Comparison of Inertial Mechanization Approaches for Inertial Aided Monocular EKF-SLAM

Markus Kleinert
Fraunhofer IOSB
Ettlingen, Germany
Email: markus.kleinert@iosb.fraunhofer.de

Christian Ascher
Karlsruhe Institute of Technology
Karlsruhe, Germany
Email: christian.ascher@kit.edu

Uwe Stilla
Technische Universität München
München, Germany
Email: stilla@bv.tum.de

***Abstract*—Localization of pedestrians becomes a difficult task in situations where no measurements with respect to an established reference system, as it is provided by satellites when using GPS, are available. One possible approach to tackle this problem is to attach a suitable sensor to pedestrians and then to run a simultaneous localization and mapping (SLAM) algorithm in order to localize the sensor. A combination of an inertial measurement unit (IMU) with a monocular camera is a promising choice of sensors for an indoor pedestrian localization system since these sensors provide complementary measurements.**

This paper discusses two approaches to integrate visual and inertial information which differ mainly in the choice of reference coordinate system. A detailed description of both approaches is given and they are compared with respect to their performance on simulated and real measurement data.

I. Introduction

The capability to localize oneself in a previously unknown environment is a prerequisite to navigate and to perform tasks within that environment. Therefore, especially blind people or first responders in disaster recovery situations would benefit from a pedestrian localization system that operates in a variety of scenarios. Satellite positioning systems like GPS are widespread, but they depend on the availability of the radio signal sent by the satellites. This precludes their usage inside buildings or in urban canyons, where the signals are often either occluded or severely disturbed due to multipath effects. Alternative positioning systems, like Honeywell's GLANSER system [1], which were developed to fill this gap, often rely on some sort of infrastructure that has to be installed at the site of operation.

Positioning systems that do not make use of external infrastructure essentially perform dead reckoning, i.e., the position can only be estimated relative to a previous position estimate. Hence, the error in the calculated solution inevitably grows over time. Inertial measurement units (IMUs) measure acceleration and angular velocity of a moving body. Thus, integration of these measurements over time yields an estimate of the body's displacement during the integration interval. However, IMUs that offer the accuracy necessary to calculate feasible estimates of a pedestrian's position are both costly and bulky and therefore not applicable to the task of localizing a pedestrian. By contrast, low-cost IMUs built with micro-electro-mechanical systems (MEMS) technologies satisfy the requirements concerning costs and size, but are subject to measurement errors that lead to rapidly growing errors in the position estimate.

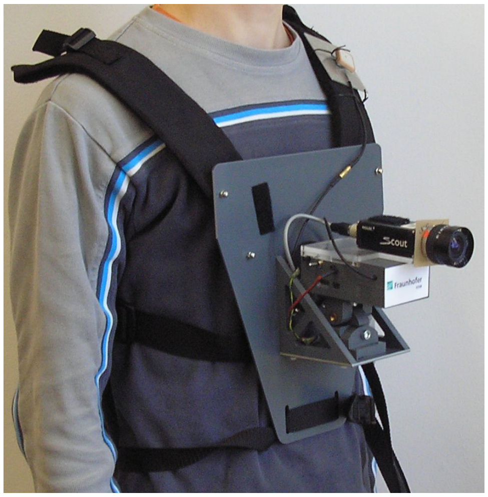


Fig. 1. Sensor system comprising an IMU and a camera attached to the torso of a person.

One possibility to overcome these problems is to use a combination of a compact low-cost IMU with a sensor that observes landmarks in the environment in order to reduce drift. Since the positions of these landmarks are not known in advance, they have to be estimated simultaneously with the pose of the sensor. This problem is commonly referred to as the simultaneous localization and mapping (SLAM) problem. This work is concerned with the integration of a camera with a low-cost IMU into a SLAM system that can be comfortably attached to a pedestrian as shown in Fig. 1. For this purpose, the extended Kalman filter (EKF) is used to estimate the positions of unknown landmarks and the pose of the sensor system simultaneously.

II. Related Work

A. Visual SLAM

Because of its importance for practical applications, not only in robotics, the SLAM problem has received a lot of attention during recent years. When monocular or stereo cameras are used as the primary sensor the procedure is usually as follows: Image coordinates that correspond to the projection of landmarks in the image plane are tracked over multiple camera images and subsequently used to estimate the sensor's trajectory as well as the map.

It is interesting to note that the same problem is addressed in different scientific communities each time for a slightly different purpose. Namely, while the focus in the robotics community is on localization of autonomous robots, cf. [2], the computer vision community often aims at efficient determination of the camera's pose for augmented reality applications [3]. Only recently the batch optimization approach to SLAM which was long known to the photogrammetric community under the term "bundle adjustment" was rediscovered and modified for real-time operation [4].

B. Inertial aiding: World-centric formulation

However, filtering using either an EKF or an unscented Kalman Filter (UKF) is still a widely adopted method for integrating visual and inertial measurements in a SLAM system. This probably stems from the fact that filtering, especially with the error-state formulation of the EKF, is the standard tool for GPS-INS integration in the navigation community. The inertial mechanization equations, i.e., the strapdown algorithm, used for error-state filtering are usually formulated w.r.t. a coordinate system that is aligned with the local direction of gravity, cf. [5]. This frame is called *navigation frame* in this work. Veth and Raquet present an inertial-aided visual SLAM system that utilizes the error state formulation w.r.t. the navigation frame by extending the state vector of an EKF with the Cartesian coordinates of observed landmarks [6]. A stereo camera system or a terrain model are used to obtain an initial estimate of the distance of new landmarks to the camera image they were first observed in.

