Stochastic cloning Kalman filter for visual odometry and inertial/magnetic data fusion

Conference Paper · January 2013

CITATIONS
6

READS
409

5 authors, including:

Michailas Romanovas
German Aerospace Center (DLR)
57 PUBLICATIONS **325** CITATIONS
SEE PROFILE

Manuel Schwaab
Hahn-Schickard-Gesellschaft - Institut für Mikro- und Informationstechnik
14 PUBLICATIONS **98** CITATIONS
SEE PROFILE

Martin Traechtler
Hahn-Schickard-Gesellschaft - Institut für Mikro- und Informationstechnik
30 PUBLICATIONS **229** CITATIONS
SEE PROFILE

Stochastic Cloning Kalman Filter for Visual Odometry and Inertial/Magnetic Data Fusion

Michailas Romanovas*, Tobias Schwarze†, Manuel Schwaab*, Martin Traechtler*
and Yiannos Manoli‡
*Hahn-Schickard-Gesellschaft e.V., Institute of Microsystems and Information Technology (HSG-IMIT)
Villingen-Schwenningen, Germany
†Institute of Measurement and Control Systems (MRT), Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
‡Fritz Huettinger Chair of Microelectronics, Department of Microsystems Engineering (IMTEK)
Albert-Ludwigs-Universitaet Freiburg, Germany

***Abstract*—The work demonstrates the fusion of the position and the orientation information from Visual Odometry (VO) with the orientation information obtained from low-cost inertial and magnetic sensors. The proposed approach is based on the stochastic cloning (SC) Kalman filter formulation which is able to incorporate independent incremental measurements in a statistically consistent way. The algorithm was tested on realistic trajectories and compared to the results of a pure VO as well as to those of a decoupled system. A drift in the heading estimation is addressed by incorporating the Earth's magnetic field measurements with associated heuristics for a robustified disturbance detection.**

***Keywords*—Visual Odometry, Kalman Filtering, Stochastic Cloning, Pedestrian Localization, Inertial Measurement Unit.**

I. Introduction

Although an outdoor localization can be often addressed with GNSS-based techniques, its usage is mainly limited to open areas with a reliable access to satellite signals. The indoor or mixed indoor-outdoor scenarios seem to be far more challenging as the signals from satellites are often too weak to penetrate the walls, suffer from multiple reflections or are not available when walking through an urban canyon. Moreover, being long-term stable, the conventional low-cost GPS receiver alone is not accurate enough to locate an entrance or find a passage between obstacles like parked cars.

A recent progress in affordable wearable computing allowed to address the localization issue by fusing in real-time the information from multiple sensors with complementary noise properties. Numerous works have been reported using inertial and/or magnetic sensors, Wi-Fi, ultrasound, Ultra Wide Band (UWB), Bluetooth etc. The approaches can be self-contained (e.g. inertial sensor-based pedometer and compass) as well as dependent on an external infrastructure (e.g. Wi-Fi) and therefore restricted to operate within a limited area. The continuous improvement in computational performance of embedded hardware permits us now to fuse several information sources in real-time which till recently was only tractable in robotics and automotive applications with PC-level computational platforms. One of such examples is a camera-based Visual Odometry (VO) which provides an incremental position and orientation. Although the absolute information is still unknown to a non-augmented system, the integrated data are short-term stable. One of the main advantages of a VO-based localization is that it is a reference-free method and thus can be applied for large scale environments. However, apart of its significant computational complexity, the performance of the approach is strongly dependent on the quality of captured video and the system can easily fail in environments with insufficient illumination and monotone walls. Within the presented work the accuracy and robustness of the VO will be addressed by its fusion with yet another self-contained sensors: inertial and magnetic. Here a 3-axis MEMS accelerometer and a 3-axis MEMS gyroscope constitute an inertial measurement unit (IMU) and are supported by a 3-axis magnetic field sensor for the heading correction. The sensor is capable to provide an absolute orientation information with low relative uncertainty at high velocities and accelerations [1]. On the other hand the cameras can track the features very accurately at low velocities although their performance drops with increased dynamics due to motion blur and camera sampling effects. The cameras with higher frame rates can be hardly considered as a panacea due to existing bandwidth constraints and limitations of the associated CPU for real-time processing.

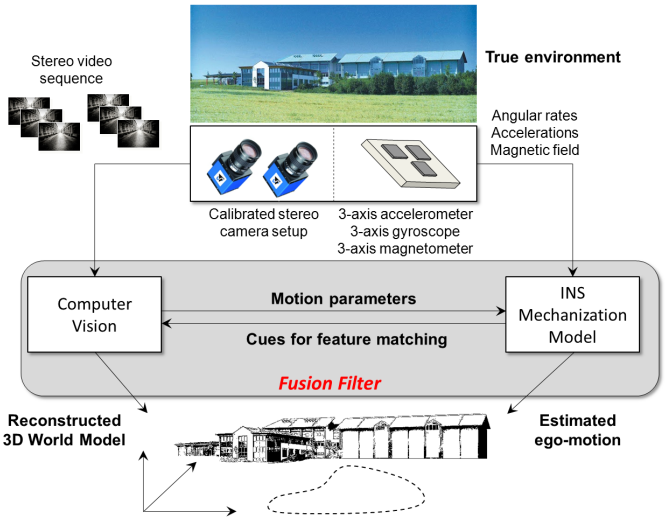

Fig. 1. A sketch of framework to visual-inertial data fusion.

A general framework for combining inertial and visual

information is shown in Fig. 1, where multiple camera and inertial measurements are fused into a common global world model with a visual reconstruction task. The motion parameters are derived from inertial and magnetic sensors using an inertial mechanization model and are supported by those inferred by the VO. Here the INS model provides the VO with dynamical cues for improved feature matching. The interaction between INS mechanization and computer vision can be done very tight ultimately resulting in a single fusion filter to process both the VO and sensor measurements. This tightly coupled approach is believed to be extremely useful compared to the decoupled filter strategy shown here.

