal{M}_t, \mathcal{X}_i)$.

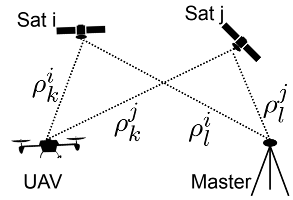

Fig. 4: Double differences are determined using the distances ρ between the known positions of the GPS satellites and the master and the approximate UAV position.

b) Integration of DD Carrier Phase Observations: To integrate the GPS double differences, the keyframe poses need to be in the GPS coordinate system. Therefore, we initially require to have a unique GPS solution, for which at least three double differences are needed. When initializing the BA, we first determine the positions of the first five keyframes with GPS coordinates and do not integrate double differences into the BA. From the 5th keyframe with a GPS position on, we estimate a similarity transformation and transform all keyframe poses and the map into the GPS system.

Then, we integrate the DD carrier phase observations by adding a factor $f(\mathcal{M}_t)$ for the L1 and L2 frequency to the factor graph. For the measurement equation, the coordinates of the GPS satellites and the master receiver is needed, see [32].

Figure 4 depicts a double differences measurement. The measurement equation for the phase carrier double differences is

$$\mathrm{DD} + v_{\mathrm{DD}} = (\rho_l^j - \rho_k^j) - (\rho_l^i - \rho_k^i) \tag{3}$$

where ρ are distances between receivers and satellites as in Figure 4 and v_{DD} the residuals. The double differences between all satellites and a reference satellite form the joint double differences measurement and thus a factor for the optimization.

c) Integration of IMU: The measured IMU rotation angles are integrated over the time between two following keyframes leading to the observed angles $\boldsymbol{\omega}$ with the rotation matrix $R(\boldsymbol{\omega})$. The measurement equation reads as $R(\boldsymbol{\omega} + \boldsymbol{v}_{\boldsymbol{\omega}}) = R_{t-1}^{\mathsf{T}} R_t$ with residuals $\boldsymbol{v}_{\boldsymbol{\omega}}$ and is integrated with factor $f(\mathcal{M}_{t-1}, \mathcal{M}_t)$.

d) Incremental Optimization using iSAM2: For our online application, the processing of a new keyframe $\mathcal{K}_t$ needs to be finished by the time the next keyframe is added. Each new information is incrementally merged with the previous information, yielding a fast statistically optimal solution for the bundle adjustment using iSAM2.

Linearization of the non-linear model at the initial values as linearization points leads to the least squares optimization problem

$$\widehat{\Delta \boldsymbol{x}} = \underset{\Delta \boldsymbol{x}}{\mathrm{argmin}} \|A \Delta \boldsymbol{x} - \boldsymbol{b}\|_{\Sigma}^2 \tag{4}$$

with the Jacobian A that contains block entries for each observation, right hand side (RHS) vector $\boldsymbol{b}$, unknown updates $\Delta \boldsymbol{x}$ for the scene point and pose parameters and for

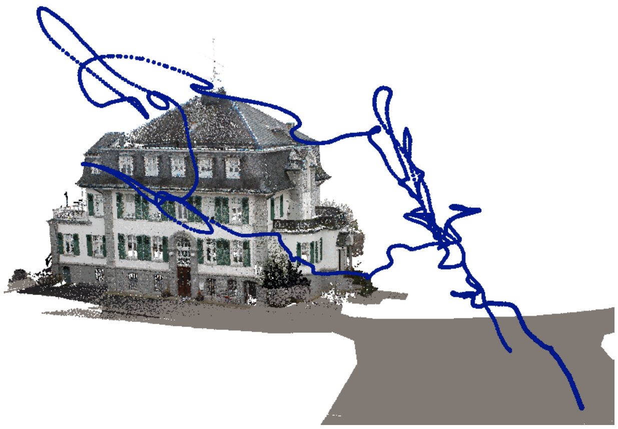

Fig. 5: Trajectory of the UAV flight overlayed with a georeferenced 3D model of a near-by building.

