(Positioning Evaluation)^2

Joakim Rydell and Erika Emilsson
Swedish Defence Research Agency, Sweden
joakim.rydell@foi.se, erika.emilsson@foi.se

***Abstract*—This paper presents a camera-based reference system which provides reliable ground truth data in relatively large indoor environments using a minimum of preinstalled infrastructure. The reference system will be used together with a foot-mounted inertial navigation system (INS) with the purpose to evaluate how errors in the foot-mounted INS grow over time. The reference system itself is evaluated and shown to provide sufficient accuracy.**

Keywords: positioning, reference system, camera, markers

I. INTRODUCTION

It is a technical challenge to create a positioning system that is accurate enough in GPS-denied environments, indoors and during harsh electromagnetic interference conditions. A priori information (e.g. building layouts, magnetic field or image information obtained through extensive pre-surveying) or pre-installed infrastructure (e.g. RFID units or WiFi base-stations) is not possible to use in all scenarios. However, a reliable positioning system with seamless outdoor-indoor coverage would increase the safety of military personnel and first responders significantly. Such a system should be light-weight, small, inexpensive and power efficient, and still provide horizontal meter-level accuracy during extended indoor operations. Correct estimates of which floor in a building a firefighter or a soldier is located on is also important [4].

This paper presents a camera-based reference system, which can be used to evaluate other positioning solutions in large indoor environments. This particular system is designed mainly for evaluation of a foot-mounted inertial navigation system. The presented system itself is not intended for navigation in actual scenarios.

First we give a general introduction to foot-mounted inertial navigation and a short description of the system we use. In Section II we present the method of our camera-based positioning system. The evaluation system is evaluated in Section III. Section IV concludes the paper.

A. Foot-mounted inertial navigation

Inertial measurement units (IMUs) deliver measurements of accelerations and angular velocities. The total acceleration is composed by a movement-induced part and the earth gravity force. The velocity is obtained by removing the gravitation effect and integrating the sensed accelerations. A second integration yields the position.

Micro electro mechanical systems (MEMS) based accelerometers and gyros are light-weight, have a low power consumption, and a size small enough to allow integration with soldier and first responder equipment. The performance of triaxial accelerometers and gyros used in MEMS IMUs is continuously improved. However, these sensors are only accurate for non-aided positioning during a few seconds due to large noise contributions and bias drifts (cubic in time) [1]–[3].

If the inertial sensors are mounted on the foot, some sensor bias errors can be estimated and compensated for using knowledge about the foot at stand-still (the stance phase), which gives the IMU zero-velocity. Information about periods of stand-still can be obtained from thresholding the gyro readings, the accelerometer readings or a combination of these measurements. During regular walking this option gives the possibility to update the measurements in the filter approximately once every second. Because the heading error is only weakly observable from the zero-velocity information, the position error will increase (linearly) with the distance. It is also influenced by a random gyro drift and the shape of each traveled path.

Foot-mounted inertial navigation is performed in a local frame (North-East-Down). This requires initialization of position, velocity and heading. The coarse alignment with gravity can give the initial roll and pitch estimates, but the heading (yaw) has to be defined by some other method (magnetometers or GPS can be used when such are available and deemed reliable).

The navigation algorithm in our foot-mounted INS is based on an Extended Kalman Filter (EKF). It estimates the 3D position, velocity and orientation of the IMU, and its error states [7].

We have tested the foot-mounted inertial navigation system with three different IMU sensor types. The 3DM-GX2 from Microstrain and the MTi from Xsens have gyro bias stabilities of around $25°/h$ and an angular random walk of around $2°/\sqrt{h}$. The newer and smaller Microstrain 3DM-GX3 sensors have somewhat better performance.

Except from normal walking there are a number of other possible movements (e.g. sprinting, stair-climbing, crawling). For acceptable position estimates, it is important that the foot-mounted inertial navigation system captures when the foot is at stand-still during all types of motion [4].