This formulation of the SLAM problem, where the sensor's pose and the map are parameterized w.r.t. the navigation frame is called *world-centric* in this work.

C. Inertial aiding: Sensor-centric formulation

A major drawback of the EKF is its susceptibility to linearization errors. Huang presents an extensive study on the effect of linearization errors in SLAM for the simplified case of a robot moving in the 2D-plane [7]. It is shown that linearization errors have the effect that theoretically non-observable directions of the SLAM state space that correspond to the sensor's pose become observable and lead to inconsistent state estimates. It is also shown that the correct observability properties can be maintained during filtering if the *sensor-centric* parameterization that was introduced by Castellanos et al. in [8] is chosen. Here, the origin of the reference coordinate system always coincides with the sensor coordinate system. This has the effect that linearization errors are significantly reduced because the derivatives w.r.t. the sensor pose are calculated close to the true linearization points. However, after each measurement update a composition step has to be performed in order to reset the origin to the updated sensor pose. Since the composition step affects the whole covariance matrix it may significantly increase the computational cost for large maps.

In [9] a similar sensor-centric landmark parameterization is presented that enables a linear measurement update for landmark observations by predicting the position of landmarks in the camera frame using inertial measurements and the unscented transformation.

In the following, the term sensor-centric will also be used to refer to parameterizations as they are used in [10] and [11]. Both formulate the inertial mechanization equations w.r.t. the sensor pose at the start of operation. For this purpose the coordinates of the gravity vector w.r.t. the first sensor pose are also estimated in order to be able to perform the strapdown computation.

D. Contribution

This work compares two variants of the strapdown algorithm, which is used to process IMU measurements during the time prediction step in error state formulations of the EKF, regarding their capability to produce consistent state estimates when employed in a visual SLAM framework in combination with an EKF. Namely, a sensor-centric formulation that allows to process landmark observations in a coordinate system that coincides with the sensor coordinate system at the beginning of the estimation process is compared to a world-centric approach where the reference frame is determined by the direction of gravity. The sensor-centric approach is similar to explicitly estimating the direction of gravity in the reference coordinate system as it also allows to start estimation with zero uncertainty in the estimate of the sensor's pose.

Furthermore, the effect of performing a composition step as originally proposed in [8] on the consistency of the filter and the overall scale estimate is investigated. Experiments with real and simulated data are presented to compare the different approaches.

III. EKF-SLAM FORMULATION

The subsequent sections present the inertial aided EKF-SLAM approach used in this work in more detail with an emphasis on the inertial mechanization equations (strapdown computation).

A. Coordinate systems

The following coordinate systems are of particular interest for the formulation of the inertial mechanization equations in secs. III-E-III-F.

- The *body or IMU-coordinate system* $\{b\}$ is aligned to the IMU's axes and therefore describes the pose of the whole sensor system. Its position and orientation are included in the filter state.
- The *camera coordinate system* $\{c\}$ is not part of the filter state. Its pose can be calculated from the IMU's pose by means of the camera-IMU transformation that only depends on the mechanical setup and is assumed to be fix and known in this work.

- The *navigation coordinate system* $\{n\}$ is the fixed world frame whose x- and y- axis point north- and eastwards while its z-axis points in the direction of local gravity. It is assumed that the distance between the body coordinate system and the navigation frame is small compared to the radius of the earth and therefore the direction of gravity can be considered constant during operation of the system. In this case the position of the navigation frame can be chosen arbitrarily.
- The *strapdown coordinate system* $\{s\}$ is the frame the inertial mechanization equations are formulated in. Also, the pose of the body frame is given in coordinates w.r.t. the strapdown frame. In the world-centric formulation this frame coincides with the navigation frame. When a sensor-centric formulation is chosen, the mechanization equations are formulated either w.r.t. the body frame at the beginning of the measurement process or w.r.t. the body frame at the time when the last composition step was performed.

Here, the term "pose" of a coordinate system refers to its position and orientation w.r.t. a given reference system. Lowercase letters are used to refer to coordinate systems in coordinate transformations. E.g., ${}^{n}\mathbf{p}_b$ refers to the position of the body frame $\{b\}$ in coordinates of the navigation frame $\{n\}$. Similarly, C_b^n denotes the direction cosine matrix (DCM) that transforms coordinate vectors in the body frame to coordinate vectors in the navigation frame. As a pair, ${}^{n}\mathbf{p}_b$ and C_b^n describe the pose of the body frame w.r.t. the navigation frame. $C(\mathbf{q})$ denotes the rotation matrix that is associated with a unit quaternion $\mathbf{q}$.