The remainder of the paper is organized as follows. In Section II we provide a brief discussion on several of the related approaches. Section III presents some of the mathematical methods employed in the work along with associated dynamical models. The system is discussed in Section IV with the results discussed in Section V. Section VI concludes the paper and outlines the future work.

II. Related Work

A representative survey on Kalman Filtering (KF) for robotic vision applications can be found in [2]. A particular idea of combining the vision with other self-contained sensors is also not new [3] with a tutorial introduction provided in [1]. The performance of both Unscented KF (UKF) and Extended KF (EKF) for fusion of vision and inertial data is compared in [4]. The ability of the EKF to integrate successfully multiple types of visual feature observations with measurements from the IMU have been also confirmed in [5] where a tightly integrated EKF was proposed for fusion of visual feature observations with the measurements from IMU for precision planetary landing. Work [1] reported on a system where the gravity vector obtained from the inertial sensors was used as a vertical reference for improved scene interpretation or employed as a center of rotation for vision-based navigation. A loosely-coupled approach for fusion of VO and IMU was proposed in [6] where the feature correspondence was improved by looking for the matching points within a disk of fixed radius from a predicted point based on the attitude change from the previous frame. A combination of a VO and IMU was also reported in [7] for a slip compensation in planetary exploration rovers.

Some works also have addressed the problem of fusing the visual and inertial data for pedestrian localization and tracking [3], [8] often with a pre-registered image mapping for the position calculation. Unfortunately such database approaches can be hardly extended for larger scale uncontrolled environments. The work [3] employed step counting for position estimation between the images while [8] reported on a visual-inertial tracking device using 3D model of the scene for feature appearance prediction.

III. Methods

A. Kalman Filtering

The position and orientation estimation is often formulated as a state estimation problem for the system as [9]:

$$x_k = f(x_{k-1}, u_k, w_k), \quad (1)$$

$$z_k = h(x_k, \epsilon_k), \quad (2)$$

with $x_k \in \mathbb{R}^n$ being state of the system at time t_k with associated measurements $z_k \in \mathbb{R}^m$ and $f(\cdot)$ and $h(\cdot)$ being the nonlinear process and measurement functions with w_k and ϵ_k being the process and measurement noises. The nonlinear angular rate integration model as well as nonlinear relationship between the state and vector observations prevent the usage of a linear KF and one of its nonlinear versions such as EKF or UKF has to be adopted.

Within the EKF both the process and measurement models are linearized through a first-order Taylor series expansion around the most recent state estimate. Although the original nonlinear process and measurement models are used for the state transition and the measurement prediction, the covariances are propagated via corresponding Jacobian matrices calculated with respect to the current state estimate, control input and noises. The algorithm structure strongly resembles the one of the linear KF. For models with weak nonlinearites the EKF performs rather well and due to its well-understood properties is often considered as a de-facto standard for tracking applications [10].

A more recent alternative to the EKF is the UKF, where the probability distribution is approximated by a set of so-called σ-points. The points are deterministically selected from the distribution to preserve the Gaussian properties under the nonlinear transformations. The points are propagated through the original nonlinear process and measurement model equations with no Jacobians to be calculated. The filter was proved to have better statistical properties [11] and can be applied for models with discontinuities or those where analytical Jacobians are hard to derive. Within the work we have tried both EKF and UKF for orientation estimation and EKF was selected for the given tracking problem due to its clear computational advantages and close to UKF performance.

B. Visual Odometry

The process of estimating the motion of a moving platform solely based on the input of cameras rigidly attached to it, is commonly referred to as Visual Odometry (VO). Since robotic systems perceiving their environment are often equipped with cameras, VO offers a valuable input to the localization task for comparably little extra computational cost. The increase of computational power enables VO computation on wearable platforms with framerates of 15 fps and more, achieving position errors as small as 1-2%. A few conditions obviously need to be met: scene illumination has to be sufficient to avoid motion blur, the captured scene needs to contain apparent static parts and camera motion has to be small enough to contain overlap with the previously captured scene.

The general goal of VO is to find the rotation matrix R and translation vector d that relate the camera pose at time t_k to the camera pose at t_{k-1} through the rigid transformation

$$T_{k,k-1} = \begin{bmatrix} R & d \\ 0 & 1 \end{bmatrix}.$$

Incremental concatenation of the transformations results in an overall transformation with respect to the initial coordinate frame with a small drift accumulating due to solely incremental measurement nature.

The transformation T can be essentially obtained from a single camera by finding 2D image feature correspondences in two subsequent camera frames. Image features have to be detected and matched across the multiple images of the same scene taken from close-by viewpoints with a whole variety of feature detectors ranging from corner detectors to blob detectors to be applied [12]. Once the features are matched across two subsequent images, a minimum of five point correspondences is sufficient to calculate the essential matrix E, from which the rotation matrix R and the translation d can be extracted up to an unknown scale [13].

For a stereo vision setup 3D features can be obtained for each image pair by triangulating stereo image features. Then, a minimum of three 3D feature correspondences is needed to find a translation that aligns the 3D features of the two consecutive stereo camera poses. Normally more features are available and can be incorporated by minimizing the L_2 distance

$$\sum_i ||X_k^i - T \cdot X_{k-1}^i||$$

between the two sets. Triangulated 3D points are disadvantageous especially in extensive environments, where they are subject to big reconstruction errors due to large point to camera distances. The common solution is to triangulate stereo features from pose $k-1$ and project them into either camera frame or pose. The sought transformation T minimizes the reprojection error

$$\sum_i ||x_k^i - K \cdot T \cdot X_{k-1}^i||$$

between the 3D and 2D image features. 2D image features in homogeneous coordinates of the current frame are denoted as x_k^i, triangulated 3D features of the previous frame as X_{k-1}^i, where K is the known camera calibration matrix.

