Global Pose Estimation using Multi-Sensor Fusion for Outdoor Augmented Reality

Gerhard Schall [1]
TU Graz, Austria

Daniel Wagner [2]
TU Graz, Austria

Gerhard Reitmayr [3]
Cambridge University, UK

Elise Taichmann [4]
TU Graz, Austria

Manfred Wieser [5]
TU Graz, Austria

Dieter Schmalstieg [6]
TU Graz, Austria

Bernhard Hofmann-Wellenhof [7]
TU Graz, Austria

ABSTRACT

Outdoor Augmented Reality typically requires tracking in unprepared environments. For global registration, Global Positioning System (GPS) is currently the best sensing technology, but its precision and update rate are not sufficient for high quality tracking. We present a system that uses Kalman filtering for fusion of Differential GPS (DGPS) or Real-Time Kinematic (RTK) based GPS with barometric heights and also for an inertial measurement unit with gyroscopes, magnetometers and accelerometers to improve the transient oscillation. Typically, inertial sensors are subjected to drift and magnetometer measurements are distorted by electro-magnetic fields in the environment. For compensation, we additionally apply a visual orientation tracker which is drift-free through online mapping of the unknown environment. This tracker allows for correction of distortions of the 3-axis magnetic compass, which increases the robustness and accuracy of the pose estimates. We present results of applying this approach in an industrial application scenario.

KEYWORDS: handheld augmented reality, inertial tracking, visual tracking, kalman filter, sensor fusion

INDEX TERMS: H.5.1 [Information Systems]: Multimedia Information Systems—Augmented Reality; I.4.8 [Image processing and computer vision]: Scene analysis—Motion, Photometry, Sensor fusion, Tracking;

1 INTRODUCTION

Augmented Reality (AR) is increasingly gaining acceptance outside of research labs in various application domains, such as industrial planning, utility maintenance or engineering.
One important enabling factor is the evolution of AR setups towards ergonomically and socially more acceptable form factors, such for example handhelds and mobile phones. A second crucial enabling factor is accurate and robust global pose estimation for delivering high-quality registration and overlays.

[1] email: schall@icg.tugraz.at
[2] email: wagner@icg.tugraz.at
[3] email: gr281@cam.ac.uk
[4] email: elise.taichmann@TUGraz.at
[5] email: wieser@geomatics.tu-graz.ac.at,
[6] email: schmalstieg@icg.tugraz.at,
[7] email: hofmann-wellenhof@TUGraz.at

For global outdoor registration, GPS is currently the best sensing technology, but its precision, update rate and quality deteriorates significantly in urban environments. We have designed and developed a hardware tracking module using Differential GPS (DGPS) or Real-Time Kinematic (RTK) based GPS. This hardware tracking module is suited for use for handheld AR devices due to its small weight and form factor. To obtain orientation estimates Inertial Measurement Units (IMU), combined with a magnetic compass are often used. To improve the orientation estimates we implemented a second Kalman filter that is able to estimate the orientation, velocities, accelerations and sensor biases by processing measurements obtained from gyroscopes (angular velocities), accelerometers (linear accelerations) and magnetometers (magnetic field). This filter eliminates the rather long transient oscillation behavior of the inertial sensor.

The simultaneous requirements of real-time performance, robustness and low latency are not met by a single approach alone and require a solution that combines the strengths of different approaches. For example, the magnetometer is very susceptible to electromagnetic interference in outdoor environments, thus leading to deviation. Magnetic deviation is the error induced in a compass by local magnetic fields such as variations caused by mountains, iron ore deposits, etc or caused by electric circuits.

A convincing tracking solution must overcome inherent limitations of individual techniques by combining different complementary methods. This work takes the common approach of combining visual trackers with inertial sensors. While the orientation information from the IMU and the position from GPS only provide complementary measurements of the camera pose, the video stream encodes relative motion information about both translation and rotation.

For compensation of the drift of the inertial sensor and of the magnetic deviation effect induced by electro-magnetic influences, we additionally apply a drift-free, deviation-free visual tracker that allows for online learning of natural features. By online mapping of the unknown environment, this tracker detects and corrects the deviation of the 3-axis magnetic compass. This visual tracker improves both accuracy and robustness of the rotation estimation.

Natural feature detection and tracking are an essential research area for AR. One promising research direction in visual tracking is model-based tracking. But these tracking systems rely on models of natural features such as architectural lines or feature points extracted from reference images. By contrast, the approach presented here does not use any models of the environment. The features to be tracked by our visual panorama tracker are dynamically determined at runtime. Still occlusions may provide wrong motion cues and generate a failure in the tracking system.

For example, a pedestrian or large vehicle crossing the field of view can induce a motion in the visual tracker filter because the tracked features are occluded. A robust system should detect and recover from such disturbances. The visual tracker creates a map of features of the environment allowing the tracker to operate in a drift-free manner and to also immediately re-initialize and continue tracking, when the moving object left the field of view.

The first contribution of this paper is the use of complex GPS and IMU data for global pose estimation. We use a differential GPS/IMU hardware module which is combined with barometric height measurements in a Kalman filter to improve the accuracy of the user's 3D position estimate. The second contribution of this paper comprises the combination of the filtered inertial tracking with a drift-free visual panorama tracker that allows for online learning of natural features. By detecting and correcting of the deviation of the magnetometer housed in the inertial sensor, this visual tracker improves both accuracy and robustness of the rotation estimation.

The paper is organized as follows. Section 2 discusses related work. In section 3 the used AR hardware platform is introduced. Section 4 highlights both the Position Kalman filtering and Attitude Kalman filtering approach, explains the visual panorama tracker and furthermore details the fusion of GPS, inertial and vision measurements. The high stability and accuracy of this multi-sensor fusion approach is demonstrated in a practical location-based application (see section 5) and evaluated against ground truth based on map data of the test site. Section 6 draws a final conclusion.

2 Related work

Outdoor augmented reality requires accurate tracking of 6D pose in unprepared environments. No single sensor provides all 6 degrees of freedom while being robust, accurate and covering a large environment. Therefore a combination of sensors is usually used, with the most popular being GPS for global position and magnetic compasses and inertial sensors for orientation [6][7][13][16].

However such systems are severely limited by the accuracy and update rate of GPS. While GPS can deliver good performance in open areas, its accuracy typically degrades severely in urban environments. Urban canyons of dense buildings block the line-of-sight to satellites and produce multi path signals giving wrong distance measurements [15]. Additionally, foliage and roof structures attenuate signals leading to drop outs in position updates. Our system improves on existing work through using a custom DGPS/RTK receiver and fuses its output with barometric height information.

For rotation estimation a magnetic compass provides a convenient absolute orientation with respect to the Earth's magnetic field. However, local electro-magnetic fields from industrial installations and general infrastructure can severely distort the field and thus produce large offsets in the estimated orientation. Better devices cannot compensate for this source of error and the resulting deviation needs to be estimated [11] with a second, redundant sensor.

Possible choices for additional sensors are gyroscopes yielding an estimate of rotational velocity that can be integrated to give relative rotation, linear accelerometers yielding an estimate of gravity or computer vision-based tracking [17][22]. Gyroscopes suffer from drift through the integration step, thus creating their own offset in the orientation estimation which requires some manual calibration to remove [11][12].

