A MEMS-based Smart Sensor System for Estimation of Camera Pose for Computer Vision Applications

Dominik Aufderheide\textsuperscript{a,b}, Werner Krybus\textsuperscript{a}, Dennis Dodds\textsuperscript{b}

\textsuperscript{a}South Westphalia University of Applied Sciences, Division Soest
Institute for Computer Science, Vision and Computational Intelligence
Luebecker Ring 2, 59494 Soest, Germany
E-Mail: \{aufderheide, krybus\}@fh-swf.de

\textsuperscript{b}The University of Bolton, School of the Built Environment and Engineering
Deane Road, Bolton BL3 5AB, U.K.
E-Mail: \{dma1bee, d.dodds\}@bolton.ac.uk

Abstract

The estimation of a camera’s egomotion during image acquisition is a mandatory task for many different computer vision applications such as Structure from Motion (SfM), Simultaneous Localisation and Mapping (SLAM) or Augmented Reality (AR). The vast majority of the proposed applications are deriving the motion parameters indirectly from the captured images. This paper suggests a smart sensor system (S\textsuperscript{3}) composed from three different micro-electromechanical (MEMS) inertial sensor types as an aiding modality for vision-based camera pose estimation. The S\textsuperscript{3} implementation contains a signal conditioning unit and a bank of Kalman filters for orientation estimation. The whole system is evaluated by using an industrial robot for the generation of specific motion patterns and the corresponding ground truth orientation measurements.

Keywords: Kalman filter, MEMS, Smart Sensor Systems, Inertial Navigation, Multi-Sensor Data Fusion, Camera Egomotion Estimation

1. Introduction

For many different algorithms in the field of computer vision (CV) it is necessary to determine the absolute or relative pose (position and orientation) of the camera during the acquisition of a monocular image stream. In most cases the computed pose of the camera is a prerequisite for further computations. One prominent example is the field of Structure from Motion (SfM) which realises the simultaneous estimation of camera egomotion and computation of a three dimensional representation of an observed scene based on the captured image sequence. A detailed description of the SfM procedure is given e.g. in Pollefeys et al. [1998], Poelman and Kanade [1997]. Another example is the field of mobile robotics where methods for simultaneous localisation and mapping (SLAM) are recently developed which allow on the one hand the localisation of a single moving robot platform based on image frames (Visual Odometry - VO) and on the other hand the synchronous mapping of the robots environment. Prominent examples can be found in Davison and Kita [2002] and Pupilli and Calway [2006]. Closely related is the field of parallel tracking and mapping (PTAM) which was initially solved by Klein and Murray [2007] for applications in Augmented Reality (AR). AR is also a typical example for the usage of the camera pose in CV applications. The general idea of many AR applications is the placement of 3D computer generated graphics (CGI) in a video captured by a standard camera. The perspective view of the 3D CGI model has to be rendered based on the actual position of the camera to generate a natural appearance of the artificial object. Examples of such procedures can be found e.g. in Schmalstieg and Wagner [2007].

The vast majority of all recently suggested procedures in those application fields are based on the successful detection and tracking of distinctive features through all frames of the captured sequence. For this it is necessary to (i) detect distinctive features in the images and (ii) track and match those features throughout all successive frames. In most cases points of interest
(PoIs) are detected by different PoI detectors suggested in literature, such as Harris-features (Harris and Stephens [1988]), SIFT-features (Lowe [2004]) or SURF-features (Bay et al. [2008]). For tracking and matching many different frameworks were suggested in literature during the last decades, whereat it is possible to distinguish between those approaches which rely only on a small subset of anchor features or those which try to maximise the number of point features even if there is the possibility for the generation of wrong matches (outliers). These two different approaches emphasise different subtasks, while for anchor feature approaches sophisticated tracking mechanisms (e.g. Hidden Markov Models (HMM) as suggested by Xie and Evans [1990]) have to be implemented, it is mainly necessary to include routines for outlier handling for the other alternative (e.g. Random Sample Consensus (RanSaC) Nister [2003]). Nevertheless the prerequisite of reliable feature tracks is often very hard to accomplish, because the feature matching between successive frames suffers from problems such as motion blur, perspective projection, repetitive patterns, less textured areas, computational complexity, etc. In Aufderheide et al. [2009] and Steffens et al. [2009] different typical problems classes of point registration are identified. Especially for long-time sequences the robust tracking of natural landmarks is an almost open question in the computer vision community. This paper suggests a smart sensor system ($S^3$) composed as a bank of different micro-electromechanical systems (MEMS). The proposed system contains accelerometers, gyroscopes and magnetometers. All of them are sensory units with three degrees of freedom (DoF). The $S^3$ contains the sensors itself, signal conditioning (filtering) and a multi-sensor data fusion (MSDF) scheme for orientation estimation. The performance of the system is evaluated by using different motion patterns generated by an industrial robot. This allows the generation of ground truth data. The system was compared against other possible fusion schemes. The remainder of this paper is organised as follows: Section 2 describes the general architecture of the proposed $S^3$. The following Sections introduce the different stages of the system, namely hardware (Section 3), signal conditioning (Section 4) and Sensor Fusion (Section 5). An overview about the experimental evaluation of the system is given in Section 6. Finally Section 7 concludes the whole paper and shows possible future work.