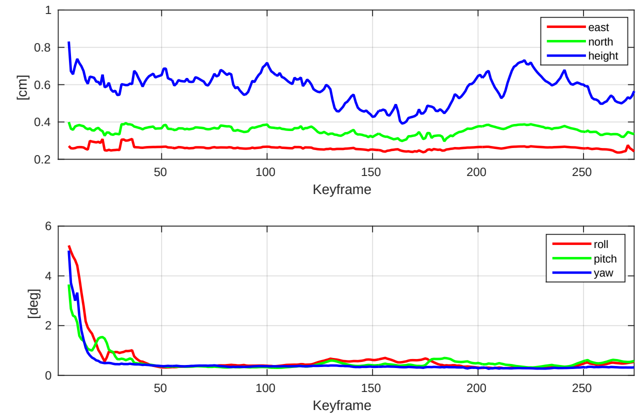

Fig. 6: Theoretical accuracy of position and orientation angles of copter at keyframes. The high long-time precision in the position is provided by the RTK-GPS, the high precision of the rotations is due to the high relative orientation accuracy obtained with bundle adjustment by considering the visual information.

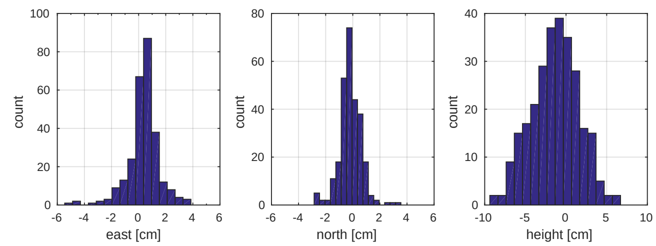

Fig. 7: Residuals between incrementally estimated positions of keyframes for the GPS double differences.

proper weighting the covariance matrix Σ that contains the covariance matrices of all observations.

As new measurements often have only a local effect and fill-in may become expensive, iSAM2 encodes the conditional density of cliques in a so-called Bayes tree, which allows for an efficient incremental reordering, just-in-time relinearization and partial solving, when parameters change only locally. For more details, see [12].

IV. Experiments

Our experimental evaluation is designed to illustrate the performance of our approach and to support the main claims made in this paper. These key claims are that our approach (i) offers accurate pose estimation for light-weight UAVs at high frequencies, (ii) exploits incomplete GPS observations with less than 4 satellites, (iii) provides highly accurate and georeferenced pose and map estimation.

For evaluation, we recorded all sensor data with our UAV under good GPS conditions, with 5 to 8 visible satellites. This allows us to manually eliminate GPS observations and evaluate the effect on the overall state estimation procedure. The flight used for this evaluation was guiding the UAV along the facade of a house, the variation in position is around 60 m and 15 m in height, see Figure 5. An exemplary frame set out of the image sequence is shown in Figure 3. The dataset contains 3,368 frame sets recorded at 10 Hz and in each image, around 200 features are tracked. The SLAM system initiates a keyframe each 1 m, resulting in 274 keyframes and online SLAM starts at take off on the ground and ends at the landing.

The first experiment is designed to show the obtained accuracy of the estimated keyframe poses during the whole flight. To obtain the theoretical accuracy, we extract the covariance information, when the bundle adjustment is incrementally solved at a new keyframe. This is too time consuming for online processing but can be done as an offline evaluation. Figure 6 shows the theoretical accuracy of the position and orientation of the copter at the estimated keyframes. The rotation angles get more accurate, if enough GPS observations constrain the rotation estimation. The highest rotational precision of 0.5° is preserved from the 50th keyframe until landing. The uncertainty is confirmed empirically with the estimated variance factor being in the order of one, assuming a image point accuracy of 2 pixel. The image point accuracy of 2 pixel may be seen as somewhat low, but results from the frame-wise KLT tracking.

In addition to that, Figure 7 shows the residuals of the GPS DD measurements after optimization, which are within the uncertainties of RTK-GPS solution. As expected, one can also see a larger uncertainty in the height estimate than in the two other directions.