Computer vision-based tracking can provide either relative measurements through optical flow or absolute measurements if a model of the observed environment is known [1][2][17]. In this work we chose a drift-free tracking method that estimates camera orientation with respect to a map of the environment which is created online [4]. As the map is not changed after initial creation, no drift is incurred. However, without any additional input the absolute orientation of this map with respect to the environment is unknown.

3 Hardware platform

We developed a handheld AR platform that is build around a Sony VAIO UX ultra-mobile PC with 1.06 GHz Pentium CPU based on the work of Kruijff [14] (see Figure 1). Alternatively, we employ the rugged UMPC Panasonic U1 featuring sunlight-viewable touch screen and sealed all-weather design for demanding outdoor use. The handheld platform is equipped with various sensors. A UEye USB 2.0 camera using a 4.2mm wide-angle lens provides the video-background and delivers the video frames, which are used as input for the visual tracker. The camera captures video frames with a resolution of 640×480 at 30 Hz. A 3DOF inertial sensor (XSens MTi-G, built-in GPS receiver is not used) is mounted at the encasing at the back of the AR platform. Today's high-tech GNSS receivers combine two standards (GPS, GLONASS) on two frequency bands (L1 and L2 for GPS and G1 and G2 for GLONASS). We use a Novatel OEMV-1 L1 DGPS/RTK receiver with an external antenna, which is eliminating multi-path signals. Furthermore, a 3G modem is used for connecting to the internet to receive NTRIP (Networked Transport of RTCM via Internet Protocol) GPS correction data via a serial connection from nearby reference stations. This continuous data stream is used to achieve positional accuracies in the sub-meter to centimeter range.

Figure 1: Handheld outdoor AR platform.

A seven-port USB hub connects all sensors with the USB port of the UMPC. The energy demands of these sensors together led us to use a special Lithium polymer battery (3200mAh), that supplies the GPS receiver and via a voltage transformer the USB hub too.

Figure 2 depicts the tracking module, which has been designed towards low kinematic and high rotational movements. This is optimally suited for pedestrian outdoor user's inspecting an environment with the handheld AR device. The update rate of the

Figure 2: Novatel OEMV-1 receiver (DGPS/RTK) (11x6x3cm) placed in the middle of connectors, antenna and battery (left). Side view of the receiver (right).

tracking module consists of the of inertial sensors update rate of 25 Hz and the GPS update rate. The NTRIP GPS correction data is received every second while the GPS receiver itself has a higher update rate. The system is able to perform at around 20 Hz using this configuration. Only a few mobile AR systems have been built, which are using differential GPS/RTK tracking. For example, a mobile system was built by Höllerer et al. [9] or Piekarski and Thomas [16]. However, these systems use rather bulky hardware and limited sensor fusion and are therefore not suited for handheld AR. To the knowledge of the authors the described device is the only handheld AR system with integrated DGPS/RTK tracking support and sensor fusion support.

4 GLOBAL POSE ESTIMATION USING MULTI-SENOR FUSION

The overall tracking framework uses Kalman filtering with a constant velocity model for fusion of DGPS/RTK with barometric heights and uses an IMU with gyroscopes, magnetometers and accelerometers to improve the transient oscillation. In the following, this approach is described in more detail. For compensation of environmental electro-magnetic influences, we additionally apply a drift-free visual tracker. By online mapping of the unknown environment, this tracker allows for detecting and correcting the deviation of the 3-axis compass, which increases the robustness and accuracy of the pose estimates.Figure 3 shows the multi-sensor fusion system architecture. Since the considered application domain shows only little kinematic motion, using GPS for supporting the orientation estimates is not useful. Considering this, position and attitude is optimized in two separate Kalman filters. A dedicated Kalman filter component for position estimation is complemented with an Attitude Kalman filter for orientation estimation. To allow for correction of both deviation and bias the visual panorama tracker is combined with the Attitude Kalman filter using a Finite State Machine. To fulfill the high requirements of the application scenario concerning positional accuracy, a DGPS/RTK receiver is employed using differential corrections from the Austrian Positioning Service (APOS). The correction data from the reference station is delivered to the handheld device in RTCM 2.3 format and thereby reduce influences such as ionospheric or tropospheric effects. This way the accuracy of the position estimation can be improved significantly [20].

Data transmission is done via a 3G modem connection. A special software module was developed handling the dial-in procedure, the data routing, the data conversion and the data transfer to the GPS receiver. If the 3G connection is lost, the software module reconnects automatically.

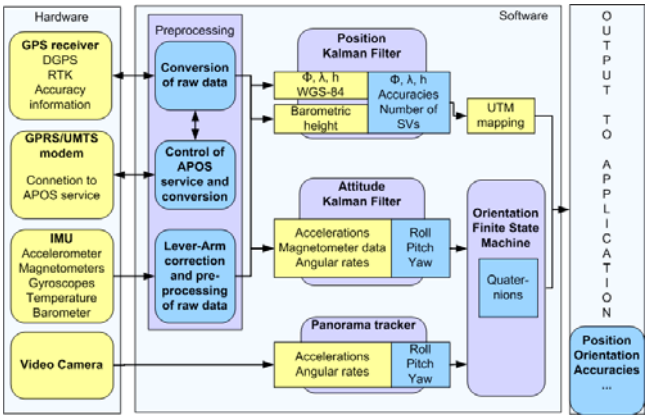

Figure 3: Multi-sensor fusion system architecture.

Moreover, lever-Arm correction is performed, the data is prepared and the GPS receiver uses this data for calculating a DGPS/RTK position estimate. Next, the Position Kalman filter fuses the position estimate with the barometric height for the final position estimate, which is transformed into Universal Transverse Mercator (UTM) format.

Raw data of the accelerometers, gyroscopes and magnetometers are preprocessed and converted. Then the Attitude Kalman filter fuses the delivered data resulting in roll, pitch and yaw as output. Two effects occur in combination with the attitude. First, the magnetic yaw is afflicted with a deviation. Second, the angular velocities of the gyroscopes show a bias, which results in a drift of the angles. The bias of the gyroscopes is considered and corrected by Kalman filter estimation.

The magnetic yaw is deducted from the 3-axis magnetometer and refers to compass north. While the variation can be corrected, the deviation effect represents an unknown location-dependent term. Magnetic yaw and the angular rate of yaw of the gyroscope should support each other. Both have a variable term namely deviation and bias. Consequently, the simultaneous estimation of the deviation and bias of yaw cannot be determined within the Attitude Kalman filter.

Only an additional input for yaw, which is without drift and without bias, would allow solving this problem. To solve this problem and achieve to also estimate the deviation of the magnetometer yaw, we introduce and apply a visual panorama tracker as additional input. By online tracking natural features and simultaneously mapping the environment, the visual tracker delivers drift-free and unbiased orientation estimates. At the beginning, the visual panorama tracker is initialized with the pitch of the inertial tracker, so that the user does not need to hold the AR device horizontally at startup. Furthermore, the visual panorama tracker is able to use the motion model of the inertial sensor to provide more accurate priors under fast motions. Classically, inertial sensors are better suited for measuring high-frequency and rapid motion while the slower vision sensor performs best with low frequency motion and provides absolute references to reset the error. This multi-sensor fusion approach allows for detection and correction of both drift and bias of the inertial sensor.