II. METHOD

The reference system is expected to provide reliable ground truth data in relatively large indoor environments (e.g. multi-story buildings), and should require a minimum of pre-installed infrastructure. Furthermore, the requirements mentioned above regarding size, weight, etc. also apply to the reference system in order not to alter how the soldier or first responder moves during the system evaluations. A low-cost reference system is desired. These requirements effectively preclude solutions based on GPS (which does not provide sufficient accuracy in most indoor environments), UWB, RFID or ultrasonic ranging (which require extensive installation of base stations, and may be imprecise in some environments due to multi-path propagation [4]) and camera- and/or IMU-based odometry systems (which do not provide long-term accuracy due to drift). Marker-based camera positioning, however, is immune to drift. Instead, the camera position is computed using one or several markers in the field of view. It may be argued that the markers constitute infrastructure, but we consider this acceptable since the installation is easy and inexpensive. It is, however, important that markers are placed at well-known positions, with good accuracy. If sufficiently precise building schematics are available, this should not present a problem. Since the reference system is designed to work together with the foot-mounted INS, a positioning error within 0.1 meter would definitely be considered acceptable. The average distance between the foot-mounted IMU and the camera is significantly larger and will contribute more to the total error.

A. *Marker-based positioning*

General camera-based positioning is a difficult problem, requiring tracking of a large number of landmarks (e.g. corners, lines or other points of interest) and/or recognition of previously visited areas as the camera moves though a scene. By installing markers at known positions, a much easier problem is obtained. In marker-based positioning, the actual localization is performed in two steps:

1) An approximate *global* position is found by determining which (if any) marker is visible. Since the global position of each marker is known, a simple table lookup provides the camera position with an accuracy of a few meters (assuming that the markers are too small to be detected and identified at larger distances).
2) Once a marker is found, more precise *local* positioning is performed by inspection of its appearance in the image. Provided that the camera is reasonably well positioned relative to the marker (with respect to distance and angles), a relative position estimate with good accuracy can be obtained.

The important difference between marker-based positioning and visual odometry-like approaches is that the marker-based methods estimate the position using only the latest image and knowledge about the marker positions. Hence no drift will occur.

Obviously a huge number of markers is required in order to obtain positioning ability throughout an entire building. In the intended application only sparse positioning is necessary; the accuracy of a foot-mounted INS can be evaluated by comparing its estimated position to ground truth at a limited number of positions along a path. (If evaluation was to be performed by following a pre-defined path, not even sparsely located markers would be necessary. Instead, the user could manually indicate when passing through specific locations. This, however, does not provide sufficient flexibility for evaluations in more realistic scenarios.)

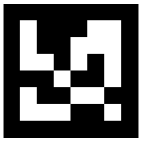


Fig. 1. Example of a marker used by ARToolKitPlus.

Fig. 2. Camera and lens.

ARToolKit is a commonly used library for positioning using visual markers. While originally developed for augmented reality applications, the toolkit also provides functions for estimating the camera position and orientation relative to one or several markers. The pattern within each marker encodes its unique ID, enabling global positioning in an environment with multiple markers. ARToolKitPlus[1] is an extension of the original ARToolKit, with improved pose estimation stability, support for more types of marker identification etc. [8]. A typical marker for use with ARToolkitPlus is shown in Figure 1. We use ARToolKitPlus in the presented work.

B. *Camera system*

The optimal choice of camera depends on the environment in which the system will be evaluated. In our case this environment will mostly consist of corridors and relatively small rooms (e.g. offices and meeting rooms). We therefore need a camera and lens providing a rather large field of view and good light sensitivity. Additionally, since the camera will either be hand-held or mounted on the head or shoulder, the camera and lens should not be too large or heavy. Based on these requirements, we selected the Flea2[2] and a lens providing a horizontal field of view of approximately 80 degrees. The camera and lens are shown in Figure 2. The total size is approximately $3 \times 3 \times 10$ cm.