For accurate VO it is inevitable to embody an outlier rejection scheme which removes false feature associations and features that belong to non-static scene parts. This is usually achieved by repeatedly estimating the transformation on a randomly chosen minimal subset of feature points, and finally choosing the transformation which is most consistent with the whole feature set. Additionally, VO is often complemented by a local bundle adjustment step over the last several poses [14].

C. Models

For the rigid body orientation we have chosen a quaternion representation due to its computational efficiency and singularity free description when compared to Euler angles. The quaternion q is a number with four real components (q_w, q_x, q_y, q_z) where q_w is the scalar part of the quaternion and (q_x, q_y, q_z) constitute the vector part. For the quaternion to be a valid orientation reprsentation it has to be of a unit norm with a side effect that the quaternion components are no longer independent and have to be handled carefully within the filter design.

The orientation estimation is based on the system state $x_k \in \mathbb{R}^7$ given as:

$$x_k = \left[q_k^T, b_{\omega,k}^T\right]^T, \tag{3}$$

where q_k is the orientation quaternion representing the rotation from body to navigation frame and $b_{\omega,k}$ is the gyroscope offset. The process model for the quaternion is a discrete integration of the input angular rate formulated as a control input in order to preserve the dynamics without an explicit angular rate modelling. The angular rate itself can be expressed as:

$$\omega_k = \tilde{\omega}_k - b_{\omega,k-1} + w_{\omega,k}, \tag{4}$$

where $w_{\omega,k} \sim \mathcal{N}(0, Q_{\omega,k})$ is the gyroscope additive noise and $\tilde{\omega}_k$ is the actual gyroscope measurement. The integration problem can be reformulated as the differential rotation during the sampling period Δt with the angle $\phi_{\Delta,k} = |\omega_k| \cdot \Delta t$ and the associated rotation axis $u_{\Delta,k} = \omega_k / |\omega_k|$. The corresponding differential rotation quaternion q_Δ is constructed as:

$$q_\Delta = \left[\begin{array}{cc} \cos\left(\frac{\phi_{\Delta,k}}{2}\right), & u_{\Delta,k} \sin\left(\frac{\phi_{\Delta,k}}{2}\right) \end{array}\right], \tag{5}$$

with the quaternion propagation model $q_k = q_{k-1} q_\Delta$. The noise disturbance is introduced in a similar way by replacing q_Δ with noisy $\tilde{q}_\Delta = q_\omega q_\Delta$:

$$q_k = q_{k-1} \tilde{q}_\Delta = q_{k-1} q_\omega q_\Delta. \tag{6}$$

Here we apply the process noise via the term q_ω constructed from the noise statistics of ω_k. The process model for the gyroscope bias propagation is assumed:

$$b_{\omega,k} = b_{\omega,k-1} + w_{b_\omega,k}, \tag{7}$$

with $w_{b_\omega,k} \sim \mathcal{N}(0, Q_{b_\omega,k})$.

The filter correction step is based on two vector measurements: gravity and the Earth's magnetic field. The acceleration measurement $\tilde{a}_k$ is taken under the assumption that no significant linear acceleration is present and the observations are generated solely by the Earth's gravity and additive sensor noise:

$$\tilde{a}_k = q_k^{-1} g q_k + \epsilon_{G,k}, \tag{8}$$

with $\epsilon_{G,k} \sim \mathcal{N}(0, R_G)$ and g is a vector quaternion constructed from a zero scalar part and the gravity vector itself. The term $\epsilon_{G,k}$ has to accomodate significant albeit unmodelled disturbances due to human motion and walking and is set higher than the intrinsic sensor noise. Simialarly, the magnetic field measurements $\tilde{m}_k$ are taken:

$$\tilde{m}_k = q_k^{-1} m q_k + \epsilon_{M,k}, \tag{9}$$

with m being a vector quaternion corresponding to the Earth's magnetic field at the given location and $\epsilon_{M,k} \sim \mathcal{N}(0, R_M)$. Again, R_M is assumed much larger than the intrinsic sensor noise in order to accomodate the unmodelled magnetic field disturbances.

An additive structure of the state correction equation of the conventional EKF could result in filter divergence as the norm of the quaternion is not preserved. Here we follow a quaternion magnitude $\tilde{q}_k$ pseudo-measurement approach [15], where the quaternion norm has to be equal 1:

$$\tilde{q}_k = 1 = \|q_k\| + \epsilon_{q,k}, \tag{10}$$

with $\epsilon_{q,k} \sim \mathcal{N}(0, R_Q)$.

D. Stochastic Cloning KF

Although the presented model-free formulation of the orientation estimation without the angular rate in the kinematic state has some clear advantages such as avoidance of explicit angular rate modeling and reduced state dimensionality, it introduces certain difficulties when combining both VO orientation and gyroscope measurements as both of them become relative measurements with the sample rate of the gyroscope being much higher than that of the VO. The gyroscope is used to predict the state x_k from the previous state x_{k-1} using the integration, whereas the incremental rotation from VO relates the quaternion q_k to the quaternion q_{k+n}[1]. The problem can be addressed by preserving a copy of the quaternion q_k until the next VO measurement is available using a framework of a stochastic cloning EKF (SC-EKF) [16].