We also evaluate the differences between the 100 Hz Kalman filter solution under GPS-friendly conditions and the bundle adjustment solution. The visual information improves the pose estimate and the differences in each axis direction of the UTM coordinate system between both estimates is on average 1.1 cm in east and north direction and 3.2 cm for the height.

The second experiment is designed to demonstrate the potential of our approach to handle underconstrained GPS situations, i.e., situations in which less than 4 satellites are available. Standard GPS receives report a GPS loss and cannot estimate a solution. Through the combination of visual SLAM and DD measurements, we can however compute a solution and exploit individual double differences. As can be seen from the trajectories shown in Figure 8, exploiting two DD measurements (3 satellites) improves the trajectory

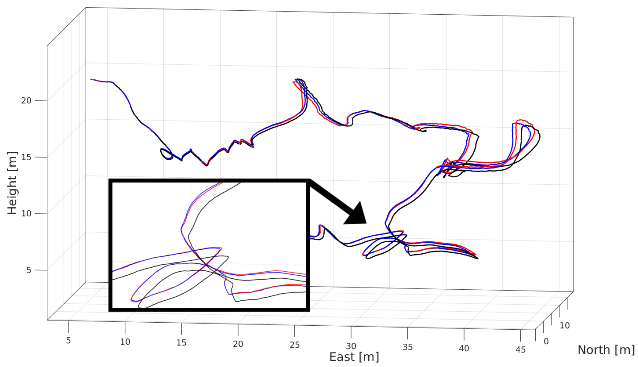

Fig. 8: The three estimated trajectories show the benefit of operating on raw GPS DD measurements. The red line represents the trajectory exploiting full GPS information (5-8 satellites) and is considered as the reference trajectory. Assuming that only 3 satellites are available, the combination of only 2 GPS DD measurements with visual information leads to the solution shown in blue that is much closer to the red reference trajectory than the GPS-free solution shown in black, which could not exploit underconstrained GPS measurements.

estimates substantially and thus is a valuable information for UAV operating in GPS-unfriendly environments.

The last experiment shows the highly accurate mapping that is possible using our system. To evaluate the quality of our map estimates, we mapped the house along which the copter flew with a terrestrial laser scanner and precisely georeferenced the recording location so that the point cloud can also be compared to a georeferenced near ground truth 3D model.

As the map build using our copter is also georeferenced, both models can be compared without any further alignment. We compare our reconstruction with the georeferenced terrestrial laser scan to evaluate the quality of the determined poses, see Figure 9. The median of the absolute differences to the nearest neighbors in each axis direction is around 1 cm. A robust MAD estimation in the component-wise deviations results in about 3 cm and 50% of all points that have a distance smaller than 5 cm to the nearest neighbor. The full distribution is given in the histogram in Figure 9. This experiment shows that our approach generated accurate georeferenced 3D point clouds online.

All computations are performed onboard the copter, which is equipped with a standard 3.6 GHz Intel CPU with 4 cores and the Kalman filter runs on the real-time board. All steps except the optimization is done at 10 Hz for all four cameras and the BA optimization runs once per second and is completed between the next keyframe is created and the next optimization would be triggered.

In sum, our experimental evaluation shows that the proposed system offers accurate pose estimation for light-weight UAVs at 100 Hz and 10 Hz. Our visual SLAM system can furthermore exploit underconstrained RTK-GPS observations with less than 4 satellites, which reduces the drift in comparison to SLAM systems with traditional GPS integration. Through the effective fusion of GPS, IMU, and visual