In order to perform positioning and provide real-time feedback to the user (for example information about whether a marker is close enough for accurate positioning), the camera needs to be connected to a computer running the positioning system. The computer can be carried in a backpack. The positioning system does not require very much processing power, and hence any small laptop computer can be used. For ease of use, the feedback can be provided using headphones.

C. *Evaluation method*

When the camera-based reference system is used to evaluate a foot-mounted INS, a number of checkpoints with markers will be passed several times. This way the error growth rate in the IMU-based system can be measured. This is possible since the error magnitude of the camera-based system is bounded.

---

[1] http://studierstube.icg.tu-graz.ac.at/handheld_ar/artoolkitplus.php

[2] http://www.ptgrey.com/

III. REFERENCE SYSTEM EVALUATION

When developing a reference system which will be used to provide ground truth for evaluating other systems, it is very important to evaluate the performance of the reference system itself thoroughly. This requires *another* reference system, which is known to provide reliable data. The Vicon motion capture system[3] can be used for accurate position measurements. This system utilizes near infrared cameras and strobes along with reflective markers, which limits its applicability to a single room where the system has been installed. Hence, the Vicon system is well suited to evaluating the local positioning performance, i.e. the camera positioning relative to a detected marker. Much simpler methods are sufficient to evaluate global performance (determination of which marker is detected). One such method is presented below.

In the absence of a system providing absolute coordinates, such as the Vicon system, a limited evaluation of local positioning performance can be performed by moving the camera along a known path at a known speed. The experiments reported here are based on this approach.

The next section presents the evaluation of global positioning, i.e. the ability to correctly identify a marker. Section III-B presents the evaluation of local positioning using ARToolKitPlus. This is further evaluated in [5]. In all experiments the marker size was 224×224 mm.

A. Global positioning performance

Since the system computes a (very crude) estimate of its position by identifying markers in the field of view, the global positioning performance is completely determined by the system's ability to detect and identify markers. We have evaluated this by repeatedly moving the camera such that it briefly sees one of a number of markers, and counted how often a marker is correctly identified, incorrectly identified, and missed. This has been carried out at three different distances between camera and markers (1.0, 1.25, and 1.5 meters; typical distances for a scenario where markers are attached to corridor walls) and at different speeds. In all experiments, the camera pointed mostly towards the markers during at least a few frames (this is similar to what can be expected in real use, when a test subject points the camera in the general direction of a marker while passing by). Close to perfect results were obtained: at all reasonable camera velocities, all markers were detected and correctly identified. At very high velocities some markers were missed, probably because of motion blur. No marker was ever incorrectly identified. An incorrectly identified marker would introduce large errors in the evaluation process, while a missed marker only reduces the amount of data available for evaluation slightly.

B. Local positioning performance

The local positioning performance was evaluated by moving the camera at constant speed (except for short acceleration and deceleration periods at the beginning and end of the trajectory) along the $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ axes of the coordinate system defined by the marker. The $\hat{\mathbf{z}}$ axis is orthogonal to the marker plane and points outwards, while the $\hat{\mathbf{y}}$ axis lies in the marker plane. Any other in-plane direction, such as the $\hat{\mathbf{x}}$ axis, should behave similarly. Movement along $\hat{\mathbf{y}}$ was performed at three different distances from the marker: approximately 1, 1.75 and 3 meters. When moving along $\hat{\mathbf{z}}$, the marker was located a few decimeters offset from the optical axis. In both cases the camera was looking along $-\hat{\mathbf{z}}$.

In each experiment, the camera moved 0.67 meters. Figure 3 shows estimated y coordinates when moving along the $\hat{\mathbf{y}}$ axis at the different distances. The first two subfigures contain three plots, each corresponding to one experiment. As these figures show, the position estimation is not perfect, but at distances of 1 and 1.75 meters the errors are below one decimeter. At 3 meters, however, the errors are very large. Hence, measurements where the distance to the marker is larger than a threshold z_{max} (between 1.75 and 3 meters) should not be used for evaluation of other systems. However, the use of a distance threshold requires the ability to accurately estimate z even when the distance is too large to accurately estimate x and y. Without this ability, we would sometimes accept a bad x or y measurement based on a z estimate $\hat{z} < z_{max}$, when the true $z > z_{max}$. Fortunately, the estimates of z are less noisy than x and y when z increases. This is illustrated in Figure 4, which shows the estimates of z obtained when moving in the $\hat{\mathbf{y}}$ direction at a distance of 3 meters from the marker.