Here at a time t_k after the VO frames have been processed, the state's quaternion $q^k_{C,k} = q_k$ is cloned to produce an augmented state x^a_k:

$$x^a_k = \begin{bmatrix} q^T_k & b^T_k & (q^k_{C,k})^T \end{bmatrix}^T. \tag{11}$$

To properly reflect the correlation between the cloned and the rest of the state, the covariance matrix P_k is also augmented:

$$P^a_k = \begin{bmatrix} P_q & P_{qb_w} & P_q \\ P^T_{qb_w} & P_{b_w} & P^T_{qb_w} \\ P_q & P_{qb_w} & P_q \end{bmatrix}, \tag{12}$$

where P_q is the covariance term for q_k, P_{b_w} - for gyroscope bias, and P_{qb_w} being the associated cross-covariance. The cloned quaternion $q^k_{C,k}$ remains unchanged during the filter prediction cycles while the covariance propagation follows an augmented state transition matrix F^a_k:

$$F^a_k = \begin{bmatrix} F_k & 0 \\ 0 & I \end{bmatrix} \tag{13}$$

and F_k being the Jacobian of the original system with no process noise assumed for the cloned quaternion. The previously described corrections using gravity and magnetic field remain unchanged, only the measurement matrix has to be adjusted slighlty i.e. $H^a_k = [H_k \quad 0]$ in order to reflect the augmentation of the state.

At the time t_{k+n} the next pair of frames is grabbed and the rotation matrix $R_{k+n,k}$ is calculated by the VO filter. The matrix is transformed into the 3x1 rotation angle vector $\alpha_{k+n,k} = \phi_{k+n,k} u_{k+n,k}$ to be employed as a pseudomeasurement in

$$h_{VO}(x^a_{k+n}, \epsilon_{VO,k}) = \alpha_{k,k+n}, \tag{14}$$

and $\epsilon_{VO,k} \sim \mathcal{N}(0, R_{VO})$. Here the function $h_{VO}(\cdot)$ for state $x^a_{k+n} = \begin{bmatrix} q^T_{k+n} & b^T_{k+n} & (q^{k+n}_{C,k})^T \end{bmatrix}$ is formulated as:

$$h_{VO}(x^a_{k+n}) = 2\cos^{-1}(q_{\Delta,w}) \frac{q_{\Delta,\{x,y,z\}}}{\sin(\cos^{-1}(q_{\Delta,w}))} + \epsilon_{VO,k}, \tag{15}$$

with $q_\Delta = q_{\Delta,k+n,k} = (q^{k+n}_{C,k})^{-1} q_{k+n}$. Thus the incremental rotation between t_k and t_{k+n} given by the VO is compared to the rotation difference between the quaternions $q^{k+n}_{C,k}$ and q_{k+n} which represents the estimated orientation change between t_k and t_{k+n} based on the gyroscope integration. Note that both $q^{k+n}_{C,k}$ and q_k represent the orientation at time k, for $q^{k+n}_{C,k}$ the measurements up to time t_{k+n} are considered and the covariance of the cloned quaternion as well as associated cross covariance to the rest of the state are properly treated in the filter ensuring that the fusion is statistically consistent. The procedure ends up with discarding and re-augmenting of the cloned part of the state and the covariance matrix.

[1]With 400Hz IMU and 15Hz VO sampling rates we have $n = 400/15 = 26.\overline{6}$. In practice we use the closest neighbour (respectively $n = 26$ or $n = 27$) to avoid the interpolation between the two states. This assumption seems to be valid due to relatively low dynamics of the head motion compared to the IMU data rate.

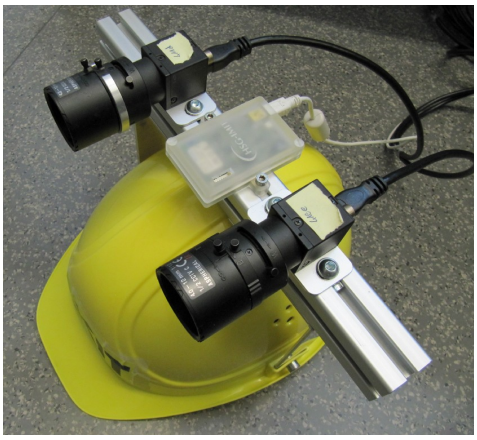

Fig. 2. Camera and IMU example used throughout the measurements.

IV. Setup

A. Camera Setup

The experimental setup as shown in Fig. 2 consists of a stereo rig with two Flea2 Firewire cameras equipped with wide-angle lenses of 12 mm focal length mounted with a short baseline of around 12 cm. Since the localization is a part of a larger system with real-time restrictions we limit the video resolution to 640x480px and capture video data with 15fps. In all our experiments the exposure time control is dedicated to the camera, which leads to considerable motion blur in indoor scenarios. The stereo cameras are calibrated using the publicly available camera calibration toolbox [17]. At the moment we do not treat the extrinsic calibration between camera and IMU sensors as the misalignment is expected to be negligible due to close mounting positions.

B. Inertial Measurement Unit

A custom-made sensor unit is shown separately in Fig. 3 and is based on low-cost commercial MEMS sensors and a low-power TI MSP430 microcontroller for sensor readout. Both acceleration and Earth's magnetic field are measured using STM LSM303DLH tilt compensated compass while the angular rate is obtained from STM L3G4200D 3-axis gyroscope. The data packets are transmitted to the host via USB at 400 Hz rate and are logged and processed by a custom written software. As the IMU and the magnetometer provide an absolute attitude, one has to ensure that reference vector measurements m and g correspond to the true "zero" orientation $q_0 = [1, 0, 0, 0]^T$ which has to be compensated within the system initialization phase (see Fig. 4). If not

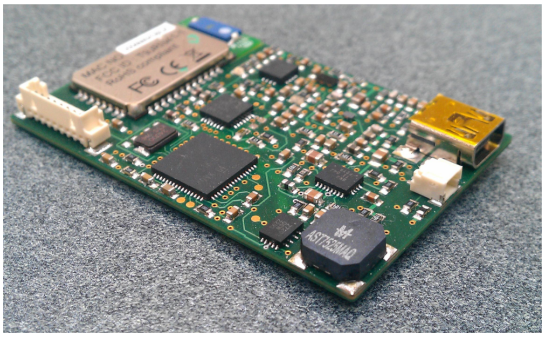

Fig. 3. Custom-made inertial and magnetic sensor unit.