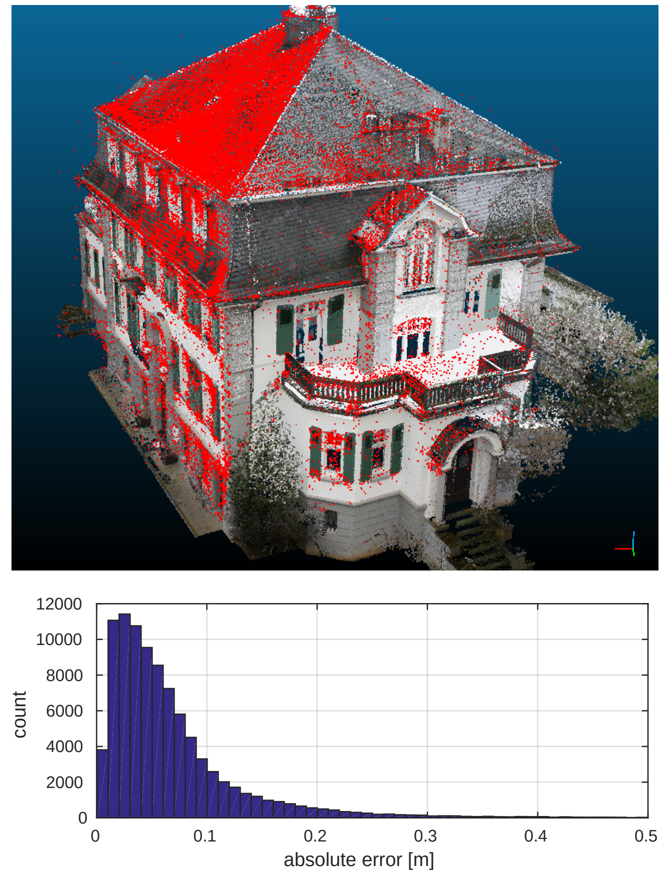

Fig. 9: Top: Point cloud from reconstruction using high resolution images (red) and point cloud from terrestrial laser scan (textured). Bottom: Histogram of the distances between the individual points from the SLAM system and the terrestrial laser scan.

information, we can compensate GPS- unfriendly situations. Finally, we compared our 3D point cloud to a georeferenced near ground truth 3D model providing an objective measure of the quality of the computed point cloud. In this way, the experimental evaluation supports our claims.

V. Conclusion

In this paper, we presented an effective system for online pose and simultaneous map estimation designed for light-weight UAVs. Our system combines real-time GPS and IMU integration at 100 Hz including GPS ambiguity resolution and provides a real-time kinematic solution. In addition to that, we presented an effective bundle adjustment solution exploiting RTK-GPS carrier phase observations, IMU data and visual SLAM in an incremental fashion at 10 Hz. The overall system yields a robust pose estimate at high frequencies and can handle underconstrained GPS situations effectively. The components running at 100 Hz have been implemented on a light-weight, real-time hardware board and the other components are software ROS modules and everything runs online on a 5 kg multi-copter. By comparing our results with models generated from georeferenced terrestrial laser scanners, we show a deviation of the median to our point clouds of less than 1 cm.

REFERENCES

[1] C.-S. Yoo and I.-K. Ahn, "Low cost GPS/INS sensor fusion system for uav navigation," in *22nd Digital Avionics Systems Conference*, 2003, pp. 8.A.1–1–8.A.1–9.

[2] D. Kingston and R. Beard, "Real-time attitude and position estimation for small UAVs using low-cost sensors," in *AIAA 3rd "Unmanned Unlimited" Technical Conference, Workshop and Exhibit*, Chicago, IL, 20-23 September 2004.

[3] J. Wendel, O. Meister, C. Schlaile, and G. F. Trommer, "An integrated GPS/MEMS-IMU navigation system for an autonomous helicopter," *Aerospace Science and Technology*, vol. 10, no. 6, pp. 527–533, 2006.

[4] M. Rieke, T. Foerster, J. Geipel, and T. Prinz, "High-precision positioning and real-time data processing of uav-systems," in *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Science, Vol. XXXVIII-1/C22, UAV-g 2011*, 2011.

[5] W. Stempfhuber and M. Buchholz, "A precise, low-cost RTK GNSS system for UAV applications," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science XXXVIII(1-C22)*, 2011, pp. 289–293.

