Simultaneous Localization and Mapping for Pedestrians using only Foot-Mounted Inertial Sensors

Patrick Robertson, Michael Angermann
Institute for Communications and
Navigation, German Aerospace Center
(DLR), D-82234 Wessling, Germany
patrick.robertson@dlr.de

Bernhard Krach
Institute for Communications and
Navigation, German Aerospace Center
(DLR); B. Krach is now with EADS.
bernhard.krach@eads.com

ABSTRACT

In this paper we describe a new Bayesian estimation approach for simultaneous mapping and localization for pedestrians based on odometry with foot mounted inertial sensors. When somebody walks within a constrained area such as a building, then even noisy and drift-prone odometry measurements can give us information about features like turns, doors, and walls, which we can use to build a form of a map of the explored area, especially when these features are revisited over time. Our initial results for our novel scheme which we call "FootSLAM" are very surprising in that true SLAM with stable relative positioning accuracy of 1-2 meters for pedestrians is indeed possible based on inertial sensors alone without any prior known building indoor layout. Furthermore, the 2D maps obtained even for just 10 minutes of walking converge to a good approximation of the true layout forming the basis for future automated collaborative mapping of buildings.

Author Keywords

Pedestrian navigation, Simultaneous Localization and Mapping, Odometry, Indoor Positioning, INS-based positioning.

ACM Classification Keywords

G.3 PROBABILITY AND STATISTICS: Probabilistic algorithms (including Monte Carlo).

General Terms

Algorithms, Theory, Experimentation.

INTRODUCTION

Recent work has shown remarkable advances in the area of indoor positioning with low cost inertial sensors worn by pedestrians. The work of Foxlin on foot mounted Inertial Measurement Units (IMUs) [5] has shown how zero velocity updates - ZUPTs- during the rest phase of a pedestrian's foot can be used to mitigate the problem of non-linear error growth in inertial integration over time. Nevertheless, errors
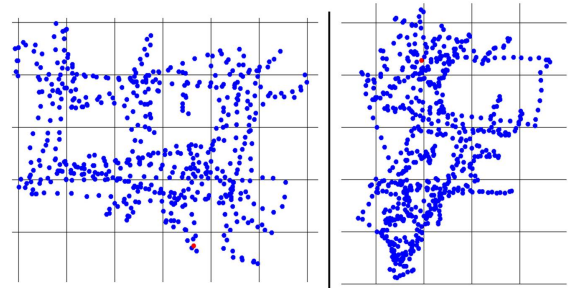
Figure 1. Plots from two walks around an office environment (one rectangular corridor circuit and rooms on the inside and outside of it). Shown is ZUPT aided inertial navigation based on a foot mounted IMU. Both these traces produce viable maps that are shown later.
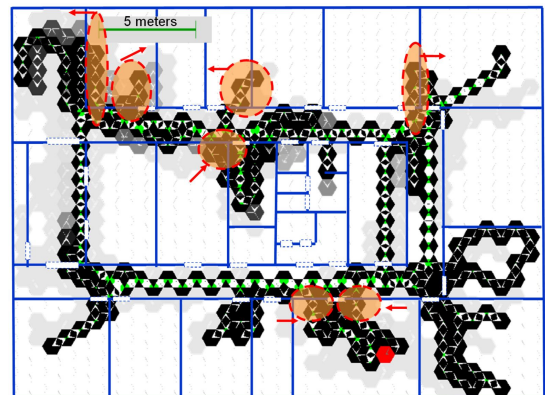

Figure 2. A map learnt from foot mounted IMU data alone plotted on the actual building layout for reference. See Figure 4 for explanations.