4.1 Position Kalman filtering approach

To obtain the required position accuracy in real-time, GPS has to be used either in differential positioning mode (DGPS) or in relative positioning mode (RTK). Due to the measurement

principle, the vertical GPS-accuracy (heights) is in general two times worse than the horizontal GPS-accuracy. After initializing the barometer (using GPS-heights as reference), the barometric height is more stable than GPS heights especially in the kinematic mode. Additional position information helps in overcoming GPS shadowing in urban regions. Therefore the Position Kalman filter performs a sensor fusion between GPS and barometer. The GPS height and the barometric height are combined in the filter with respect to their accuracies. Accurate measurements get higher weights. During the initialization step, the barometric height gets a small weight in such a way that the filtered height is exclusively affected by the GPS height. The equations of the Kalman filter are well-known and are therefore not repeated here [10]. The filter input consists of:

- GPS position (geographical coordinates)
- Barometric height (offset corrected)

The **observation vector** z and the state vector x are defined as followed.

$$z = \begin{bmatrix} \phi_{GPS} & \lambda_{GPS} & h_{GPS} & h_{BARO} \end{bmatrix}^T$$
$$x = \begin{bmatrix} \phi & \lambda & h & \dot{\phi} & \dot{\lambda} & \dot{h} \end{bmatrix}^T \quad (1)$$

The barometric height is used as absolute height after correcting for the barometric offset. The barometric offset is estimated at the beginning. During the offset estimation the barometric height is not integrated in the filtering.

The **observation equations** are linear and stated below.

$$\begin{aligned} \phi_{GPS} &= \phi \\ \lambda_{GPS} &= \lambda \\ h_{GPS} &= h \\ h_{BARO} &= h \end{aligned} \quad (2)$$

The application which is covered in this paper implicates little dynamics. Therefore it is sufficient to consider a uniform motion for the dynamic model. In addition to the geographical coordinates itself, also their velocities have to be estimated in order to perform prediction. The dynamic model simply updates the position through integrating the velocities.

4.2 Attitude Kalman filtering approach

The Attitude Kalman filter performs a sensor fusion of gyroscopes, accelerometers and magnetometer. Pre-processed quantities of these sensors form the filter input and are described in detail later on:

- Roll and pitch derived from triaxial accelerometers
- Magnetic yaw
- Gyroscopic angular rates of roll, pitch and yaw

The **observation vector** z and the state vector x are defined in the following way.

$$z = \begin{bmatrix} \varphi_{ACC} & \vartheta_{ACC} & \psi_{MAG} & \dot{\varphi}_{GYR} & \dot{\vartheta}_{GYR} & \dot{\psi}_{GYR} \end{bmatrix}^T$$
$$x = \begin{bmatrix} \varphi & \vartheta & \psi & \dot{\varphi} & \dot{\vartheta} & \dot{\psi} & \ddot{\varphi} & \ddot{\vartheta} & \ddot{\psi} & b_\varphi & b_\vartheta & b_\psi \end{bmatrix}^T \quad (3)$$

Gyroscopic angular rates have biases which are estimated in this filter to avoid a temporal drift of the attitude angles. The linear **observation equations** are the following.

$$\begin{aligned} \varphi_{ACC} &= \varphi \\ \vartheta_{ACC} &= \vartheta \\ \psi_{MAG} &= \psi \end{aligned} \quad (4) \qquad \begin{aligned} \dot{\varphi}_{GYR} &= \dot{\varphi} + b_\varphi \\ \dot{\vartheta}_{GYR} &= \dot{\vartheta} + b_\vartheta \\ \dot{\psi}_{GYR} &= \dot{\psi} + b_\psi \end{aligned} \quad (5)$$

For a sufficiently small time interval in accordance with the measurement update rate (here 25 Hz), the biases are assumed to be constant. A uniform angular acceleration of the attitude angles is assumed for the dynamic model. Therefore, the angular rates and the angular accelerations of the attitude angles have to be estimated within the Kalman filter. The filtered magnetic yaw is still affected by magnetic deviation and magnetic declination.

Gyroscope measurement model

A triaxial gyroscope measures raw angular rates along the input axis in the body frame (BF).

$$\omega^{BF} = \begin{bmatrix} \omega_x^{BF} & \omega_y^{BF} & \omega_z^{BF} \end{bmatrix}^T \quad (6)$$

These measurements can be converted into angular rates of roll, pitch and yaw. The angular rate of roll corresponds directly to the measured angular rate along the x-axis.

$$\dot{\varphi} = \dot{\varphi}\left(\omega_x^{BF}\right) = \omega_x^{BF} \quad (7)$$

The angular rate of pitch is not only a function of measured angular rates along the y- and z-axis but also a function of the roll angle. It can be derived as followed.

$$\dot{\vartheta} = \dot{\vartheta}\left(\omega_y^{BF}, \omega_z^{BF}, \varphi\right) = \omega_y^{BF} \cdot \cos\varphi - \omega_z^{BF} \cdot \sin\varphi \quad (8)$$

The angular rate of yaw is a function of all the raw angular rates of the tripod and the leveling angles roll and pitch.

$$\dot{\psi} = \dot{\psi}\left(\omega_x^{BF}, \omega_y^{BF}, \omega_z^{BF}, \varphi, \vartheta\right) = -\omega_x^{BF} \cdot \sin\vartheta + \omega_y^{BF} \cdot \sin\varphi\cos\vartheta + \omega_z^{BF} \cdot \cos\varphi\cos\vartheta \quad (9)$$

Magnetometer measurement model

A triaxial magnetometer is needed to derive heading information. Thereby magnetic field strengths along three input axis in the body frame are measured.

$$m^{BF} = \begin{bmatrix} m_x^{BF} & m_y^{BF} & m_z^{BF} \end{bmatrix}^T \quad (10)$$

In this context, the heading information derived from the magnetometer is called magnetic yaw (MY). The magnetic yaw does not equal the true yaw (TY). Due to the fact that a magnetometer corresponds to compass north, we prefer to stay with the predicate ‘magnetic’. The magnetic yaw is infected by magnetic variation (VAR) as well as magnetic deviation (DEV).

$$TY = MY + DEV + VAR \quad (11)$$

The magnetic yaw can be derived, if the leveling angles roll and pitch are known according to Caruso [3].

$$\psi = \psi\left(m_x{}^{BF}, m_y{}^{BF}, m_z{}^{BF}, \varphi, \vartheta\right) = \tan^{-1}\left(\frac{-H_y}{H_x}\right)$$
$$H_x = m_x{}^{BF} \cdot \cos\vartheta + m_y{}^{BF} \cdot \sin\vartheta\sin\varphi + m_z{}^{BF} \cdot \sin\vartheta\cos\varphi$$
$$H_y = m_y{}^{BF} \cdot \cos\varphi - m_z{}^{BF} \cdot \sin\varphi \tag{12}$$

Accelerometer measurement model

The principle to obtain roll and pitch from a triaxial accelerometer according to Groves [8] is called leveling. This method implies that the observed accelerations along the input axis in the body frame are exclusively due to the gravitational acceleration.