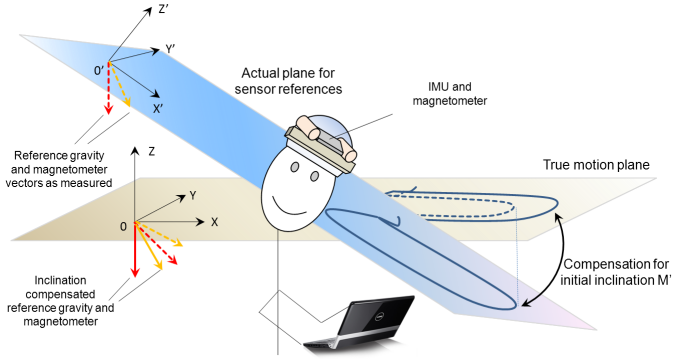

Fig. 4. Sensor reference compensation mechanism.

handled properly, this results in a trajectory within the plane X'Y' which in general is different from the one assumed for the visualization. The correction is done by employing the pitch and roll information (equivalent to rotation M') from the accelerometer data during the startup phase.

C. Algorithm Configuration

The general data flow for the proposed SC-EKF filter is shown in Fig. 5, where the substitution filter can be obtained by removing the block with VO information and covariance input to the KF. The VO estimation is carried out by the LIBVISO2 library [18], which implements the motion estimation based on 3D to 2D feature correspondences using features matched between all four images of two consecutive stereo camera poses. The VO and IMU filter are run independently and due to low performance of the MEMS sensors no position information is provided by the inertial part. No special calibration has been performed for the sensor board except of soft ironing compensation [19]. The position increment d from the VO is transformed from camera to IMU coordinate frame and is integrated to obtain the position using the orientation quaternion from the separate orientation filter. The substitution filter completely discards the VO orientation information and uses only incremental position with IMU-based orientation estimation. However, ignoring such an important and complementary piece of information results in an estimation performance far from optimal. The SC-EKF approach retains the basic structure of the substitution filter and assumes a fixed noise covariance for VO data. Note that the latter assumption can be hardly justified as the VO error is in general non-Gaussian, can be correlated

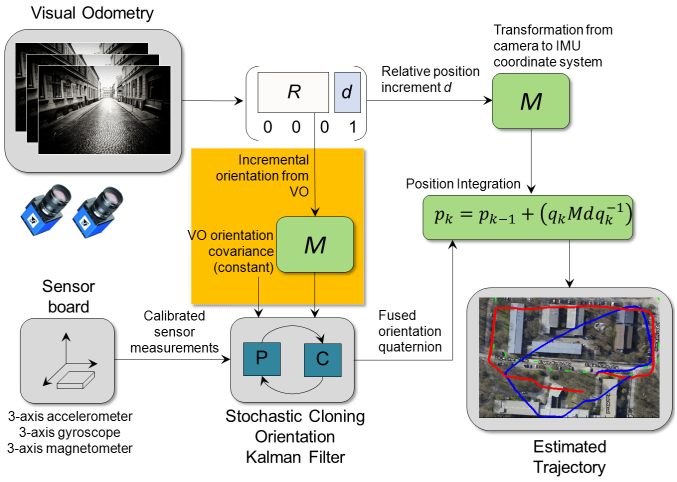

Fig. 5. Data flow within the filter.

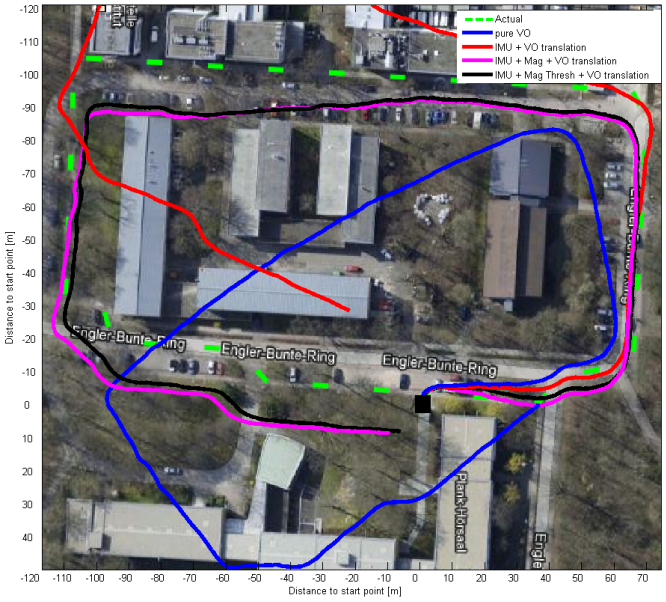

Fig. 6. Estimated XY trajectory for the first trajectory using the substitution filter. The black square denotes the start of the true path. Total path length is appr. 530 meters.

between the consecutive frames and varies with the actual video quality. Currently the recorded data are processed off-line while online implementation on the embedded platform is planned for the future.