2. General $S^3$ architecture

The general architecture of the $S^3$ is shown in the following Figure 1. Whereat the overall architecture contains mainly the sensory units as described in subsection 3.1. A single micro controller is used for analog-digital-conversion (ADC), signal conditioning (SC) and transfer of sensor data to a PC (see Section 4). The actual sensor fusion scheme, as described in Section 5, for the estimation of orientation is realised in the PC for a better visualisation.

![Figure 1: General architecture of the inertial $S^3$](image)

3. Hardware

The main hardware which is used to build the $S^3$ is of course the bank of inertial sensory units in MEMS technology. An additional micro controller ($\mu C$) was added mainly for signal conditioning purposes.

3.1. Sensory units

The general architecture of the proposed $S^3$ system is based on three different types of inertial measurement units (IMUs). The whole system consist of three orthogonal arranged accelerometers which measure a three dimensional acceleration $a^b = [a_x \ a_y \ a_z]^T$ and three gyroscopes which measure the angular velocities $\omega^b = [\omega_x \ \omega_y \ \omega_z]^T$ around the sensitivity axes of the accelerometers. As an addition three magnetometers are able to determine the earth magnetic field in 3 DoF $m^b = [m_x \ m_y \ m_z]^T$. All of the quantities are measured with reference to the body coordinate frame (here indicated by subscript $b$) which is
rigidly attached to the IMU-platform. The following figure shows the general configuration of all sensory units and the corresponding measured entities.

Figure 2: General architecture of the IMU

Due to the fact that the whole unit should be available for low-costs only off-the-shelf sensors are used, whereat a single-axis gyroscope LY530AL and a LPR530AL dual-axis gyroscope from STMicroelectronics are used to measure the angular velocities around all three axes. Analog Devices provides with the ADXL345 a full 3-DoF accelerometer unit in a single chip. For measuring the earth’s magnetic field, Honeywell provides a 3-DoF magnetometer (HMC5843).

3.2. Processing units

The different sensory units are connected to the micro controller by using either I²C-Bus (Accelerometer and Magnetometer) or direct ADC of voltage output signals (Gyroscopes). Due to the different shape of the signals the complete signal conditioning was placed inside the μC in digital domain. We use a ATMEGA 328 processor from AVR to collect data from all information channels and subsequent signal conditioning and transfer by USB to the PC.

4. Sensor Modelling and Signal Conditioning

Measurements from MEMS devices in general and inertial MEMS sensors in particular are suffering from different error sources. Due to this it is necessary to implement both an adequate calibration routine and a signal conditioning routine.

The calibration of the sensory units is only possible if a reasonable sensor model is available beforehand. The sensor model should address all possible error sources. Here the proposed model from Skog and Händel [2006] was utilised and adapted for the given context. It can be shown that the main influences for the occurrence of measurement errors are the following (see Petkov and Slavov [2010]):

- Misalignment of sensitivity axes - Ideally the three independent sensitivity axes of the sensor should be orthogonal. Due to imprecise construction of MEMS-based IMUs this is not the case for the vast majority of sensory packages. The misalignment can be compensated by finding a matrix $M$ which transforms the non-orthogonal axis to a orthogonal setup as shown in Dorobantu [1999].