$$f^{BF} = \begin{bmatrix} f_x{}^{BF} & f_y{}^{BF} & f_z{}^{BF} \end{bmatrix}^T \tag{13}$$

The sensor must not be affected by additional accelerations. That accounts static measurements. With regard to this application the condition for using this method will be partly fulfilled. In case of sensor movement, the derived leveling angles get lower weight for the filtering. The leveling equations are stated below.

$$\varphi = \varphi\left(f_y{}^{BF}, f_z{}^{BF}\right) = \tan^{-1}\left(\frac{f_y{}^{BF}}{f_z{}^{BF}}\right) \tag{14}$$

$$\vartheta = \vartheta\left(f_x{}^{BF}, f_y{}^{BF}, f_z{}^{BF}\right) = \tan^{-1}\left(\frac{-f_x{}^{BF}}{\sqrt{f_y{}^{BF\,2} + f_z{}^{BF\,2}}}\right) \tag{15}$$

4.3 Visual tracking approach

This subsection describes the implementation of the visual panorama tracker. For compensation of the deviation of the inertial sensor, which is induced by electro-magnetic influences, we additionally apply this visual tracker for detecting and correcting the deviation of the 3-axis magnetic compass. This visual tracker improves both accuracy and robustness of the rotation estimation.

The visual panorama tracker assumes a purely rotational motion, ignoring any translational movement. A pure rotation does not create a parallax effect and hence the environment can be mapped onto a closed two-dimensional surface, such as a cube, sphere or cylinder. This technique is well known in computer graphics under the names environment mapping, reflection mapping or sky-box. Our visual tracker maps the environment onto a cylinder with a height of Π/2 relative to the radius, and can therefore map ~76.3 degrees vertically. Conceptually, the radius does not matter, so for convenience we set it to 1.

Starting with a predefined initial direction, the tracker maps the environment on the fly, while the camera is moving. A similar technique has been presented by DiVerdi [5]. However, our mapping and tracking technique is much more efficient: DiVerdi's approach requires intensive GPU processing to run in real-time, whereas our approach runs in real-time with minimal memory and CPU resources only. This is a significant characteristic, since makerless systems generally suffer from high computational costs. Compared to high-end PCs, small handheld devices, such as UMPCs have slow CPUs and GPUs. Hence, an efficient solution is mandatory. Our tests showed that the visual tracker requires between 1.5ms and 2.5ms per frame on a notebook with a 2.5GHz CPU and between 4.0ms and 6.5ms per frame on the handheld device. The speed depends on the number of new pixels that are drawn into the map. For a completed map the tracker therefore runs in ~4.0ms per frame on the UMPC, leaving enough processing power for other tasks.

Even though its technical details are entirely different from traditional approaches of simultaneous localization and mapping (SLAM), the basic approach is similar: For each frame, the tracker first estimates a new pose from the camera image and then enters new features into the map. A major difference to classic SLAM systems is that our approach creates a dense map, but entries are not updated, once they are mapped. This is viable, because under a pure rotational motion the whole 2D state of a map feature is directly observable and the motion is sufficiently constrained.The orientation update step uses 2D-2D point correspondences between the environment map and the camera image. The point correspondences are matched using normalized cross correlation (NCC) on warped 8x8 pixel patches. The locations of the interest points are selected in the map using the FAST [18] corner detector in areas of the map that have already been finished. These keypoint locations are then projected into the camera image and searched in the close proximity to yield sub-pixel accuracy. It is important to notice that we do not apply a keypoint detector, such as FAST on the whole camera image, which is one reason for the high speed of our approach.

Once enough correspondences have been found (usually the tracker finds at least a several hundred), the tracker updates the rotation using Gauss-Newton iteration: We basically perform the same optimization as for a full camera pose, but ignore the position and calculate the Jacobians only for the three rotation parameters. An M-estimator is used to deal with re-projection errors. The final 3x3 system is then solved using Cholesky decomposition to yield the update vector for the rotation. Since the starting point is already close to the final solution only few (3-5) iterations are required.

After the rotation has been updated to match the current camera image, we project the camera image into the environment map. We use run-length encoded pixel spans to keep track of which parts of the map have already been mapped and which haven't. Hence, every pixel of the map is filled only once. When a pixel in the map has been selected to be filled, we intersect its 3D cylinder position with the camera image, undistort the image coordinate, use bilinear filtering to extract the pixel intensity and finally correct for vignetting using a simple radial falloff model.

The visual tracker works at a camera resolution of 320x240 pixels. The map is created at a resolution of 2048x512 pixels. For a typical camera field of view of 60 degrees, the camera's resolution is therefore close to the map's resolution: 320 pixels / 60 · 360 = 1920 ≈ 2048 pixels. The angular resolution of the map is therefore 360 degrees / 2048 pixels = 0.176 degrees per pixel with an average error of ~0.002 degrees per pixel (see Section 5.3). This is much more accurate than a gyroscope sensor can deliver. Although errors accumulate over the map, the tracker is inherently free of drift.The visual tracker can perform loop closing to remove the accumulated error. The tracker therefore uses a map that is larger than 360 degrees horizontally (e.g. 405 degrees) in order to create overlapping regions. Once enough overlap is available, the tracker uses template matching on the keypoints in the map to identify the exact overlap. Since no error model is available, we simply scale and shear the map to close the gap.

4.4 Fusion of attitude with visual tracking

We have implemented the integration of the inertial sensors attitude with the visual tracker orientation output using a state machine. This state machine is denoted as SM (Σ, S, S_0, δ) and shown in Figure 4. Σ describes the five input conditions, S is the set of four states, S_0 is the initial state, an element of S, δ is the state-transition function: $\delta : Sx\Sigma \rightarrow S$. Table 1 lists the combinations of states and transition conditions. ΔP (delta pose) denotes the difference between the orientations of visual and inertial tracker.

The state machine starts in state S_0, in which both the inertial and the visual tracker are valid. Despite a fix offset between the orientations of both trackers, the deviation is zero. The visual tracker uses the motion model of the inertial tracker to provide more accurate priors under fast motions. The final orientation is calculated by fusing the orientation of the visual tracker with the orientation of the inertial tracker. We use timestamps for assuring a correct synchronization of both trackers.

A threshold set on ΔP is responsible for deciding whether the two trackers are diverging or not. This threshold is empirically determined and is less than a few degrees. A second threshold, tracking the number of detected features in the image (20 in our experiments), is used for assessing the validity of the visual tracker. If more features than the threshold have been found in the image, the visual tracker can be trusted.

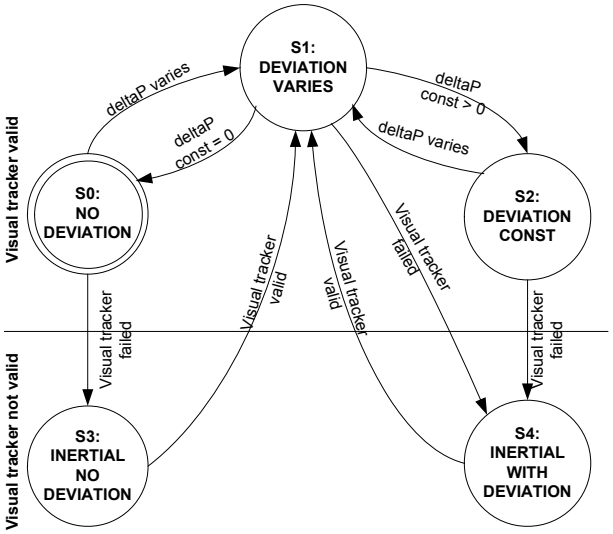

Figure 4: State machine SM {Σ, S_0, S1, S2, S3, S4, δ}.