B. SLAM state parameterization

Since the goal is to determine the pose of the body frame and a sparse map of point landmarks, the EKF state vector comprises parameters which describe the IMU's motion and biases as well as the coordinates of observed landmarks:

$$\mathbf{s}_t \quad = \quad \left[\mathbf{s}'^{\mathrm{T}}\ \mathbf{m}^{\mathrm{T}}\right]^{\mathrm{T}} \tag{1}$$

Here, $\mathbf{s}'$ contains the state variables that describe the motion of the body frame w.r.t. the reference frame that was chosen for the formulation of the inertial mechanization equations. It is described in more detail in sec. III-E and III-F. The map vector $\mathbf{m}$ subsumes the coordinates of all landmarks that are included in the filter's state:

$$\mathbf{m} \quad = \quad \left[\mathbf{Y}_1^{\mathrm{T}}\ \ldots\ \mathbf{Y}_N^{\mathrm{T}}\right]^{\mathrm{T}} \tag{2}$$

In the following, estimated values are denoted by a hat $(\hat{\cdot})$ and a tilde $(\tilde{\cdot})$ is used to indicate the error, i.e., the deviation between a true value $(\cdot)$ and its estimate: $(\tilde{\cdot}) = (\cdot) - (\hat{\cdot})$.

C. Error state formulation

Since the EKF relies on a truncation of the Taylor series expansion of the measurement equation as well as the time update step after the first derivative, it can be regarded as an estimator for the state error $\tilde{\mathbf{s}}$. This is the basis for the error state formulation of the EKF which is commonly used for GPS-INS integration, cf. [5, pp. 199-222]. Therefore, the covariance matrix associated with the filter state describes the distribution of $\tilde{\mathbf{s}}$ under the assumption that the errors follow a normal distribution. It is given by

$$P \quad = \quad \left[\begin{array}{cc} P_{\tilde{\mathbf{s}}',\tilde{\mathbf{s}}'} & P_{\tilde{\mathbf{s}}',\tilde{\mathbf{m}}} \\ P_{\tilde{\mathbf{m}},\tilde{\mathbf{s}}'} & P_{\tilde{\mathbf{m}},\tilde{\mathbf{m}}} \end{array}\right] . \tag{3}$$

The error of the estimated orientation can be written in terms of the incremental orientation that aligns the estimated coordinate system with the unknown true coordinate system:

$$\mathbf{q}_d^c \quad = \quad \mathbf{q}(\mathbf{\Psi}_d^c) * \hat{\mathbf{q}}_d^c \quad , \quad \mathbf{q}(\mathbf{\Psi}) \quad \approx \quad \left[1\ \frac{\mathbf{\Psi}}{2}^{\mathrm{T}}\right]^{\mathrm{T}} \tag{4}$$

Where $*$ denotes quaternion multiplication and $\{c\}$, $\{d\}$ are arbitrary coordinate frames.

D. IMU measurement model

The IMU measures acceleration and angular velocity w.r.t. an inertial frame in its own reference frame $\{b\}$. A reference frame that is fixed to the earth's surface cannot be an inertial frame because gravitational and Coriolis forces, which result from the earth's mass and rotation, affect any body that rests in such a coordinate frame. While the effect of gravity cannot be neglected, it is assumed that the Coriolis effect cannot be distinguished from noise with the low-cost inertial sensor used in this work. Similarly, it is assumed that the effect of the earth's rotation cannot be observed in the angular rate measurements.

In addition, inertial measurements are usually subject to systematic errors (biases), which have to be compensated before integrating inertial measurements.

Thus, the IMU's measurements are modeled by the following equations:

$$\begin{aligned} {}^{b}_{m}\mathbf{a} \quad &= \quad {}^{b}\mathbf{a} + \mathbf{b}_a + \mathbf{n}_a & (5) \\ {}^{b}_{m}\omega \quad &= \quad {}^{b}\omega + \mathbf{b}_g + \mathbf{n}_g & (6) \end{aligned}$$

In the above equations, $\mathbf{b}_a$ and $\mathbf{b}_g$ are the biases, which affect the acceleration and angular rate measurements respectively. Furthermore, $\mathbf{n}_a$ and $\mathbf{n}_g$ are white Gaussian noise terms pertaining to acceleration and angular rate measurements ${}^{b}_{m}\mathbf{a}$ and ${}^{b}_{m}\omega$. It is assumed, that the biases change in time according to a random walk process driven by white Gaussian noise terms $\mathbf{n}_{b_a}$ and $\mathbf{n}_{b_g}$.

E. Inertial mechanization: World-centric formulation

The inertial mechanization equations serve two purposes: They describe how the sensor system moves according to the inertial measurements and how the covariance matrix that describes the IMU's pose ($\mathbf{s}'$) uncertainty is propagated in time.