- Biases - The output of a sensor should be exactly zero if the $S^3$ is not moved. Also this is not true, because it was shown e.g. in Gulmammadov [2009], that there is a time-varying offset. Here Aslan and Saranli [2008] differentiate $g$-independent biases (e.g. for gyroscopes) and $g$-dependent biases. For the later there is a relation between the applied acceleration and the bias. The bias is modelled by incorporation of a bias vector $b$

- Measurement noise - Of course also the general measurement noise has to be taken into account, whereat it is assumed here as a white noise term $n$.

- Scaling factors - In most cases there is an unknown scaling factor between the measured physical quantity and the real signal. The scaling can be compensated by introducing a scale matrix $S = diag(s_x, s_y, s_z)$.

Based on these error classes a general error-model based on the findings in Skog and Händel [2006] was used in this work. The general model is valid for the different inertial sensors. A block-diagram of the general sensor model is shown in the following figure.
Based on this it is possible to define three separate sensor models for all three sensor types\(^1\), as shown in the following equations:

\[
\begin{align*}
\omega_b &= M_g \cdot S_g \cdot \omega^*_b + b_g + n_g \quad (1) \\
a_b &= M_a \cdot S_a \cdot a^*_b + b_a + n_a \\
m_b &= M_m \cdot S_m \cdot m^*_b + b_m + n_m \quad (3)
\end{align*}
\]

It was shown that \(M\) and \(S\) can be determined by sensor calibration routines which move the sensor array to different known locations to determine the calibration parameters. So presented Hwangbo [2008] a calibration approach based on the factorisation of a measurement matrix which is inspired by methodologies from classical SfM. The noise and the bias terms can not be determined a-priori due to their time-varying character. The signal conditioning step on the \(\mu\)C takes care of the measurement noise by integrating an FIR digital filter structure. The implementation realises a low-pass FIR filter based on the assumption that the frequencies of the measurement noise are much higher than the frequencies of the signal itself. The complete filter was realised in software on the \(\mu\)C, whereby the different cutoff-frequencies for the different sensory units were determined in an experimental evaluation. Based on the conditioned signals it is now possible to fuse the measurements from the different sensors for attitude estimation as described in the next section.

5. Multi-Sensor Data Fusion

The general idea of the multi-sensor data fusion (MSDF) step is based on the redundancy in the measurements delivered by the bank of inertial sensors. This is important, because due to the immense influence of noise and biases it is not possible to rely only on one source of information. In MSDF it is possible to interpret the different sensors as independent information channels, as suggested by Mitchell [2007]. Classical approaches for inertial navigation are stable-platform systems which are isolated from any external rotational motion by specialised mechanical platforms. In comparison to those classical stable platform systems the MEMS sensors are mounted rigidly to the device (here: the camera). In such a strap-down system it is necessary to transform the measured quantities of the accelerometers into a global coordinate system by using known orientations computed from gyroscope measurements. In general the mechanisation of a strapdown inertial navigation systems (INS) can be described by the computational elements indicated in Figure 4. The necessary computation of the orientation \(\xi\) of the \(S^3\) based on the gyroscope measurements \(\omega_b\) and a start orientation \(\xi(t_0)\) can be described as follows:

\[
\xi = \xi(t_0) + \int \omega_b dt \quad (4)
\]

The integration of the measured rotational velocities would lead to an unbounded drifting error in the absolute orientation estimates. Figure 5 shows two examples for this typical drifting behaviour for all three Euler angles. For the two experiments shown in Figure 5 the \(S^3\) was not moved, but even after a short period of time (here: \(6000 \cdot 0.01s = 60s\)) there is an absolute orientation error of up to 4 recognisable. For

---

\(^1\)The different sensor types are indicated by the subscript indices at the entities in the different equations.
inertial reference frame \( \mathbf{a}_i \) only by double integration:

\[
\varphi = \varphi(t_0) + \int \int \mathbf{a}_i dt
\]  

(5)

On the other hand possible errors in the orientation estimation stage would lead also to a wrong position due to the necessity to transform the accelerations in the body coordinate frame \( \mathbf{a}_b \) to the inertial reference frame (here indicated by the subscript \( i \)).