still accrue over time, especially the heading error which is only weakly observable from the zero velocity update. Figure 1 shows the data from the IMU attached to a shoe processed in a typical Inertial Navigation System (INS) using ZUPTs: The drifts (particularly in heading) lead to large divergence from the true path. Three groups independently showed that this problem can be addressed by employing known building layouts [10, 3, 6]. All used "particle filtering" (PF) methods where the building layout information is used to constrain particle movement to within the areas accessible to a pedestrian. As a result, long term error stability can be achieved when the map is sufficiently accurate and the layout sufficiently constrains the motion. The goal of this work is to *cope with situations where the floor plan is unknown*. One useful application would be the ability to use

data collected from people wearing IMUs on their shoes as they walk about in some environment during their lives (such as their office premises, while traveling, or at home) and to build maps automatically which can later be combined with other maps of the same or other areas and shared with navigating users. See Figure 2 for an example corresponding to Figure 1. These maps could then be used by people requiring assistance using positioning, such as rescue personnel, people with disabilities, or visitors to a company.

The robotics community has long used numerous sensors such as sonar, laser ranging scanners and cameras to perform high precision positioning of robots. The SLAM (Simultaneous Localization and Mapping) problem was addressed to allow robots to navigate in a-priori unknown environments [9]: A moving robot explores its environment and uses its sensor information to build a "map" of landmarks.

This paper builds on prior work on pedestrian positioning using foot mounted IMUs as well as the SLAM approach used in robotics. *The novelty here is the fact that we use no visual or other exteroceptive sensors of any kind*; the only sensors we used are the foot mounted IMU. We show that a pedestrian's location and the building layout can be jointly estimated by using the pedestrian's odometry alone, as measured by the foot mounted IMU. We begin by presenting the problem formulated as a dynamic Bayesian estimation problem and a Rao-Blackwellized particle filtering approach that follows the FastSLAM principle [8]. Subsequently we propose a probabilistic map that represents human motion in a two-dimensional hexagonal grid (somewhat similar to an occupancy grid [4] but with a different purpose). We then show qualitative and quantitative results obtained from real data obtained from a person walking in a building.

DYNAMIC BAYESIAN NETWORK FORMULATION
We will formulate the problem as a dynamic Bayesian network (DBN). The key is to suitably represent the actor in the system; i.e. the pedestrian. When people walk in a constrained environment, they rely mainly on visual cues in order to avoid walls and other obstacles. Such a person might be walking towards a particular target destination such as an office, or might just be walking randomly in the accessible space in an office during a conversation.

In robotic SLAM the robot is controlled by a series of inputs U that control its movement and these inputs are used in the SLAM estimation as inputs to a probabilistic motion model. Here, we can state that the human visual system is interpreting the environment and using it to guide motion: Observed direct physical constraints such as walls will influence intentions which will result in a person deciding which steps U to take. We obviously have no means of directly observing the human visual system **Vis**, nor can we directly measure where the person might actually want to go next (**Int**). We do have sensors that (differentially) measure the steps a person takes, i.e. a foot mounted IMU. Figure 3 shows a dynamic Bayesian network (DBN) that encodes relevant aspects of the system. Random Variables (RVs) are denoted in bold face. The step transition vector $\mathbf{U}_k$ has a special property: given the old and new poses $\mathbf{P}_{k-1}$, $\mathbf{P}_k$ the step transition $\mathbf{U}_k$ is determined entirely; just as knowledge of any two of the state variables $\mathbf{P}_{k-1}$, $\mathbf{P}_k$ and $\mathbf{U}_k$ determines the third.
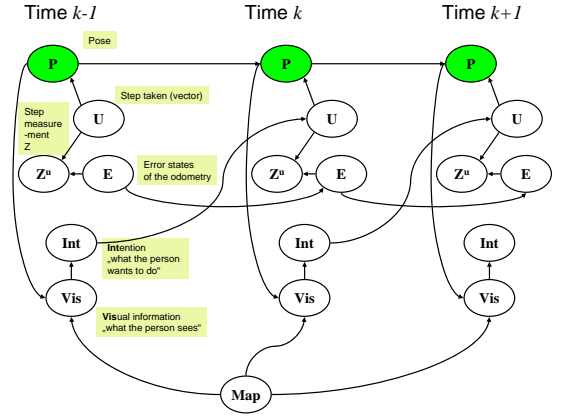

Figure 3. DBN for FootSLAM showing three time slices. The Map can include any features and information to let the pedestrian choose Int.