In the world-centric formulation these equations are written w.r.t. the navigation frame, which hence becomes the strapdown frame in this case. The IMU's motion can thus be described with the following state variables:

$$\mathbf{s}' \quad = \quad \left[{}^n\mathbf{p}_b^{\mathrm{T}}\,{}^n\mathbf{v}_b^{\mathrm{T}}\,\mathbf{b}_a^{\mathrm{T}}\,\mathbf{q}_b^{n\,\mathrm{T}}\,\mathbf{b}_g^{\mathrm{T}}\right]^{\mathrm{T}} \tag{7}$$

Inertial measurements are integrated to update position and velocity estimates according to the following equations:

$$\begin{aligned} {}^n\hat{\mathbf{a}} &= C(\hat{\mathbf{q}}_b^n)\cdot({}^b_m\mathbf{a}-\hat{\mathbf{b}}_a)+{}^n\mathbf{g} \\ {}^n\hat{\mathbf{p}}_{b,t+\tau} &= {}^n\hat{\mathbf{p}}_b+{}^n\hat{\mathbf{v}}_b\cdot\tau+\tfrac{1}{2}{}^n\hat{\mathbf{a}}\cdot\tau^2 \\ {}^n\hat{\mathbf{v}}_{b,t+\tau} &= {}^n\hat{\mathbf{v}}_b+{}^n\hat{\mathbf{a}}\cdot\tau \end{aligned} \tag{8}$$

Where τ is the time interval between consecutive inertial measurements. Note, that the gravity vector ${}^n\mathbf{g}$ does not need special treatment since the equations are given w.r.t. the navigation frame where the direction of gravity is known. To integrate angular rate measurements, a quaternion that describes the incremental rotation in the body frame is formed from the angular rate measurements and subsequently used to update the orientation estimate:

$$\begin{aligned} \hat{\omega} &= {}^b_m\omega-\hat{\mathbf{b}}_g \\ \hat{\mathbf{q}}_{b,t+\tau}^n &= \hat{\mathbf{q}}_b^n * \mathbf{q}(\hat{\omega}) \end{aligned} \tag{9}$$

In the error space formulation the Rodrigues vector takes the place of the quaternion in (7):

$$\tilde{\mathbf{s}}' \quad = \quad \left[{}^n\tilde{\mathbf{p}}_b^{\mathrm{T}}\,{}^n\tilde{\mathbf{v}}_b^{\mathrm{T}}\,\tilde{\mathbf{b}}_a^{\mathrm{T}}\,\mathbf{\Psi}_b^{n\,\mathrm{T}}\,\tilde{\mathbf{b}}_g^{\mathrm{T}}\right]^{\mathrm{T}} \tag{10}$$

The uncertainty of the error state is propagated according to a first order differential equation that corresponds to the physical model in (8) and (9):

$$\dot{\tilde{\mathbf{s}}}' = F\cdot\tilde{\mathbf{s}}' + G\cdot\mathbf{n} \tag{11}$$

Here, vector $\mathbf{n}$ summarizes the noise terms. The entries of the matrices F and G are determined by the coefficients of the time derivatives of the error state:

$${}^n\dot{\tilde{\mathbf{p}}}_b={}^n\tilde{\mathbf{v}}_b \tag{12}$$

$$\begin{aligned} {}^n\dot{\tilde{\mathbf{v}}}_b=&\left\lfloor -C(\hat{\mathbf{q}}_b^n)\cdot({}^b_m\mathbf{a}-\hat{\mathbf{b}}_a)\right\rfloor_{\times}\cdot\mathbf{\Psi}_b^n+ \\ &C(\hat{\mathbf{q}}_b^n)\cdot\mathbf{n}_a - C(\hat{\mathbf{q}}_b^n)\cdot\tilde{\mathbf{b}}_a \end{aligned} \tag{13}$$

$$\dot{\tilde{\mathbf{b}}}_a=\mathbf{n}_{b_a} \tag{14}$$

$$\dot{\mathbf{\Psi}}_b^n=C(\hat{\mathbf{q}}_b^n)\cdot\tilde{\mathbf{b}}_g \tag{15}$$

$$\dot{\tilde{\mathbf{b}}}_g=\mathbf{n}_{b_g} \tag{16}$$

Where $\lfloor\mathbf{v}\rfloor_{\times}$ is the skew-symmetric cross product matrix with entries from $\mathbf{v}$.

It is important to realize, that these derivatives depend on estimated values. In case of the velocity and the orientation error this leads to coefficients in F that depend on the estimated state. As noted in [12] this is an important source of linearization related errors in inertial aided SLAM systems.

Given the time derivatives of the error state, the covariance is propagated as follows for each inertial measurement:

$$\Phi = \exp(F\cdot\tau) \approx I_{15\times15} + F\cdot\tau \tag{17}$$

$$P'_{t+\tau} = \Phi\cdot P_{\tilde{\mathbf{s}}',\tilde{\mathbf{s}}'}\cdot\Phi_t^{\mathrm{T}} + \Phi\cdot G\cdot Q\cdot G^{\mathrm{T}}\cdot\Phi^{\mathrm{T}}\tau \tag{18}$$

$$P_{t+\tau} = \left[\begin{array}{cc} P'_{t+\tau} & \Phi\cdot P_{\tilde{\mathbf{s}}',\mathbf{m}} \\ P_{\tilde{\mathbf{m}},\tilde{\mathbf{s}}'}\cdot\Phi^{\mathrm{T}} & P_{\tilde{\mathbf{m}},\tilde{\mathbf{m}}} \end{array}\right] \tag{19}$$

In the expression above, Q is the power spectral density matrix which characterizes the noise vector $\mathbf{n}$.