V. Results

In order to evaluate the approaches we recorded different datasets consisting of basic outdoor as well as more complex indoor scenarios comprising e.g. staircases. The estimation results of the substitution filter with and without magnetic field for heading compensation are shown in Fig. 6 and 7. In both the cases we have used $Q_{b_\omega,\{x,y,z\}}$ = 1e-8 and $Q_{\omega,\{x,y,z\}}$ = 1e-4. The measurement noises were taken $R_{G,\{x,y,z\}} = 0.25$ and $R_{M,\{x,y,z\}} = 0.1$ and $R_{Q,\{x,y,z\}} = 0.1$ with the initial uncertainty set to $P_{0,q,\{x,y,z\}}$ = 1e-1 for the quaternion

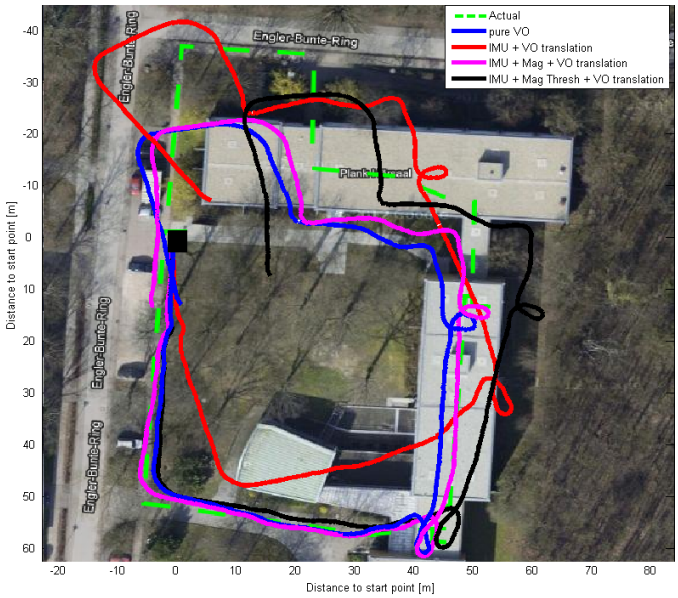

Fig. 7. Estimated XY trajectory for the second trajectory using the substitution filter. The black square denotes the start of the true path. Total path length is appr. 267 meters.

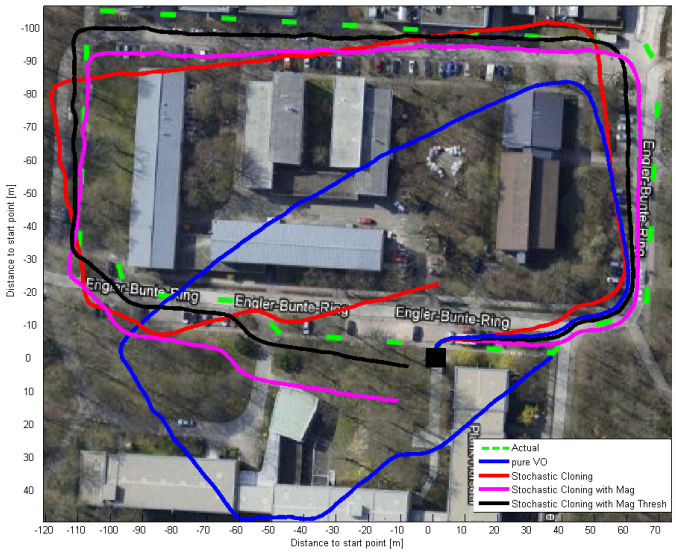

Fig. 8. Estimated XY trajectory for the first trajectory using the SC-EKF filter.

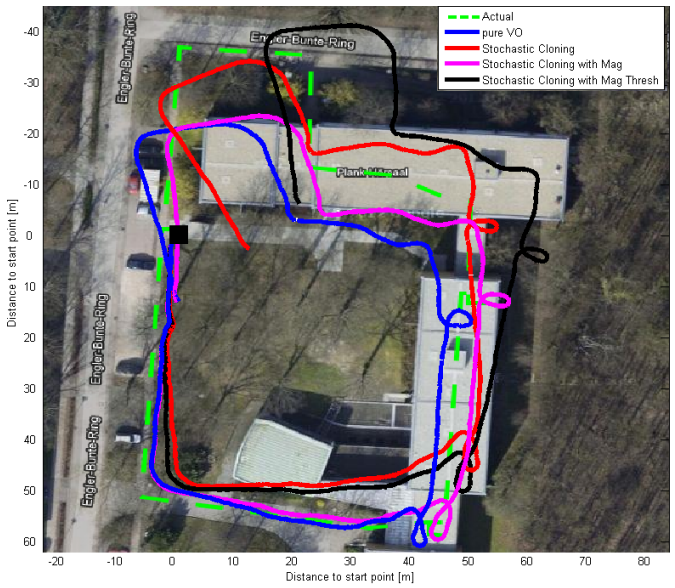

Fig. 9. Estimated XY trajectory for the second trajectory using the SC-EKF filter.

and $P_{0,b_\omega}, \{x, y, z\}$= 1e-4 for the bias. For the SC-EKF a constant value of $R_{VO,\{x,y,z\}}$=1e-6 worked rather well. The magnetic field amplitude threshold was set to 1% (e.g. range 0.99...1.01) when applicable.

Although pure VO provides a reasonably good orientation for the second path (see Fig. 7) it shows significant drift for the 1st trajectory (Fig. 6). A direct substition of the VO orientation with the one from pure inertial filter does not result in a performance improvement, although the position drift follows a different pattern due to intrinsic inability of the filter to correct the heading drift. Although in theory the heading drift can be corrected with the Earth's magnetic field measurements, in practice these measurements are strongly distorbed in a proximity of larger metallic objects with the situation often much worse for indoor scenarios due to electronic equipment and building structures. Fig. 8 and 9 present the results for an associated SC-EKF where it outperforms the substitution method for pure inertial sensing as expected due to its inherent mechanism for VO and IMU noise trade-off. However for the approach with continuous magnetic field measurements a significant improvement can be observed only for the first trajectory with a marginal improvement for the second one. We believe this behavior is caused by the magnetic field disturbances for indoor walking segments where the noise assumptions are not valid anymore.