Step measurements $\mathbf{Z}_k^U$ are obtained in a manner as described in [6] and we assume that a suitable Kalman filter or similar algorithm was used to estimate the step vector $\mathbf{U}_k$. From the viewpoint of our DBN this estimate is a *measurement*. The only other influence on the measurement $\mathbf{Z}_k^U$ is a state variable $\mathbf{E}_k$ that encodes the correlated errors of the IMU.

DERIVATION OF BAYESIAN FILTER FOR GENERIC MAPS
Our goal is to estimate the states and state histories of the DBN given the series of all observations $\mathbf{Z}_{1:k}^U$ from the foot-mounted IMU. In a Bayesian formulation we compute the joint posterior $p(\mathbf{P}_{0:k}\mathbf{U}_{0:k}\mathbf{E}_{0:k}, \mathbf{M}|\mathbf{Z}_{1:k}^U)$ which factorizes into

$$p(\mathbf{M}|\mathbf{P}_{0:k}) \cdot p(\{\mathbf{P}\ \mathbf{U}\ \mathbf{E}\}_{0:k}|\mathbf{Z}_{1:k}^U). \qquad (1)$$

The expression for the map probability in (1) simplifies because the assumed knowledge of $\mathbf{P}_{0:k}$ makes the map $\mathbf{M}$ conditionally independent of $\mathbf{U}_{0:k}$, $\mathbf{E}_{0:k}$, and the measurements $\mathbf{Z}_{1:k}^U$; as follows from the DBN in Figure 3 and the relationship linking $\mathbf{P}_{k-1}$, $\mathbf{P}_k$ and $\mathbf{U}_k$. It is important to point out that the additional states - encoding vision and intention - never actually have to be explicitly used - they are in the model only for completeness and understanding. We shall express the right hand term in (1) recursively in the sense of a Bayesian filter. It can be shown that the recursive formulation is $p(\{\mathbf{P}\ \mathbf{U}\ \mathbf{E}\}_{0:k}|\mathbf{Z}_{1:k}^U)$

$$\propto \quad p(\mathbf{Z}_k^U|\mathbf{U}_k\mathbf{E}_k) \cdot p(\{\mathbf{PU}\}_k|\{\mathbf{P}\ \mathbf{U}\}_{0:k-1}) \cdot$$
$$p(\mathbf{E}_k|\mathbf{E}_{k-1}) \cdot p(\{\mathbf{P}\ \mathbf{U}\ \mathbf{E}\}_{0:k-1}|\mathbf{Z}_{1:k-1}^U). \quad (2)$$

The recursion is usually begun with the pose $\mathbf{P}_0$ set to an arbitrary fixed state, since performing SLAM without any true heading or location information is invariant over rotation and translation. We assume pertinent post-processing to resolve rotation, translation (and scale) transformations.

It is clear from the DBN that the map must play a role in determining the second term of (2), the pose and step transition probability. Marginalizing over $\mathbf{M}$ we can express this as

$$I \hat{=} \int_{\mathbf{M}} p(\{\mathbf{PU}\}_k|\mathbf{P}_{k-1}, \mathbf{M}) \cdot p(\mathbf{M}|\mathbf{P}_{0:k-1})d\mathbf{M}. \quad (3)$$

If we are able to compute $p(\mathbf{M}|\mathbf{P}_{0:k-1})$ and the influence of a map on the transition from $\mathbf{P}_{k-1}$ to $\mathbf{P}_k$, then we can perform

sequential Bayesian estimation of the map and pose.