F. Inertial mechanization: Sensor-centric formulation

The idea of using a sensor-centric formulation is to reduce the dependency on estimated values in the covariance propagation process by formulating the mechanization equations with respect to a strapdown frame that is known with high certainty, e.g., a recent sensor coordinate system [12]. Since the direction of gravity is not known in this frame, the transformation between the strapdown frame and the navigation frame is included in the state vector:

$$\mathbf{s}' \quad = \quad \left[{}^s\mathbf{p}_b^{\mathrm{T}}\,{}^s\mathbf{v}_b^{\mathrm{T}}\,\mathbf{b}_a^{\mathrm{T}}\,\mathbf{q}_b^{s\,\mathrm{T}}\,\mathbf{b}_g^{\mathrm{T}}\,{}^n\mathbf{p}_s^{\mathrm{T}}\,\mathbf{q}_s^{n\,\mathrm{T}}\right]^{\mathrm{T}} \tag{20}$$

Here, ${}^n\mathbf{p}_s$ and $\mathbf{q}_s^n$ are the position and orientation quaternion of the strapdown frame w.r.t. the navigation frame. With the corrected acceleration

$${}^s\hat{\mathbf{a}} \quad = \quad C(\hat{\mathbf{q}}_b^s)\cdot({}^b_m\mathbf{a}-\hat{\mathbf{b}}_a)+C(\hat{\mathbf{q}}_n^s)\cdot{}^n\mathbf{g} \tag{21}$$

inertial measurements can be integrated in the $\{s\}$-frame as shown in (8)-(9) for the navigation frame.

The time derivatives for the error state in this case are:

$${}^s\dot{\tilde{\mathbf{p}}}_b={}^s\tilde{\mathbf{v}}_b \tag{22}$$

$$\begin{aligned} {}^s\dot{\tilde{\mathbf{v}}}_b=&\left\lfloor -C(\hat{\mathbf{q}}_b^s)\cdot({}^b_m\mathbf{a}_b-\hat{\mathbf{b}}_a)\right\rfloor_{\times}\cdot\mathbf{\Psi}_b^s + C(\hat{\mathbf{q}}_b^s)\cdot\mathbf{n}_a- \\ &C(\hat{\mathbf{q}}_b^s)\cdot\tilde{\mathbf{b}}_a + C(\hat{\mathbf{q}}_s^n)^{\mathrm{T}}\cdot\lfloor{}^n\mathbf{g}\rfloor_{\times}\cdot\mathbf{\Psi}_s^n \end{aligned} \tag{23}$$

$$\dot{\tilde{\mathbf{b}}}_a=\mathbf{n}_{b_a} \tag{24}$$

$$\dot{\mathbf{\Psi}}_{\mathbf{b}}^{\mathbf{s}}=C(\hat{\mathbf{q}}_b^s)\cdot\tilde{\mathbf{b}}_g \tag{25}$$

$$\dot{\tilde{\mathbf{b}}}_g=\mathbf{n}_{b_g} \tag{26}$$

$${}^n\dot{\tilde{\mathbf{p}}}_s=0_{3\times3} \tag{27}$$

$$\dot{\mathbf{\Psi}}_s^n=0_{3\times3} \tag{28}$$

With these derivatives a covariance propagation step is performed as described by (17)-(19).

Note, that the inertial integration equations together with the covariance updates define the time update step of the EKF for both mechanizations.

G. Measurement update

Salient image regions are continuously extracted and tracked in the image stream. The coordinates of all features extracted in one image are stacked together to form the measurement vector that is subsequently used to update the state. The observation model for a landmark $\mathbf{Y}_i$ consists of a coordinate transformation and subsequent projection onto the image plane:

$$\begin{aligned}\mathbf{z} &= \mathbf{h}(\mathbf{s}) + \mathbf{v} \\ &= \pi(C_s^c \cdot (\mathbf{Y}_i - {}^s\mathbf{p}_c)) + \mathbf{v} \end{aligned} \tag{29}$$

Here, $\mathbf{z}$ are the observed image coordinates, $\mathbf{v}$ is the zero mean white Gaussian measurement noise, and $\pi(\cdot)$ is the projection function. The Jacobian of the observation model w.r.t. the state variables is:

$$H_i \quad = \quad J_\pi \left[J_\mathbf{p} \; 0_{3\times 6} \; J_\Psi \; \ldots \; 0_{3\times 3\cdot(i-1)} \; J_\mathbf{Y} \; 0_{3\times 3\cdot(N-i)} \right] . \tag{30}$$

Where $J_\mathbf{p}$, J_Ψ and $J_\mathbf{Y}$ are the derivatives of the coordinate transformation w.r.t. the position of the body frame, its orientation, and the position of the landmark, respectively. Furthermore, J_π is the Jacobian of the projection function.

Similarly to the measurement vector, the Jacobian H for the EKF update step is obtained by stacking the Jacobians H_i for individual landmarks. Finally, an EKF update step is performed to estimate the error $\tilde{\mathbf{s}}$, cf. [13]. The error estimate is then used to correct the state where quaternions are corrected as described by (4).

The observation model does not depend on the chosen mechanization. However, when a mechanization w.r.t. an arbitrary frame is used, the additional state variables have to be updated as well.