In case magnetic deviation occurs, the orientation of the inertial tracker will change in respect to the orientation of the visual tracker. As long as the deviation varies, the visual tracker, which is now using its constant velocity motion model, is trusted more and its tracking results are taken as final orientation result (S1). In case the deviation decreases back towards zero, the transition is made to S_0 (e.g. if a transient deviation occurs such as a large vehicle driving by). In case the introduced deviation stays constant and bigger than zero, the state machine changes to state S2. In this state also the visual tracker is trusted, since the inertial tracker is precise, but not accurate now (e.g. a constant deviation

Table 1: State transition table

	Current State (S)				
Condition (δ)	S0	S1	S2	S3	S4
ΔP varies	S1	S1	S1	-	-
ΔP const = 0	S0	S0	-	-	-
ΔP const > 0	-	S2	S2	-	-
Visual tracker failed	S3	S4	S3	S4	S4
Visual tracker valid (re-initialized)	-	-	-	S1	S1

can be induced if the user moves into an area with different magnetic fields or if a magnetic field is introduced around a static user, for example, by switching on electric circuits).Also in state S2 the motion model of the inertial tracker is used for vision tracking.

If the visual tracker fails, which can happen in states S0, S1 and S2, a transition to state S3 or S4 is performed, in which the inertial tracker must be assumed valid, since else no further tracking would be possible until the visual tracker re-initializes. The only difference between the state S3 and S4 is that in the latter state the previous known deviation is taken into account with the inertial measurements. Reasons for the failing of the visual tracker can be an abrupt or very fast rotational motion of the AR device. While the video stream of the camera delivers blurred images, the visual tracker will not find meaningful features. After the rotational motion decreases, the visual panorama tracker cannot be used again until it is re-initialized. Note, that during operation, the visual tracker builds a map of the environment on the fly. In case the part of the environment the camera is facing has already been mapped, the tracker is able to re-initialize immediately to the correct pose. If the current environment is new to the tracker, the orientation values of the inertial tracker need to be considered for re-initialization.

5 RESULTS

We have conducted extensive experiments to test the single tracking solutions separately as well as the overall multi-sensor fusion approach. We tested our system on live video using the hardware presented in section 3. Our application domain is the visualization of underground infrastructures, such as water mains and electricity lines [21]. Outdoor field workers are seeking to locate buried underground assets. The field worker equipped with the AR device walks to the location of interest, which needs to be determined with sub-meter accuracy (not much kinematic movement happens when the user arrives at the location). From this static location, the field worker inspects the underground infrastructure around him by scanning his surroundings with the handheld AR device, resulting in a circular motion.

We have chosen a test site near our campus. For this experiment at the test site we semi-automatically generated an urban 3D model from data from geospatial databases delivered by the local utility company as described in [19]. The urban 3D model (see Figure 5) includes digital elevation model (DEM) of the test site, extruded building footprints, water and electricity lines, pavement border lines, street middle axis as well as surveyed reference points that can act as ground truth data.

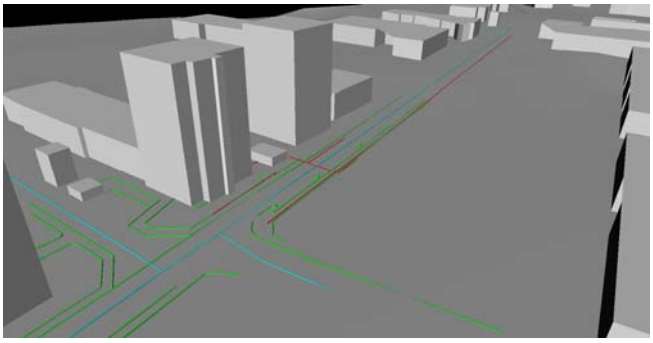

Figure 5: 3D model of the test site. The model includes a DEM, extruded building footprints and underground water and electricity lines.

5.1 Position Kalman filter

This test series uses the Position Kalman filter, which integrates the GPS and barometer sensor. A first test assessed the positional accuracy and precision of the GPS receiver and the APOS service in a static scenario. The comparison in Figure 6 shows the GPS C/A Code solution and the DGPS solution. Qualitatively it is visible that a more stable position can be calculated by using correction signals. Table 2 lists the measurements according to the plots in Figure 6.

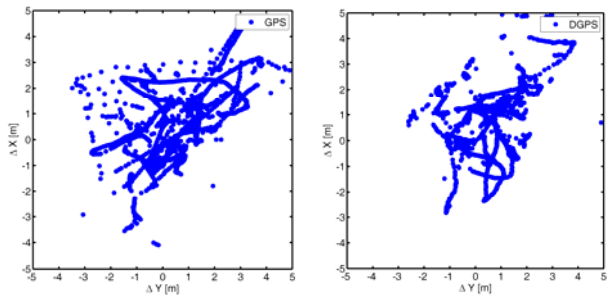

Figure 6: GPS C/A code solution (left), DGPS solution (right).

Table 2: Comparison of GPS and DGPS measurement solutions.

	GPS C/A code		DGPS	
	Mean	Std dev	Mean	Std dev
X	1.058	2.600	0.800	1.848
Y	0.617	1.296	0.765	1.553

Furthermore, a test of the Position Kalman filter was performed in a dynamic scenario, in which the user moves along a path and passes over known reference points, which are drawn in red color (see Figure 7). The test shows that the filtered DGPS position (in blue color) satisfies the accuracy requirements. After one minute, instead of the C/A Code solution, a DGPS position solution with higher accuracy is calculated by the receiver. After five minutes, positions estimates with sub-meter accuracy are calculated by the static receiver. Then, the known reference points were passed over with sub-meter accuracy. It can be observed, that when the user changes the direction of movement, a short post-pulse oscillation appears (see grey ellipsoids).

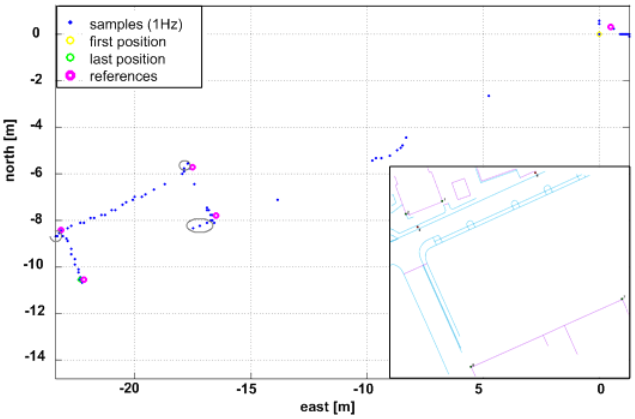

Figure 7: Position estimates along a path using the Position Kalman filter. Gray ellipsoids depict post-pulse oscillation of the filter. Inlay: view of the cadastral map of the test site (compare to 3D model of Figure 5). Blue lines are vector map features of streets and pavements. Violet lines show building footprints with reference points at their corners.