As an alternative we have designed a heuristic approach of detecting the magnetic field disturbances by setting the validation gate (threshold) of 1% on the amplitude of the magnetic field. Unfortunately this approach does not always result in performance improvement as one can see in the second walking scenario. Although further filter tuning is always possible, we believe that the performance is a result of the "hard thresholding" scheme and more advanced detection schemes have to be designed instead. Recall that in all the cases both IMU and VO do not provide an absolute reference for the position and therefore the accumulated error cannot be completely avoided within the presented system. The algorithm simultaneously estimates the position in Z direction and the vertical position consistency is also important for the local 3D environment to be perceived correctly. Fig. 10 provides an associated XZ projection for the first trajectory and significant improvement in Z position can be seen for both substitution EKF and SC-EKF approaches.

The importance of augmentation of the VO with independent sensor data can be clearly seen in Fig. 11 where the VO failed to provide a proper orientation within the region "A" as the user had climbed stairs 2 floors up in a corridor with low illumination. This is a rather well-known problem of the VO caused by insufficient features found and tracked between

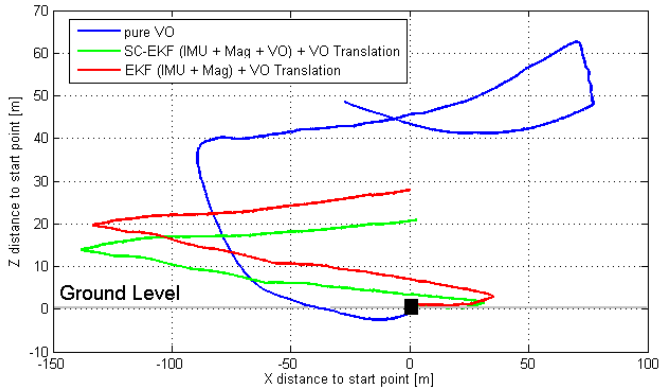

Fig. 10. Estimated XZ position for the 1st trajectory using the substitution filter and the SC-EKF approach. Black square denotes the start of the path. The true path should be always close to Z = 0 meters level.

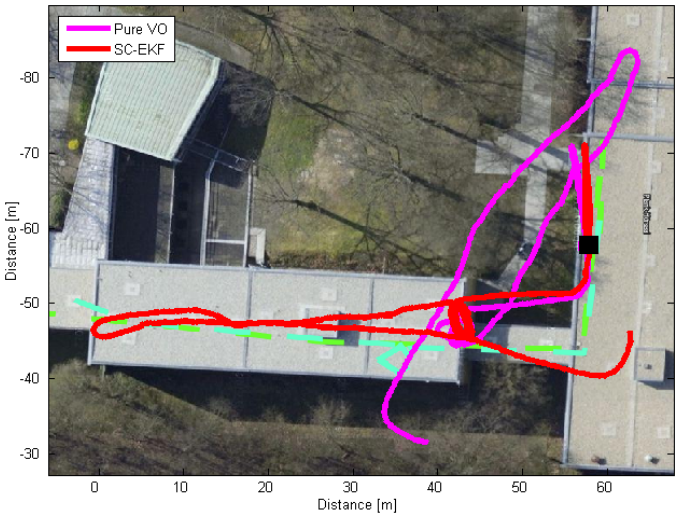

Fig. 11. An example of an indoor scenario where VO fails due to low illumination and poor quality of features extracted. Black square denotes a start of the trajectory. SC-EKF is used with constant magnetic field (no threshold). Green line denotes a true walking path

images and is typical for indoor scenarios with monotone or textureless walls. Differently the SC-EKF showed to be robust and was able to preserve the orientation in spite of the VO failure. Moreover, a pure VO approach showed a significant vertical drift (see Fig. 12) when walking through darker corridors. Note that the VO in "A" region still resulted in relatively good vertical position displacement although the heading estimation failed.

Although both camera data and IMU measurements are assigned a proper system time stamp by the logging software, there could be a non-negligible delay before the VO output is available due to limited processing power of a typical embedded system. Interestingly, such a delay almost has no effect on the performance of the substitution filter although the SC-EKF showed an increased sensitivity to the delay (Fig. 13). This is caused by an actual discrepancy between the incremental orientation measurements from both sensors which is not present in a substitution filter. The delay could become of a crucial importance for the coupled approaches

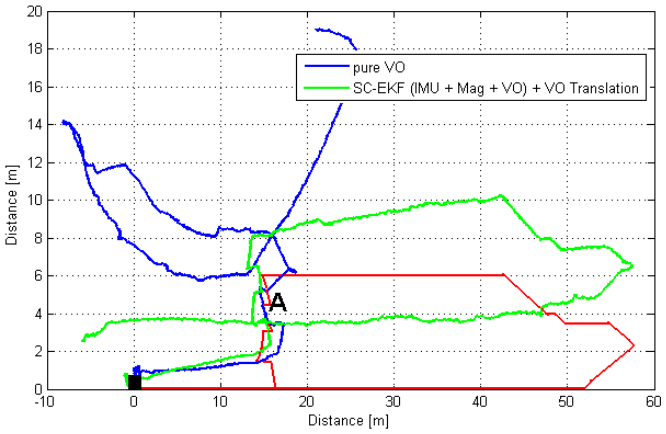

Fig. 12. An indoor scenario where VO fails due to low illumination and low quality of features extracted. Projection on XZ plane. Red line denotes a true walking path.

and therefore has to be properly addressed within the real-time implementation. Although a straightforward solution can be found by buffering the measurements and processing them later in a batch mode, a more elegant solution can be envisioned by extrapolating the measurements as suggested in [20]. The computational complexity of the presented decoupled approaches is mainly limited by that of the VO processing as the orientation EKF or SC-EKF adds only a little overhead to the total demand.

VI. Conclusion

The work proposes a method for localization using VO augmented with affordable inertial and magnetic sensors. The system is self-contained and allows to navigate over significant distances without external referencing. The approach is based on a stochastic cloning EKF and combines the multi-rate incremental measurements of both sensor modalities. The technique was tested for both indoor and outdoor environments and was augmented with magnetic disturbance detection to robustify its performance for indoor scenarios. The approach showed a significant reduction of the positioning error compared to a pure VO approach and improved robustness for the cases with low illumination and monotone walls.