[6] M. Bäumker, H.-J. Przybilla, and A. Zurhorst, "Enhancements in UAV flight control and sensor orientation," in *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Science XL-1/W2, UAV-g 2013*, 2013, pp. 33–38.

[7] M. Rehak, R. Mabillard, and J. Skaloud, "A micro aerial vehicle with precise position and attitude sensors," *PFG - Photogrammetrie, Fernerkundung, Geoinformation*, vol. 4, pp. 239–251, 2014.

[8] SBG, *Ellipse-d: Miniature dual GPS INS*, http://www.sbg-systems.com/products/ellipse-d-dual-antenna-gps-inertial-system, 15.01.2016, 2015.

[9] Applanix, *APX-15 UAV*, http://www.applanix.com/products/airborne/apx15uav.html, 15.01.2016, 2015.

[10] F. Lu and E. Milios, "Globally consistent range scan alignment for environment mapping," *Autonomous Robots*, vol. 4, 1997.

[11] F. Dellaert and M. Kaess, "Square Root SAM: Simultaneous localization and mapping via square root information smoothing," *Int. Journal of Robotics Research*, vol. 25, no. 12, pp. 1181–1204, 2006.

[12] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, "iSAM2: Incremental Smoothing and Mapping Using the Bayes Tree," *Intl. Journal of Robotics Research (IJRR)*, vol. 31, no. 2, pp. 217–236, 2012.

[13] E. Olson, J. Leonard, and S. Teller, "Fast iterative optimization of pose graphs with poor initial estimates," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2006, pp. 2262–2269.

[14] G. Grisetti, C. Stachniss, and W. Burgard, "Non-linear constraint network optimization for efficient map learning," *IEEE Transactions on Intelligent Transportation Systems*, 2009.

[15] R. Newcombe, S. Lovegrove, and A. Davison, "Dtam: Dense tracking and mapping in real-time," in *Proc. of the IEEE Intl. Conf. on Computer Vision (ICCV)*, 2011.

[16] J. Stühmer, S. Gumhold, and D. Cremers, "Real-time dense geometry from a handheld camera," in *Proc. of the Symposion of the German Association for Pattern Recognition (DAGM)*, 2010, pp. 11–20.

[17] M. Aubry, K. Kolev, B. Goldluecke, and D. Cremers, "Decoupling photometry and geometry in dense variational camera calibration," in *Proc. of the IEEE Intl. Conf. on Computer Vision (ICCV)*, 2011.

[18] J. Engel, J. Sturm, and D. Cremers, "Semi-dense visual odometry for a monocular camera," in *Proc. of the IEEE Intl. Conf. on Computer Vision (ICCV)*, Sydney, Australia, 2013, pp. 1449–1456.

[19] M. Pizzoli, C. Forster, and D. Scaramuzza, "Remode: Probabilistic, monocular dense reconstruction in real time," in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2014.

[20] C. Forster, M. Pizzoli, and D. Scaramuzza, "Air-ground localization and map augmentation using monocular dense reconstruction," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2013.

[21] A. Harmat, M. Trentini, and I. Sharf, "Multi-camera tracking and mapping for unmanned aerial vehicles in unstructured environments," *Journal of Intelligent and Robotic Systems (JIRS)*, vol. 78, no. 2, pp. 291–317, 2015.

[22] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'07)*, 2007.

[23] P. Rudolf, M. Wzorek, and P. Dotherty, "Vision-based pose estimation for autonomous indoor navigation of micro-scale unmanned aircraft systems," in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2010, pp. 1913–1920.

[24] T. Tomic, K. Schmid, P. Lutz, A. Dömel, M. Kassecker, E. Mair, I. Grixa, F. Ruess, M. Suppa, and D. Burschka, "Toward a fully autonomous uav: Research platform for indoor and outdoor urban search and rescue," *IEEE Robotics and Automation Magazine (RAM)*, vol. 19, no. 3, pp. 46–56, 2012.

[25] C. Ellum, "Integration of raw gps measurements into a bundle adjustment," in *ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXV-B4, Istanbul, Turkey, Jul 2004, pp. 933–938.