5.2 Attitude Kalman filter

Moreover, we performed a series of tests to assess the accuracy of the inertial sensor as well as the stability and behavior over time of the Attitude Kalman filter, which is described in section 4.2. One test was conducted to observe the attitude during a test of the relative angular accuracy. During this experiment the inertial sensor was fixed on mount that was rotated in steps of 90 degrees between the five measurements of yaw. The duration of one measurement was 15 minutes. Table 3 shows that the relative accuracy of yaw is better than one degree under ideal conditions.

Table 3: Attitude during test of relative angular accuracy.

	Magnetic yaw [deg]		Measured rotation of magnetic yaw [deg]
Deg	Mean	Std dev	
180	178.44	0.12	0
270	269.12	0.09	90.68
360	359.29	0.07	90.17
90	89.08	0.04	89.79
180	178.52	0.17	89.44

Next, we present a selection of interesting results from a series of measurements comparing the output of the inertial sensor with the Kalman filtered output. In the following scenario a user is observed under realistic conditions. In situation (a), the user holds the handheld AR device in her hands while taking the measurements. The following figures show a dataset in which the sensor was experiencing small vibrations from holding it in the hands. Additionally, at a later time the user turned the AR device by 90 degrees. No drift occurs during the test shown in dataset of Figure 8. In the diagrams the drift is marked by green dots. The results of situation (b) in which transient drift occurs, are shown in Figure 9.

Figure 10 depicts the measurements of situation (c) in which permanent drift appears. The accuracy of the filtered yaw stays better than 10 degrees in all 3 situations.

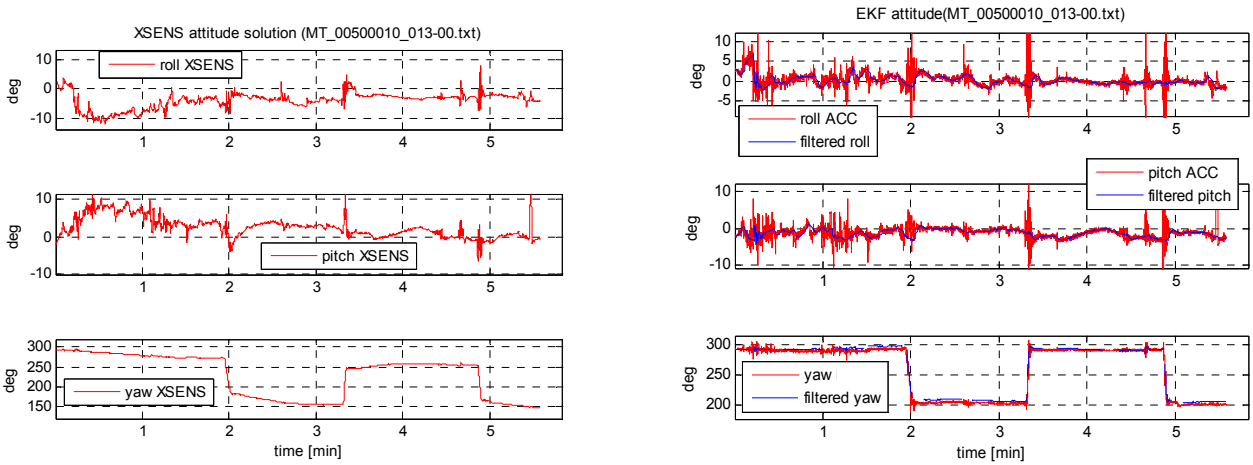

Figure 8: Attitude during situation (a): Rotation of user, no drift (left: inertial sensor, right: Kalman filter attitude).

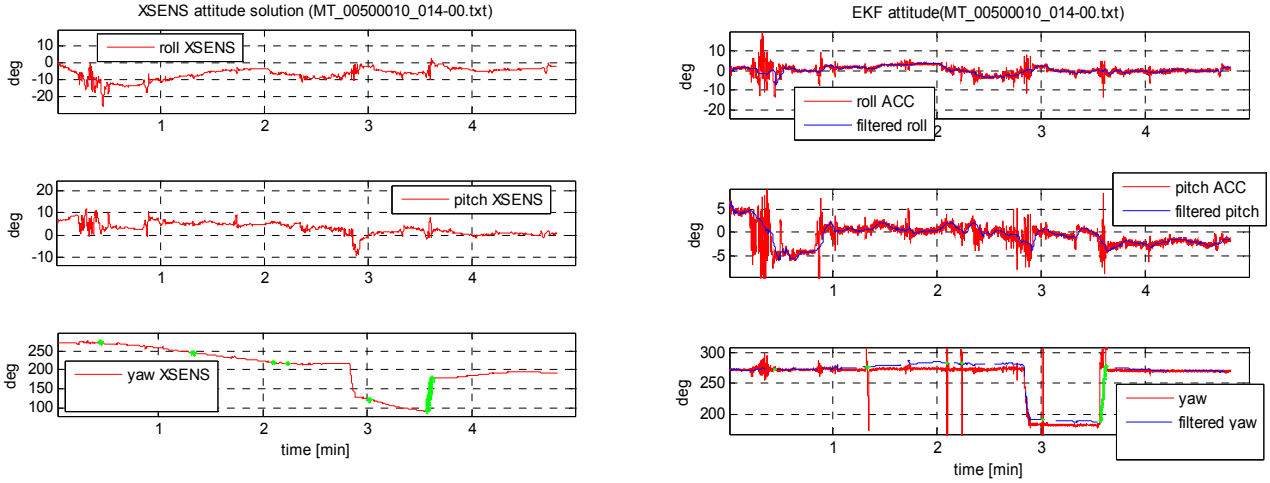

Figure 9: Attitude during situation (b): Rotation of user, transient drift (left: inertial sensor, right: Kalman filter attitude).

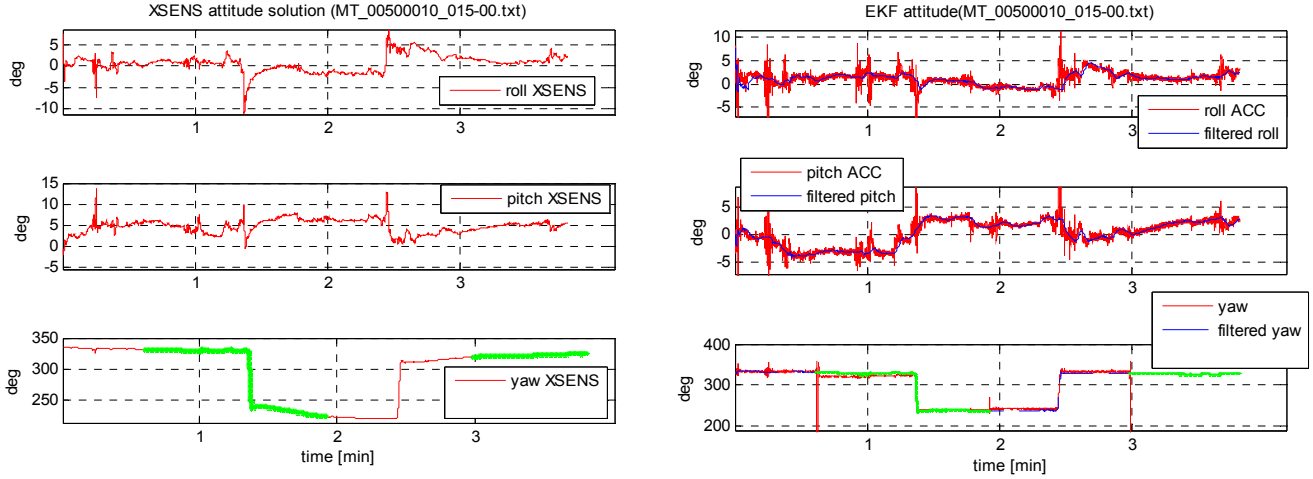

Figure 10: Attitude during situation (c): Rotation of user, permanent drift (left: inertial sensor, right: Kalman filter attitude).