The following figure gives an impression about the typical drifting error for the absolute position (one axis) computed by using the classical strapdown methodology. It can be easily seen that after 20 s the error is already drifted to approximately 13 m for a not moved device.

Figure 6: Drifting error for absolute position estimates based on classical strapdown mechanisation of an inertial navigation system (left: acceleration measurements; right: absolute position estimate)

By using only gyroscopes there is actually no possibility to bound the drifting error for the orientation in a reasonable way. At this point it is necessary to use other information channels. The general idea for compensating the drift error of the gyroscopes is based on using the accelerometer as an additional attitude sensor for generating redundant information. Due to the fact that the 3-DoF accelerometer measures not only (external) translational motion, but also the influence of the gravity it is possible to calculate the attitude based on the single components of the measured acceleration. This is of course only true if no external force is accelerating the sensor. So there are two questions which have to be answered:

1. How it is possible to calculate the attitude from accelerometer measurements?
2. How external translational motion can be handled?

Both problems can be solved by following a two-stage switching behaviour inspired by work presented in Rehbinder and Hu [2004]. At this point it should be pointed out that measurements from the accelerometers can only provide roll and pitch angle and the heading angle has to be derived by using the magnetometer instead.

The following figure gives an illustration about the geometrical relations between measured accelerations due to gravity and the roll and pitch angle of the attitude. By this it follows that the angles can be determined by following relations:

\[
\theta = \arctan\left(\frac{a_x^2}{\sqrt{(a_y + a_z)^2}}\right)
\]  

(6)

\[
\phi = \arctan\left(\frac{a_y^2}{\sqrt{(a_x + a_z)^2}}\right)
\]  

(7)

The missing heading angle can be recovered by using the readings from the magnetometer and the already determined roll and pitch angles. Here it is important to consider that the measured elements of the earth magnetic field have to be transformed to the local horizontal plane (tilt compensation). Figure 8 is indicating the corresponding relations as shown in Caruso [2000]:

\[
X_h = m_x \cdot c\varphi + m_y \cdot s\theta \cdot s\varphi - m_z \cdot s\theta \cdot s\varphi
\]

\[
Y_h = m_y \cdot c\theta + m_z \cdot s\theta
\]

\[
\psi = \arctan2(Y_h, X_h)
\]  

(8)

Based on these findings a discrete Kalman filter bank (DKF-bank) is implemented which is responsible for the estimation of all three angles of the system. For the pitch and the roll angle the same DKF-architecture is used, as indicated in Figure 9-(a). In comparison to that the heading angle is estimated by a alternative architecture as shown in Figure 9-(b).
All DKFs are mainly based on the classical structure of a Kalman filter (see Bishop [2007]) which consists of a first prediction of states and subsequent correction, where the two states are the unknown angle \( \xi \) and the bias of the gyroscope \( b_{\text{gyro}} \). The Kalman filtering itself is composed from the following classical steps, where the following descriptions are simplified to a single angle \( \xi \).

5.1. Computation of an a priori state estimate \( \vec{x}_{k+1}^- \)

As already mentioned the hidden states of the system are \( x = [\xi, b_{\text{gyro}}]^T \). The a priori estimates are computed by following the following relations:

\[
\begin{align*}
\hat{\omega}_{k+1} &= \omega_{k+1} - b_{\text{gyro}}
\xi_{k+1} &= \xi_k + \int \hat{\omega}_{k+1} dt \\
\end{align*}
\]

(9)

Here the actual measurements from the gyroscopes \( \omega_{k+1} \) are corrected by the actually estimated bias \( b_{\text{gyro}} \) from the former iteration, before the actual angle \( \xi_{k+1} \) is computed.

5.2. Computation of a priori error covariance matrix \( P_{k+1}^- \)

The a priori covariance matrix is calculated by incorporating the Jacobi matrix \( A \) of the states and the process noise covariance matrix \( Q_K \) as follows:

\[
P_{k+1}^- = A \cdot P_k \cdot A^T + Q_K
\]

(10)

The two steps 1) and 2) are the elements of the prediction step as indicated in Figure 9.