H. Composition step

The composition step, which is introduced in [8], is used in this work in order to reset the strapdown frame, which is also the reference frame for the SLAM algorithm, at regular intervals. It is basically a coordinate transformation that is applied to the whole state vector and the covariance matrix with the effect that the entries in the covariance matrix that correspond to the sensor's pose w.r.t. the strapdown frame are zeroed out while the uncertainty that is associated with the remaining state entries is adjusted accordingly. The coordinate transformation that resets the strapdown frame can be written as follows:

$$T_s^b \quad = \quad \begin{bmatrix} C_b^{s\,\mathrm{T}} & -C_b^{s\,\mathrm{T}} \cdot {}^s\mathbf{p}_b \\ 0 & 1 \end{bmatrix} \tag{31}$$

This transformation is applied to all state variables except for the biases, which do not depend on a chosen reference frame, and the pose of the strapdown frame w.r.t. the navigation frame. The latter is adjusted by the composition

$$T_{s_2}^n \quad = \quad T_s^n \cdot T_b^s . \tag{32}$$

Here, $\{s_2\}$ is the new strapdown frame after the composition step, which coincides with the body frame. Let J_c be the Jacobian of the function that results from applying the above coordinate transformations to the whole state. The new covariance matrix for the whole state is then obtained by first order error propagation:

$$P_c \quad = \quad J_c \cdot P \cdot J_c^\mathrm{T} \tag{33}$$

In the experiments presented in the next section a composition step is only performed whenever more than four new features are included in the filter state simultaneously. This strategy was chosen to speed up calculations.

IV. Experimental results

A. Simulation results

For the simulation experiments, a L-shaped trajectory was generated that resembles a walk along a hallway with a sharp turn into a narrow sideway at the end. Acceleration and angular rate measurements were derived from the generated trajectory. These were artificially corrupted with white Gaussian noise whose standard deviation was determined from the noise statistics that were measured by the sensor system used for the real data experiment presented in the next section. Landmark observations were generated by projecting the coordinates of known points on the image plane using a central projection model and adding white Gaussian noise with one pixel standard deviation. Again, the chosen camera parameters resemble the values of the system used in the real data experiment.

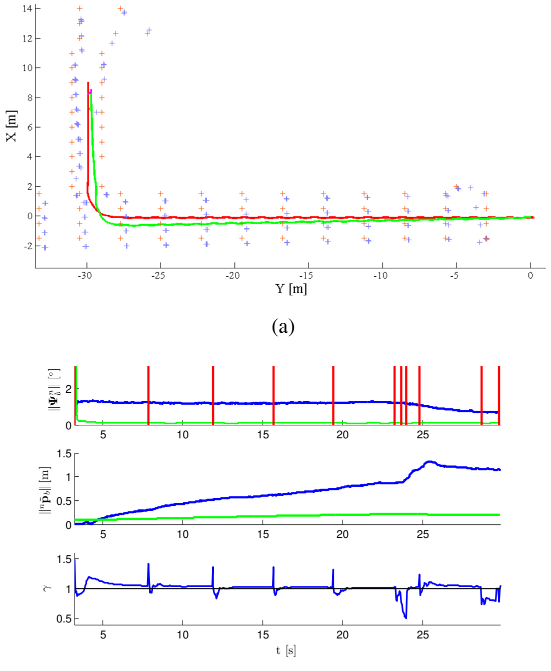
In order to investigate in how far inertial measurements facilitate the estimation of scale, the ratio of average distances between estimated landmark positions to the average distances between ground truth landmark positions is calculated during the simulation runs:

$$\gamma = \frac{\sum_{i=1}^{N} \sum_{j=i}^{N} \|\hat{\mathbf{Y}}_i - \hat{\mathbf{Y}}_j\|}{\sum_{i=1}^{N} \sum_{j=i}^{N} \|\mathbf{Y}_i - \mathbf{Y}_j\|} \tag{34}$$

Results for one simulation run are shown in Figs. 2-4. When a world-centric mechanization is used, the filter becomes inconsistent after a short period of time. Since the initial uncertainty of the sensor's pose describes the expected registration error at the beginning of the measurement process, this precludes the correction of these initial errors when the estimates of the IMU's acceleration biases improve. Consequently, the initial pitch angle error is not corrected, and an substantial error in the height estimate occurs. The world-centric approach also yields the worst scale estimates for the main part of the scene compared to the two alternatives.

The plots for the sensor-centric mechanizations also include the transformation from the $\{s\}$-frame to the $\{n\}$-frame. Since the initial position ${}^n\mathbf{p}_s$ is not correlated to the other state variables, it is not adapted in the filtering process and hence not included in Fig. 3. However, correlations arise when the composition step is performed. Thus, ${}^n\mathbf{p}_s$ is included in Fig. 4.

The results indicate that a sensor-centric mechanization improves the consistency of the algorithm. Nevertheless, the filter becomes inconsistent approximately after 23 s. This is the time when the system enters the narrow sideway, and a lot of new landmarks are introduced to the filter state. Because new landmarks are initialized with a fix depth of 8 m and a



(a)

(b)

Fig. 2. Simulation results for the world-centric mechanization. (a) Top view of the simulated trajectory. Red: Reference trajectory, Green: Estimated trajectory, Blue: Reference landmark positions Orange: Estimated landmark positions. (b) Error plots. Blue: Error, Green: 3σ-bounds as estimated by the filter. Red bars indicate the points in time when new landmarks are introduced.