The experiments showed that the accuracy of yaw can be significantly increased because transient and permanent drift can be detected and their influence on yaw can be reduced.

Figure 11 depicts roll, pitch and yaw of the inertial sensor that was assessed in a practical test, which used identical data for the comparison. Again, the AR platform was held in the user's hands

and shivery motions affect the inertial sensor. In this practical test, the AR platform was rotated by 90 degrees in yaw direction.

Results show that the Kalman filter attitude is more stable than the inertial sensor attitude. Especially, at the beginning a large difference between the attitudes occurs due to a slow transient oscillation. According to the manufacturer a lead time of 15 minutes is suggested. Using the Kalman filtered approach, no lead time is necessary and the attitude can be used immediately.

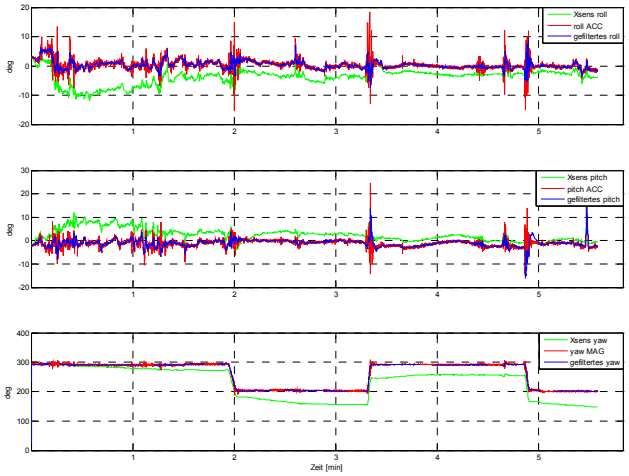

Figure 11: Comparison of attitude (ACC - is derived from accelerometers, MAG is derived from magnetometer)

5.3 Visual tracking

Table 4 lists results from experiments assessing the accuracy of the visual tracker after having calibrated the camera. During the test, the camera, which is using a 4.2mm wide-angle lens, is fixed on a mount and rotated in steps of 30 and 90 degrees to the left and right starting at 0 degrees. At each step the orientation delivered by the tracker was measured for 15 minutes.

Table 4: Yaw measurements of visual tracker.

	Yaw of Visual tracker [deg]		Measured Yaw of Visual
Deg	Mean	Std dev	tracker [deg]
-120	-118.68	0.11	0
-90	-89.20	0.07	29.48
0	0.002	0.008	89.202
90	88.96	0.32	88.94
120	118.05	0.17	29.09

A small bias towards underestimating the rotation is present in the visual tracking. This is corrected as soon as a loop-closure happens in the panorama.

5.4 Combination of attitude with visual tracking

This section presents results from applying the multi-sensor fusion system at the outdoor test site with disruptions from electro-magnetic fields. In this scenario, the user went on-site to a location, held the AR platform in her hands and was performing orientation movements. The user's orientation was tracked using the state machine (described in section 4.4), which combines the orientation estimates of both the Kalman filtered inertial sensor and the visual panorama tracker.

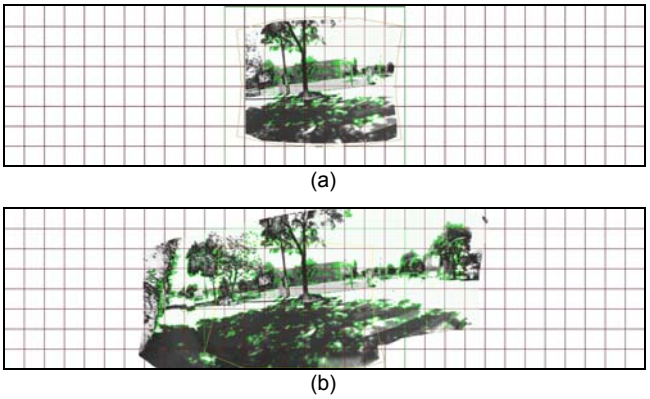

Figure 12: Map of the outdoor environment created by the visual panorama tracker. (a) first image used for mapping the test site and calculating the features. (b) after rotational movements the tracker has mapped a larger area of the environment.

Nearly 180 degrees of the horizontal panorama was mapped, since the full length of the image in Figure 12 represents 360 degrees. Simultaneously, the Attitude Kalman filter delivers improved orientation estimates. Figure 13 shows results from the multi-sensor fusion approach, plotting the orientation values from both visual tracking and inertial tracking, together with the state of the fusion state machine.

At the beginning, both trackers are valid and combined for calculating the final orientation. Within frames #25 to #75, the user rotates the handheld device. Both trackers continue to deliver accurate estimates. During frames #120 to about #210 the inertial tracker experiences transient deviation, caused by electro-magnetic influences. Now, the visual tracker provides the final orientation. After the deviation disappeared, both trackers are combined again. At sample #345 the user performs fast, abrupt rotations. Hence, the video camera delivers blurred images, which causes the visual tracker to fail. Now, the inertial orientation is taken as final orientation. The user rotates the handheld device back again and near sample #460 the visual tracker re-initializes and continues tracking. Switching states depends on thresholds and exhibits some noise, but overall performs well.

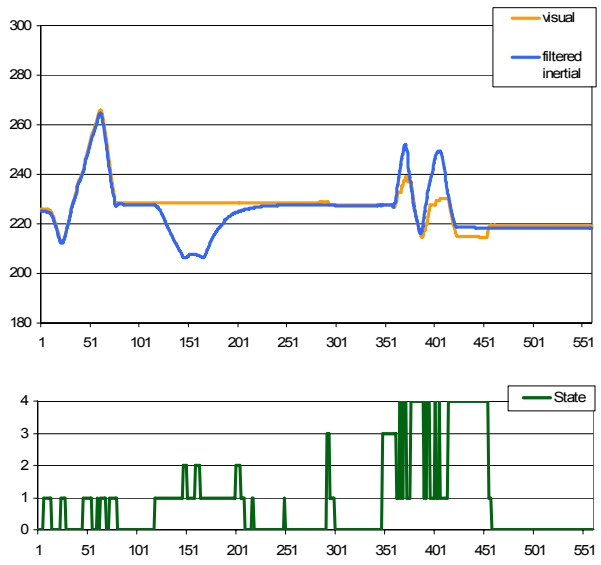

Figure 13: Comparison of yaw of the Kalman filtered inertial sensor with yaw of the visual panorama tracker under various conditions.