Interestingly, some error contributions are similar between different experiments; the three plots in Figure 3a are almost identical. Also, the errors are larger when y is close to zero. Most probably, the latter is explained by poor estimation of the camera orientation when it is located at $x \approx 0$ or $y \approx 0$, and it is looking directly at the marker. In this configuration, the perspective effect is very small, and random perturbations due to noise influence the estimation heavily. The performance of ARToolKitPlus in such difficult configurations is improved compared to the original ARToolKit [6], but some issues seem to remain. It should be noted that the magnitude of these errors is influenced by the light conditions. The shown results were obtained using artificial lighting at night; in brighter and more uniform daylight the errors are smaller.

In our application, the encountered inaccuracies (except at large distances, as shown in Figure 3c) are too small to be of any consequence. If more precision is needed, samples where x and/or y is too close to 0 could simply be discarded.

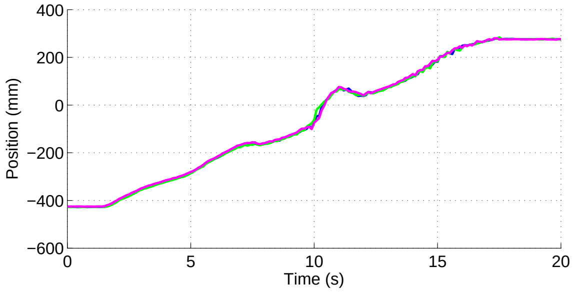
Figure 5 shows the estimated position along the $\hat{\mathbf{z}}$ and $\hat{\mathbf{x}}$ axis when moving 0.67 meters along $\hat{\mathbf{z}}$. As the distance from the marker increases, so does the position noise level, particularly for the x coordinate. Additionally, the noise level while moving (when $2.5 < t < 18$) is considerably higher than when the camera is stationary. However, even the worst noise level (± 5 cm) is within the acceptable range in our application.

IV. SUMMARY

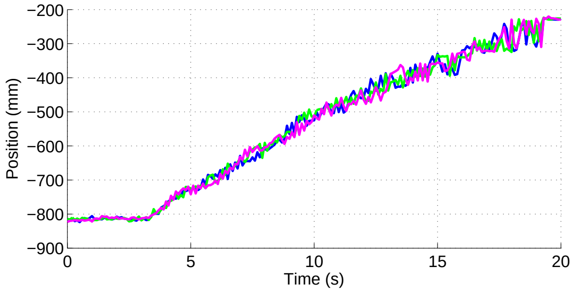
A marker-based reference system for evaluation of indoor positioning solutions has been presented. The reference system has been evaluated and shown to provide adequate performance for use in typical indoor situations. Future work will focus on more extensive evaluation, as well as on further

---

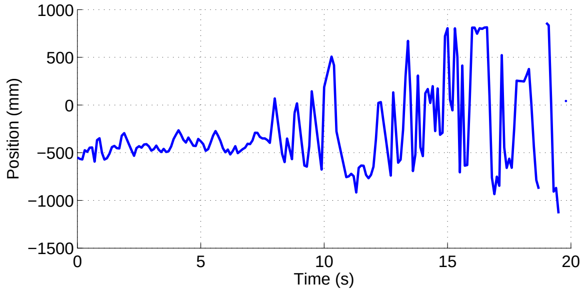
[3] http://www.vicon.com/



(a) $z = 1$ m (estimated motion approximately 0.70 m)



(b) $z = 1.75$ m (estimated motion approximately 0.58 m)



(c) $z = 3$ m (estimated motion undefined)

Fig. 3. Estimated y coordinates when moving along $\hat{\mathbf{y}}$ at three different distances from the marker.