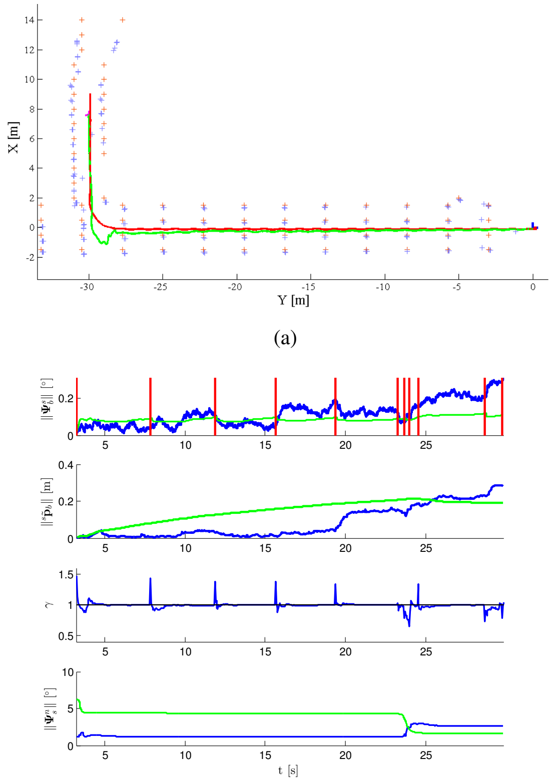
large uncertainty in the direction of the projection ray, this causes the peaks that can be observed in the plots of the scale ratio γ. This probably also explains the observed inconsistency after 23 s: the fix depth used for initialization deviates even more from the true depth of observed landmarks in the narrow sideway.

The composition step improves the consistency of the filter w.r.t. the local strapdown frame. Note, that error and uncertainty of the estimated sensor pose in the $\{s\}$ frame are reduced to zero after each composition step while the uncertainty associated with the pose of the $\{s\}$ frame w.r.t. the $\{n\}$ frame increases until the acceleration biases become observable due to the turn after 23 s. However, overall the estimates for the pose of the $\{s\}$ frame w.r.t. the $\{n\}$ frame are still inconsistent at the end of the trajectory.

B. Indoor experiment

The three approaches were also tested on an indoor dataset that was recorded in an office building using the system shown in Fig. 1. The sensor system is composed of MEMS accelerometers with 5-10 mg RMS noise characteristics and gyroscopes that are subject to a 0.0056 ° / (sec $\cdot\sqrt{\text{sec}}$) angular velocity random walk according to the manufacturers. In addition, a camera records video images with a resolution of 1398x1080 pixels at 28 Hz. These images were scaled down to half size.

Because ground truth is not available for these experiments, the approaches have to be evaluated by comparing the reconstructed trajectory to the building's floor plan as shown in



(a)

(b)

Fig. 3. Simulation results for the sensor-centric mechanization without composition. See Fig. 2 for explanation.

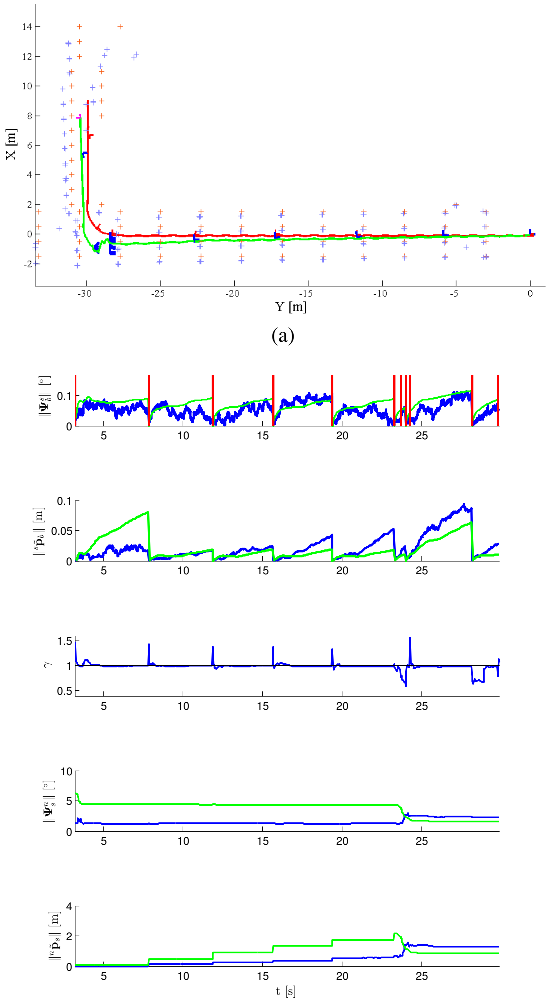
Fig. 5. The initial orientation of the sensor system w.r.t. the map was determined with an integrated compass.

Obviously, the sensor-centric approach using the composition step performs best on this dataset. This holds especially for the scale factor that is severely underestimated when the composition step is not applied. If a world-centric mechanization is used, the system fails approximately after the first quarter of the trajectory. Moreover, a substantial heading angle error can be observed in this case.