Sequential Monte-Carlo - Particle Filtering
In order to obtain an online particle filter we sample from the "likelihood PF" proposal density shown in the algorithm summary below. Following [2], it can be shown that to a very good approximation $w_k^i \propto w_{k-1}^i \cdot I^i$. We therefore apply a Rao-Blackwellized PF [8] based on (1): Each particle i represents $\{\{\mathbf{PUE}\}_k^i, p(\mathbf{M}|\mathbf{P}_{0:k}^i)\}$.

ADOPTING PROBABILISTIC TRANSITION MAPS
We will now introduce a probabilistic map, based on the probability of the pedestrian crossing transitions in a regular 2D grid of adjacent hexagons of radius r. We chose hexagons because they are the polygons with the greatest number of sides that can be arranged to cover a 2D area completely without overlap. Furthermore, six angular transitions appear an appropriate number of choices for which human motion would be reasonably independent between angles.

We restrict this space to the region visited by any particle and define H_h as one of N_H hexagons, where h uniquely references a hexagon's position. Our map RV $\mathbf{M}$ is defined as the set comprising all $\mathbf{M}_h$ where $\mathbf{M}_h$ is a RV vector of length 6 with each component RV denoting the transition probability:

$$\mathbf{M}_{h(\mathbf{P}_{k-1})}^{e(\mathbf{U}_k)} \hat{=} P(\mathbf{P}_k \in H_j | \mathbf{P}_{k-1} \in H_h) \ ; \text{ where } j \neq h, \quad (4)$$

for leaving the h-th hexagon over the edge e via $\mathbf{U}_k$ to hexagon H_j. We assume that the map factors into local, conditionally independent components, and step $\mathbf{U}_k$ is only dependent on $\mathbf{P}_{k-1}$ and the local map $\mathbf{M}_{h(\mathbf{P}_{k-1})}$. Writing $\tilde{h}$ for outgoing hexagon $h(\mathbf{P}_{k-1})$, and $\tilde{e}$ for $e(\mathbf{U}_k)$, we compute the integral I by integrating over $\mathbf{M}_{\tilde{h}}^{\tilde{e}}$ for the respective edge.

Learning the Transition Map
Learning is straightforwardly based on Bayesian learning of probabilities of discrete RVs. Each time a particle makes a transition $\mathbf{P}_{k-1}^i \rightarrow \mathbf{P}_k^i$ across edge $\tilde{e}$ we count this transition in the local map of hexagon $H_{\tilde{h}}$. When computing the counts we assume that observing a certain transition from an outgoing hexagon to an incoming one allows us to increment the counts for the outgoing as well as the incoming one. This is the same as assuming that a person is likely to walk in either direction and improves convergence. In order to easily incorporate prior information we assume that $p(\mathbf{M}_{\tilde{h}}^{\tilde{e}}|\mathbf{P}_{0:k}^i)$ follows a beta distribution and integrating I^i yields

$$I^i \ \alpha \ \frac{N_{\tilde{h}}^{\tilde{e}} + \alpha_{\tilde{h}}^{\tilde{e}}}{N_{\tilde{h}} + \alpha_{\tilde{h}}} \qquad (5)$$

where $N_{\tilde{h}}^{\tilde{e}}$ is the number of times particle i crossed the transition, $N_{\tilde{h}}$ is the sum of the counts over all edges of the hexagon in this particle's map counters; and $\alpha_{\tilde{h}}^{\tilde{e}}$ and $\alpha_{\tilde{h}} = \sum_{e=0}^{5} \alpha_h^e$ are the prior counts.

SUMMARY OF THE ALGORITHM
1. Initialize all N_p particles to $\mathbf{P}_0^i = (x, y, h = 0)$ where x, y, h denote the pose location and heading in 2D; draw $\mathbf{E}_0^i$ from a suitable initial distribution for the error states.