Further work is planned in improving the compensation mechanism for magnetic field disturbances and the misalignment calibration between the sensors. A gain is also expected when a loosely/tightly coupled approach is adopted [1] with a proper treatment of non-Gaussian noise properties of VO and magnetic field for indoor walking segments. The method is expected to help in feature mapping for highly dynamical motion conditions, reduce the search space and improve the consistence of the estimates as well as overall robustness of the scheme. Moreover, a tighter coupling can provide a fair performance for reduced video data rates and permit implementation on embedded paltforms, while existing schemes seem to be computationally prohibitive for a moderate performance hardware (e.g. ARM processors). Some improvements are also envisioned from implementing a complete inertial mechanization instead of a simple orientation filter, although the performance gain could be marginal due to performance

Fig. 13. Influence of the delay between VO and inertial/magnetic data. Negative delay value means delay of the VO observation with respect to IMU ones. All three lines of substitution EKF almost overlap confirming low sensitivity of the approach to a slight delay between VO and INS data.

of the commercial low-cost sensors employed in the existing hardware.

ACKNOWLEDGMENT

Presented research activities are funded by the Federal Ministry of Education and Research (BMBF, Germany) under the contract 13EZ1126D.

REFERENCES

[1] P. Corke, J. Lobo, and J. Dias, "An introduction to inertial and visual sensing," *The International Journal of Robotics Research*, vol. 26, no. 6, pp. 519–535, 2007.

[2] S. Chen, "Kalman filter for robot vision: A survey," *Industrial Electronics, IEEE Transactions on*, vol. 59, no. 11, pp. 4409–4420, Nov. 2012.

[3] M. Kourogi and T. Kurata, "A method of personal positioning based on sensor data fusion of wearable camera and self-contained sensors," in *Multisensor Fusion and Integration for Intelligent Systems, MFI2003. Proceedings of IEEE International Conference on*, July-1 Aug., pp. 287–292.

[4] P. Gemeiner, P. Einramhof, and M. Vincze, "Simultaneous motion and structure estimation by fusion of inertial and vision data," *Int. J. Rob. Res.*, vol. 26, no. 6, pp. 591–605, Jun. 2007.

[5] A. Mourikis, N. Trawny, S. Roumeliotis, A. Johnson, A. Ansar, and L. Matthies, "Vision-aided inertial navigation for spacecraft entry, descent, and landing," *Robotics, IEEE Transactions on*, vol. 25, no. 2, pp. 264–280, April.

[6] S. Saripalli, J. Roberts, P. Corke, G. Buskey, and G. Sukhatme, "A tale of two helicopters," in *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, vol. 1, Oct., pp. 805–810 vol.1.

[7] D. M. Helmick, S. I. Roumeliotis, Y. Cheng, D. S. Clouse, M. Bajracharya, and L. H. Matthies, "Slip-compensated path following for planetary exploration rovers," *Advanced Robotics*, vol. 20, no. 11, pp. 1257–1280, 2006.

[8] G. Bleser and D. Stricker, "Advanced tracking through efficient image processing and visual-inertial sensor fusion," in *Virtual Reality Conference, 2008. VR '08. IEEE*, March, pp. 137–144.

[9] D. Fox, J. Hightower, L. Liao, D. Schulz, and G. Borriello, "Bayesian filtering for location estimation," *Pervasive Computing, IEEE*, vol. 2, no. 3, pp. 24 – 33, july-sept. 2003.

[10] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, September 2005.

[11] R. van der Merwe and E. Wan, "Sigma-point Kalman filters for integrated navigation," in *Proceedings of the 60th Annual Meeting of The Institute of Navigation (ION)*, Dayton, Ohio, June 2004, pp. 1 – 14.

[12] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," *Found. Trends. Comput. Graph. Vis.*, vol. 3, no. 3, pp. 177–280, Jul. 2008. [Online]. Available: http://dx.doi.org/10.1561/0600000017

[13] D. Nister, "An efficient solution to the five-point relative pose problem," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 6, pp. 756–770, 2004.

[14] F. Fraundorfer and D. Scaramuzza, "Visual odometry : Part II: Matching, robustness, optimization, and applications," *Robotics Automation Magazine, IEEE*, vol. 19, no. 2, pp. 78–90, 2012.

[15] I. Y. Bar-Itzhack, J. Deutschmann, and F. L. Markley, "Quaternion normalization in additive EKF for spacecraft attitude determination," in *Flight Mechanics/Estimation Theory Symposium, SEE N92-14070 05-13*, 1991, pp. 403–421.

[16] S. Roumeliotis and J. Burdick, "Stochastic cloning: a generalized framework for processing relative state measurements," in *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, vol. 2, pp. 1788–1795 vol.2.

[17] A. Geiger, F. Moosmann, O. Car, and B. Schuster, "Automatic camera and range sensor calibration using a single shot," in *International Conference on Robotics and Automation (ICRA)*, St. Paul, USA, May 2012.

[18] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3D reconstruction in real-time," in *IEEE Intelligent Vehicles Symposium*, Baden-Baden, Germany, June 2011.

[19] L. Klingbeil, M. Romanovas, P. Schneider, M. Traechtler, and Y. Manoli, "A modular and mobile system for indoor localization," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, sept. 2010, pp. 1 –10.

[20] T. Larsen, N. Andersen, O. Ravn, and N. Poulsen, "Incorporation of time delayed measurements in a discrete-time kalman filter," in *Decision and Control, 1998. Proceedings of the 37th IEEE Conference on*, vol. 4, Dec, pp. 3972–3977 vol.4.