[26] J. Schneider and W. Förstner, "Real-time accurate geo-localization of a mav with omnidirectional visual odometry and gps," in *ECCV Workshop: Computer Vision in Vehicle Technology*, ser. L.N. in Computer Science, vol. 8925, 2014, pp. 271–282.

[27] J. Schneider, T. Läbe, and W. Förstner, "Incremental real-time bundle adjustment for multi-camera systems with points at infinity," in *ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-1/W2, 2013, pp. 355–360.

[28] L. Klingbeil, M. Nieuwenhuisen, J. Schneider, C. Eling, D. Droeschel, D. Holz, T. Läbe, W. Förstner, S. Behnke, and H. Kuhlmann, "Towards autonomous navigation of an UAV-based mobile mapping system," in *4th International Conference on Machine Control & Guidance*, Braunschweig, Germany, March 19-20 2014, pp. 136–148.

[29] C. Eling, M. Wieland, C. Hess, L. Klingbeil, and H. Kuhlmann, "Development and evaluation of a UAV based mapping system for remote sensing and surveying applications," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science, Volume XL-1/W4*, 2015, pp. 233–239.

[30] C. Eling, L. Klingbeil, M. Wieland, and H. Kuhlmann, "Direct georeferencing of micro aerial vehicles - system design, system calibration and first evaluation tests," *PFG - Photogrammetrie, Fernerkundung, Geoinformation*, vol. 4, pp. 227–237, 2014.

[31] C. Eling, L. Klingbeil, and H. Kuhlmann, "Real-time single-frequency gps/mems-imu attitude determination of lightweight uavs," *IEEE Sensors Journal*, vol. 15, no. 10, pp. 26 212–26 235, 2015.

[32] B. Hofmann-Wellenhof, H. Lichtenegger, and E. Wasle, *GNSS Global Navigation Satellite Systems*. Springer Wien New York, 2008.

[33] P. Teunissen, "The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation," *Journal of Geodesy*, vol. 70, pp. 65–82, 1995.

[34] D. Titterton and J. Weston, *Strapdown Inertial Navigation Technology, 2nd Edition*, ser. 17, N. Stewart and H. Griffiths, Eds. IEE Radar, Sonar and Navigation, 2004.

[35] X.-W. Chang, X. Yang, and T. Zhou, "MLAMBDA: A modified LAMBDA method for integer least squares estimation," *Journal of Geodesy*, vol. 79, pp. 552–565, 2005.

[36] C. Eling, E. Heinz, L. Klingbeil, and H. Kuhlmann, "Cycle slip detection in the context of RTK GPS positioning of lightweight UAVs," in *4th International Conference on Machine Control & Guidance*, Braunschweig, Germany, 19-20 March 2014, pp. 148–155.

[37] S. Abraham and T. Hau, "Towards autonomous high-precision calibration of digital cameras," in *Proc. of SPIE Videometrics*, vol. 3174, San Diego, USA, Jul 1997, pp. 82–93.

[38] S. Abraham and W. Förstner, "Fish-eye-stereo calibration and epipolar rectification," *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)*, vol. 59, no. 5, pp. 278–288, 2005.

[39] J. Schneider and W. Förstner, "Bundle adjustment and system calibration with points at infinity for omnidirectional camera systems," *Photogrammetrie – Fernerkundung – Geoinformation (PFG)*, vol. 4, pp. 309–321, Aug 2013.

[40] J. Shi and C. Tomasi, "Good features to track," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, Jun 1994, pp. 593–600.

[41] J. Bouguet, "Pyramidal implementation of the lucas kanade feature tracker," Intel Corporation, Microprocessor Research Labs, Tech. Rep., 2000.

[42] J. Schneider, F. Schindler, T. Läbe, and W. Förstner, "Bundle adjustment for multi-camera systems with points at infinity," in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. I-3, Aug 2012, pp. 75–80.

[43] P. Huber, *Robust Statistics*. New York: John Wiley, 1981.