5.3. Computation of Kalman gain \( K_{k+1} \)

As a prerequisite for computing the a posteriori state estimate the Kalman gain \( K_{k+1} \) has to be determined by following Equation 11.

\[
K_{k+1} = P_{k+1}^- \cdot H_{k+1}^T \cdot \left( H_{k+1} \cdot P_{k+1}^- \cdot H_{k+1}^T + R_{k+1} \right)^{-1}
\]

(11)

5.4. Computation of a posteriori state estimate \( x_{k+1}^+ \)

The state estimate can now be corrected by using the calculated Kalman gain \( K_{k+1} \). Instead of incorporating the actual measurements as in the classical Kalman structure the suggested approach is based on the computation of an angle difference \( \Delta \xi \). The difference is a comparison of the angle calculated from the gyroscope measures and the corresponding attitude as derived from the accelerometers, respectively the heading angle from the magnetometer, as already introduced in the introduction of this chapter. So the relation for \( x_{k+1}^+ \) can be formulated as:

\[
x_{k+1}^+ = x_{k+1}^- - K_{k+1} \cdot \Delta \xi
\]

(12)

At this point it is important to consider the fact that the attitude measurements from the accelerometers are only reliable if there is no external translational motion. For this an external acceleration detection mechanism is also part of the fusion procedure. For this reason the following condition (see Rehbinder and Hu [2004]) is evaluated continuously:

\[
\|\mathbf{a}\| = \sqrt{(a_x^2 + a_y^2 + a_z^2)} = 1
\]

(13)
If the relation is fulfilled there is no external acceleration and the estimation of the attitude from accelerometers is more reliable than the one computed from rotational velocities as provided by the gyroscopes. Noteworthy for real sensors an adequate threshold $\epsilon_g$ is introduced to define an allowed variation from this ideal case. If the camera is not at rest the observation variance for the gyroscope data $\sigma^2_g$ is set to zero. So by incorporating the magnitude of the acceleration measurements as $\|\mathbf{a}\|$ and the earth gravitational field $\mathbf{g} = [0, 0, -g]^T$ the observation variance can be defined by following Equation 14.

$$
\sigma^2_g = \begin{cases} 
\sigma^2_g, & \|\mathbf{a}\| - \|\mathbf{g}\| < \epsilon_g \\
0, & \text{otherwise}
\end{cases} \quad (14)
$$

A similar approach is chosen to overcome the problems with the magnetometer measurements in magnetically distorted environments for the DKF for the heading angle. Instead of gravity $\mathbf{g}$ the magnitude of the earth magnetic field $\mathbf{m}$ is evaluated as shown in the following relation:

$$
\sigma^2_g = \begin{cases} 
\sigma^2_g, & \|\mathbf{m}\| - m_{des} < \epsilon_m \\
0, & \text{otherwise}
\end{cases} \quad (15)
$$

5.5. Computation of posteriori error covariance matrix $\mathbf{P}_{k+1}^+$

Finally the error covariance matrix is updated in the following way:

$$\mathbf{P}_{k+1}^+ = \mathbf{P}_{k+1}^- - \mathbf{K}_{k+1} \cdot \mathbf{H}_{k+1} \cdot \mathbf{P}_{k+1}^- \quad (16)$$

6. Results

as already mentioned our approach was evaluated by using an ABB IRB1400 industrial robot. The $S^3$ was attached to the robot and moved along predefined motion patterns. Thus the ground truth data of the movement is available for a comparison. The tests consider besides the comparison against ground truth also a comparison against other inertial navigation algorithms:

- **Gyrosopes alone (Gyro)** - Here we tested the naive implementation of a simple integration of gyroscope measures as indicated in Equation 4, whereat the initialisation of the starting orientation was computed by using accelerometer and magnetometer measurements.

• **Complementary Filtering (CF)** - The CF approach as suggested by Euston et al. [2008] or Baerveldt and Klang [1997] combines the two information channels (gyroscopes and accelerometers) by using a simple adder, but the two signal sources are filtered before by two complimentary filters. So the accelerometer measurements are filtered by a low-pass filter (here: first-order) and the gyroscope signals by a high-pass filter.

• **Weighting Filter (Est)** - The weighting filter approach as suggested by Bluemel [2010] is a simple straightforward combination of accelerometer and gyroscope measurements by using fixed weights.

For the test different motion patterns were used: rotation around a single axis, consecutive rotation around two axis and simultaneous rotation around two axes. The following subsections summarise the results of the comparison.