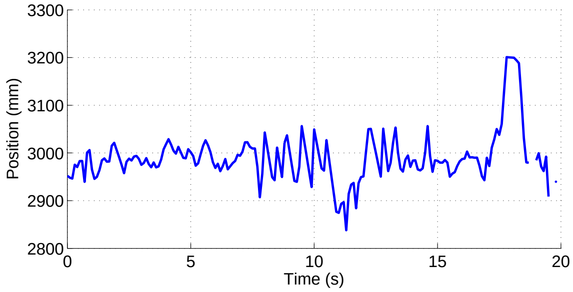
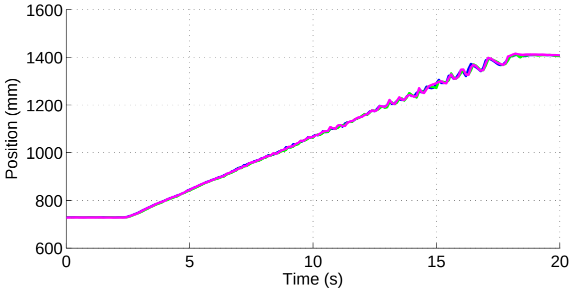
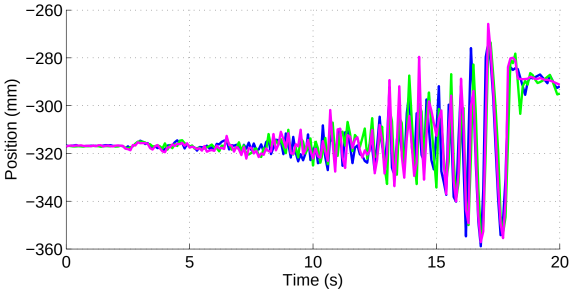


Fig. 4. Estimated z coordinates when moving along $\hat{\mathbf{y}}$ at a distance of 3 meters from the marker.



(a) z coordinate (estimated motion approximately 0.68 m)



(b) x coordinate

Fig. 5. Estimated positions during motion along $\hat{\mathbf{z}}$.

improvements of positioning performance under difficult conditions, such as far from markers and in low-light-conditions.

REFERENCES

[1] E. Foxlin. Pedestrian tracking with shoe-mounted inertial sensors. *IEEE Computer Graphics and Applications*, 5(6):38–46, 2005.

[2] S. Godha, G. Lachapelle, and M.E. Cannon. Integrated GPS/INS system for pedestrian navigation in a signal degraded environment. In *In Proc. ION GNSS*, 2006.

[3] L. Ojeda and J. Borenstein. Non-GPS navigation for emergency responders. In *International Joint Topical Meeting: Sharing Solutions for Emergencies and Hazardous Environments*, 2006.

[4] J. Rantakokko, P. Strömbäck, J. Rydell, P. Händel, J. Callmer, D. Törnqvist, F. Gustafsson, M. Jobs, and M. Grudén. Accurate and reliable soldier and first responder indoor positioning: Multi-sensor systems and cooperative localization. *IEEE Wireless Communications Magazine*, 2011.

[5] P. Rudol, M. Wzorek, and P. Doherty. Vision-based pose estimation for autonomous indoor navigation of micro-scale unmanned aircraft systems. In *In Proc. IEEE International Conference on Robotics and Automation*, pages 1913–1920, 2010.

[6] G. Schweighofer and A. Pinz. Robust pose estimation from a planar target. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006.

[7] P. Strömbäck, J. Rantakokko, S.-L. Wirkander, M. Alexandersson, K. Fors, I. Skog, and P. Händel. Foot-mounted inertial navigation and cooperative sensor fusion for indoor positioning. In *Proc. ION International Technical Meeting (ITM)*, 2010.

[8] D. Wagner and D. Schmalstieg. ARToolKitPlus for pose tracking on mobile devices. In *Proceedings of 12th Computer Vision Winter Workshop (CVWW'07)*, 2007.