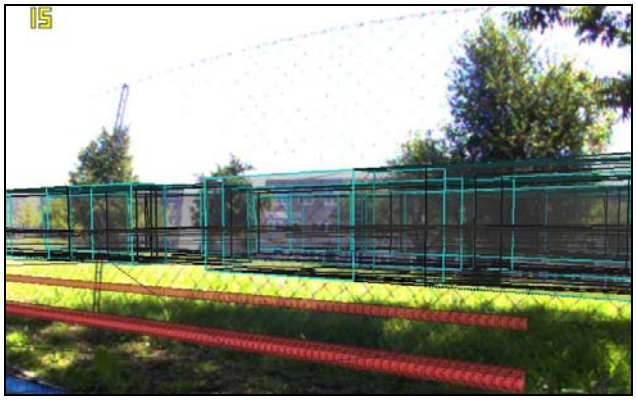

Figure 14: Screenshot taken on AR platform showing underground assets (water mains are blue, electricity lines are in red colour) as well as wire frames of buildings.

The bars underneath the diagram in Figure 13 indicate which tracker is used for calculating the final orientation, which is used for the augmentation. The green bar indicates that both trackers are used. The orange bar stands for visual tracking (V) and the blue bar stands for Kalman filtered inertial tracking (I). Figure 14 gives an example screenshot taken of the application on the AR platform that shows superimposed models of buries assets and wire frame buildings supporting the user in locating these assets.

6 Conclusion and Future Work

We presented a multi-sensor fusion system with integrated inertial and vision tracking technologies for registration of three-dimensional overlays on the real environment. While providing sub-meter accuracy position estimates using differential GPS, the orientation estimates are significantly more robust and accurate.

Inertial tracking has the advantages of range, and a system that is passive and self-contained. Its major disadvantage is its lack of accuracy and drift over time. The first effect of time-depended drift of the accelerometers angular rates is corrected by the Attitude Kalman filter. The second effect of location-depended deviations of yaw can be detected and corrected by using the visual panorama tracker. No model of the environment is needed for the visual tracker. This improves both the accuracy and the robustness of the tracking system and supports well aligned visual overlays to the user.Next steps will include performing a tight GPS/IMU coupling. Future work will also consider the extension of the visual tracker by taking kinematic movements into account.

Acknowledgments

This work was funded by the Austrian Research Promotion Agency (FFG) under contract no. BRIDGE 811000, FIT-IT 820922, POMAR3D ALR-OEWP-CO-417/07, the European Union under contract no. FP6-2004-IST-427571, the EU FP6-2004-IST-4-27731 (PRESENCCIA), and the Austrian Science Fund (FWF) under contract no. Y193 and W1209-N15. Moreover, we would like to express our gratitude to our industrial project partner GRINTEC GmbH and Graz AG - Stadtwerke für kommunale Dienste, Stadtvermessungsamt Graz for providing both real-world test datasets and valuable input to our research activities from the end-user perspective.

References

[1] R. Azuma, J. W. Lee, B. Jiang, J. Park, S. You, and U. Neumann. Tracking in unprepared environments for augmented reality systems. Computer & Graphics, 23(6):787-793, 1999.

[2] G. Bleser, Stricker, D., Advanced tracking through efficient image processing and visual-inertial sensor fusion. In Proc. of IEEE VR 2008, pp. 137-144 2008.

[3] M. J. Caruso, Application of Magnetoresistive Sensors in Navigation Systems, Sensors and Actuators. SAE SP-1220, pp. Feb. 1997.

[4] J. Civera, A. J. Davison, J. A. Magallon, and J. M. M. Montiel. Drift-free real-time sequential mosaicing. Int. Journal of Computer Vision, 81(2):128-137, 2009.

[5] S. DiVerdi, J. Wither, T. Höllerer. "Envisor: Online Environment Map Construction for Mixed Reality," *In IEEE Virtual Reality,* Mar. 8-12, 2008.

[6] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. In Proc. ISWC'97, pages 74, Cambridge, MA, USA, October 13, 1997.

[7] E. Foxlin, Pedestrian Tracking with Shoe-Mounted Inertial Sensors, IEEE Computer Graphics and Applications, p.38-46, Nov 2005.

[8] P. D. Groves, Principles of GNSS, Inertial and Multisensor Integrated Navigation Systems. ARTECH HOUSE INC, Boston London, 2008.

[9] T. Hoellerer, S. Feiner, Tachio Terauchi, Gus Rashid, and Drexel Hallaway. Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system. Computer & Graphics, 23(6):779-785, 1999.

[10] B. Hofmann-Wellenhof, K. Legat, and M. Wieser, Navigation - Principles of Positioning and Guidance. Springer, Wien/New York, 2003.

[11] B. Hoff and R. Azuma. Autocalibration of an electronic compass in an outdoor augmented reality system. In Proc. ISAR 2000, 159-164

[12] X. Hu, Y. Liu, Y. Wang, Y. Hu, and D. Yan. Autocalibration of an electronic compass for augmented reality. In Proc. ISMAR 2005, pages 182-183, Vienna, Austria, October 5-8 2005. IEEE and ACM.

[13] S. Julier, Y. Baillot, M. Lanzagorta, D. Brown, and L. Rosenblum. BARS: Battlefield augmented reality system. In Proc. NATO Symposium on Information Processing Techniques for Military Systems, Istanbul, Turkey, October 9-11 2000.

[14] E. Kruijff., Veas E., Vesp'R – Transforming Handheld Augmented Reality. Proc. 6th IEEE International Symposium on Mixed and Augmented Reality (ISMAR'07), 2007.

[15] G. Lachapelle. GNSS indoor location technologies. Journal of Global Positioning Systems, 11 2004.

[16] W. Piekarski and B. H. Thomas. Tinmith-metro: New outdoor techniques for creating city models with an augmented reality wearable computer. In Proc. IEEE ISWC'01, pages 31- 38, Zurich, Switzerland, 8-9 October 2001.

[17] G. Reitmayr and T. W. Drummond. Going out: Robust tracking for outdoor augmented reality. In Proc. ISMAR 2006, pages 109-118, Santa Barbara, CA, USA, October 22-25 2006. IEEE and ACM.

[18] Rosten, E., Drummond, T., Machine learning for high-speed corner detection. In Proc. of 9th European Conference on Computer Vision (ECCV2006), pp. 430-443, 2006

[19] G. Schall , Schmalstieg D., Interactive Urban Models generated from Context-Preserving Transcoding of Real-Wold Data, Proc. of the 5th International Conference on GIScience, abstracts volume, Park City, Utah, USA, 23.-26. September 2008.

[20] G. Schall, Berglez P., Taichmann E., Kühtreiber N., POMAR 3D project report, Graz 2008.

[21] G. Schall, Mendez E., Schmalstieg D. Virtual Redlining for Civil Engineering in Real Environments", In Proc. ISMAR´08, Cambridge, UK, 15.-18. September 2008.

[22] S. You, U. Neumann, and R. Azuma. Orientation tracking for outdoor augmented reality registration. IEEE Comp. Graph. Appl., 19(6), November 1999