6.1. Rotation around a single axis

The first motion pattern contains rotations of the roll/pitch angle as indicated in the following figure. The motion pattern was tested for the roll and pitch angle while the orientation estimation was computed by using the suggested method based on a bank of Kalman filters and the three naive methods described above. All results were tested against the ground truth, thus an absolute error angle was computed for all the algorithms. Figure 11 and Figure 12 show the results of this test for the roll and pitch angle. Here $Rx$ indicates the weighting filter, $gyro$ the naive integration of rotational velocities, $CF$ the complimentary filtering and $KF$ the Kalman filter approach.

![Figure 10: Motion pattern for roll/pitch angle](image-url)
The typical drifting behaviour of the gyroscope measures can be directly identified in the orientation estimates delivered only by gyroscope measures. The suggested KF approach outperforms the other filtering methods.

6.2. Consecutive rotation around two axes

The second motion pattern contains a rotation of 90° around the roll-axis and a consecutive rotation of 90° around the yaw-axis. The following Figure 13 gives an impression about the performance of the different filtering strategies for this kind of motion. It can be seen that especially the CF approach got immense problems during the times of motion. Also for this test the KF approach delivers the best results, but the simple weighting approach delivers comparable results but with less computational complexity. The gyroscopes alone show the same drifting results as for the previous experiments.

6.3. Simultaneous rotation around two axes

Finally we tested the motion pattern with a simultaneous movement around two axes. The results are summarised in Figure 14, whereat again a comparison against the other methods was carried out.

The suggested KF approach shows the best result in terms of accuracy and long-time stability.

7. Conclusion and future work

It was shown that the suggested approach which utilises a bank of Kalman filters is able to outperform other classical methods for orientation estimation. In this context it was proved that the usage of a smart sensor system containing a sensor array, signal conditioning devices and a sensor fusion scheme is able to deliver reliable information about a cameras pose. This information can be fed into classical computer vision algorithms as an aiding modality for camera egomotion estimation.

Future work will consider mainly possibilities for position estimation based on the same MEMS-based sensor array and the combination of inertial and visual measurements. We already proposed a framework for a visual-inertial system for scene reconstruction in Außerheide and Krybus [2011]. In this context a parallel fusion network was suggested in Außerheide
Figure 14: Comparison of different filtering techniques for simultaneous motion pattern - (a1): Orientation estimates for weighting filter; (a2): Absolute error for weighting filter; (b1): Orientation estimates for CF; (b2): Absolute error for CF; (c1): Orientation estimates for KF; (c2): Absolute error for KF.

Acknowledgements

The authors acknowledge support from all other fellows and students at the Laboratory for Image Processing Soest (LIPS) and the Institute for Computer Science, Vision and Computational Intelligence, especially Mr. Dominik Bluemel for his participation in the project. Furthermore we like to thank all people at the Electronics Department of the School of the Built Environment and Engineering at the University of Bolton for their kind support.

Authors Biographies

Dominik Aufderheide is an active researcher in the area of multi-sensor image processing and computer vision. Currently he is a research fellow at the Institute of Computer Science, Vision and Computational Intelligence (CV&CI), where he is working towards a Ph.D. in cooperation with the University of Bolton.

Werner Krybus is a professor for data systems engineering and signal processing at South Westphalia University of Applied Sciences. He is founder of the Laboratory for Image Processing Soest within the Institute for Computer Science, Vision and Computational Intelligence. His primary research interests include embedded systems, computer vision and sensor fusion.

Dennis Dodds is an academic teaching fellow at the University of Bolton. He graduated with a Ph.D. at the University of Salford and gained practical experience in a position as a research engineer at Ferranti Semiconductors Ltd. His research interests can be summarised as non-contacting measurement systems, biomedical and current-mode analogue electronics.

References


C. Harris, M. Stephens, A combined corner and edge detector, volume 15, Manchester, UK, pp. 147–151.


I. Skog, P. Händel, Calibration of a MEMS inertial measurement unit, in: XVII IMEKO World Congress on Metrology for a Sustainable Development, Rio de Janeiro, Brazil.


G. Aslan, A. Saranli, Characterization and Calibration of Mems Inertial Measurement Units, EURASIP.


M. Caruso, Applications of magnetic sensors for low cost compass systems, pp. 177 –184.