V. Conclusion

Two approaches to formulate the inertial mechanization equations are compared w.r.t. their applicability to inertial aided visual SLAM using an EKF. It is shown that consistency and robustness can be improved by formulating the mechanization equations w.r.t. the initial pose of the system and performing composition steps at regular intervals.

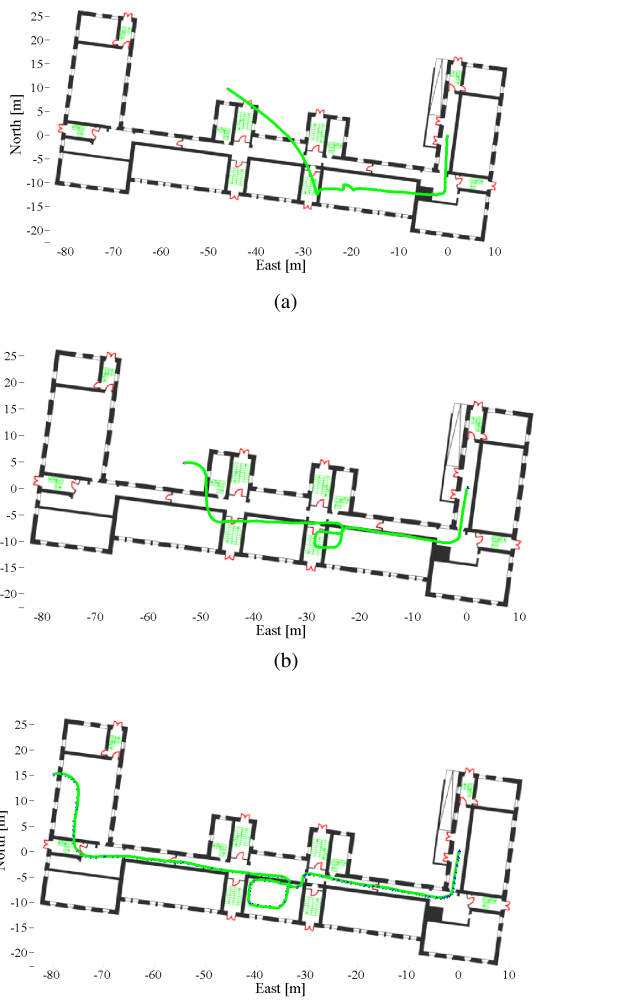
However, the filter is still inconsistent in some situations, especially when a lot of new landmarks are included to the state with arbitrary initial depth estimates. It is a fundamental flaw of the Kalman filter algorithm that it is not able to relinearize when the depth estimate for newly introduced landmarks improve. Furthermore, these linearization errors are accumulated over time when poorly estimated state variables are marginalized.



(a)

(b)

Fig. 4. Simulation results for the sensor-centric mechanization with composition. See Fig. 2 for explanation.



(a)

(b)

(c)

Fig. 5. Experimental results for an indoor trajectory. (a) World-centric mechanization (b) Sensor-centric without composition (c) Sensor-centric with composition.

References

[1] R. McCroskey, P. Samanant, W. Hawkinson, S. Huseth, and R. Hartman, "Glanser - an emergency responder locator system for indoor and gps-denied applications," in *23rd International Technical Meeting of the Satellite Division of The Institute of Navigation, Portland, OR, September 21-24, 2010*, 2010.

[2] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. The MIT Press, 2005.

[3] A. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. Ieee, 2003, pp. 1403–1410.

[4] H. Strasdat, A. J. Davison, J. M. M. Montiel, and K. Konolige, "Double window optimisation for constant time visual slam," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2011.

[5] J. Farrell and M. Barth, *The Global Positioning System & Inertial Navigation*. McGraw-Hill, 1999.

[6] M. J. Veth and J. F. Raquet, "Fusion of low-cost imaging and inertial sensors for navigation," in *Proceedings of the ION meeting on Global Navigation Satellite Systems*, 2006.

[7] G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis, "Observability-based rules for designing consistent ekf slam estimators," *The International Journal of Robotics Research*, vol. 29, no. 5, pp. 502–528, 2010. [Online]. Available: http://ijr.sagepub.com/content/29/5/502.abstract

[8] J. A. Castellanos, R. Martinez-Cantin, J. D. Tardós, and J. Neira, "Robocentric map joining: Improving the consistency of ekf-slam," *Robotics and Autonomous Systems*, vol. 55, pp. 21–29, 2007.

[9] M. George and S. Sukkarieh, "Inertial navigation aided by monocular camera observations of unknown features," in *IEEE International Conference on Robotics and Automation, ICRA*, 2007.

[10] E. S. Jones and S. Soatto, "Visual-inertial navigation, mapping and localization: A scalable real-time causal approach," *The International Journal of Robotics Research*, vol. 30, no. 4, pp. 407–430, 2011. [Online]. Available: http://ijr.sagepub.com/content/30/4/407.abstract

[11] J. Kelly and G. S. Sukhatme, "Visual-inertial simultaneous localization, mapping and sensor-to-sensor self-calibration," in *IEEE Transactions on Robotics, 24(5)*, 2009.

[12] T. Lupton, "Inertial slam with delayed initialisation," Ph.D. dissertation, School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney, 2010.

[13] A. Gelb, *Applied Optimal Estimation*. The MIT Press, ISBN: 0262570483, 1995.