2. for each time step increment k:
   (a) Particles are drawn from the proposal density $p(\mathbf{E}_k|\mathbf{E}_{k-1}^i) \cdot p(\mathbf{U}_k|\mathbf{Z}_k^U, \mathbf{E}_k^i)$ from left to right..
   (b) The pose $\mathbf{P}_k^i$ is is the vector addition of $\mathbf{U}_k^i$ to $\mathbf{P}_{k-1}^i$.
   (c) The particle weight updates are $w^i \ \alpha \ w_{k-1}^i \cdot I^i$; the counts used in I^i are those that are computed up to step $k-1$; weights are then normalized,
   (d) Increment each particle's $N_{\tilde{h}}^{\tilde{e}}$ for the transitions of both hexagons joined by the edge $\tilde{e}^i$ w.r.t. the outgoing hexagon $\tilde{h}^i$; perform resampling if required.

So far an increment of the time index k is associated with a step that leads from one hexagon to an adjacent one. In reality a step might keep us in the hexagon or it might lead us over several. To address this we simply perform a weight update only when we have stepped out of the last hexagon and apply multiple products of (5) for all edges crossed. Similarly, we update the counts of all edges crossed. We also incorporated a minor correction term in the weight update (raising it to a power depending on the step's angle in the grid) to account for the fact that straight tracks with different angles traversing the grid will yield slightly different total number of hexagon edge transitions per distance unit (otherwise particles with some directions are slightly favored).

EXPERIMENTS, RESULTS AND CONCLUSIONS
We performed three runs where a pedestrian walked around an office layout for roughly 10 minutes with a foot mounted IMU and $r = 0.5$m. The data sets were recorded and processed offline. Results for the learned maps (translated, rotated and scaled) are shown in Figures 2 and 4. The maps in those areas that were visited reflect the real path and are accurate to about 1-3 meters, with better accuracy in the corridors that were frequented more often. Both walks in Figure 4 intentionally remained out of "loop closure" in the corridor for some time. The particles start converging once the user backtracks or revisits a region for about 10 meters. Hexagons have to be re-visited once or twice in each main undirected traversal axis (remember we increment bidirectionally) for a usable map to emerge, and this governs the required duration of a walk; it fits within what a person typically covers in an office day, say. Accuracy in any case will be limited by the average physical structure dimension, such as corridors and doors which is about 1-2 m. Note that map errors were often in different regions of the individual map, suggesting that map combination from many walks could be valuable. We recorded ground truth for two positions at opposite corners of the main corridor. The error growth (as the walk progessed) for FootSLAM processing is shown in Figure 5. Our coordinate system origin was both the starting point and one of the reference points and we restricted the hexagon area to remove rotation ambiguity. With sufficient particles we achieve an accuracy of relative position to within two meters at the two reference points. Without FootSLAM we see error growth after some time - the INS coasted without too much error for about 300 seconds. Durations from 30s - 300s are typical and suggest that without maps the PF can bridge areas like large halls where
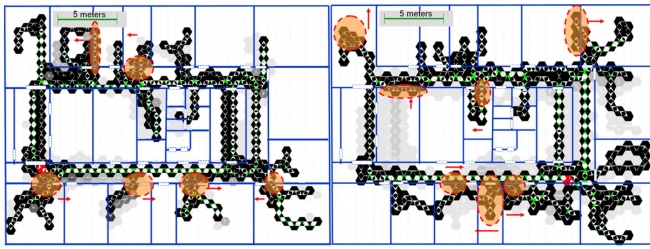
Figure 4. Two other maps learnt from IMU data alone. Shown are an overlap of the posterior (i.e. weighed average) map (in shades of grey) and the more useful MAP map (i.e. the map of the "best" particle) - black. Hexagons which the pedestrian has visited more often are more open, unvisted ones are not shown. More white within the hexagons themselves means more frequently crossed transitions. Ovals indicate errors - the arrow indicating roughly the error vector (see the scale).

there are no features for FootSLAM to map; at least to let someone find a destination or till a restricting door/corridor is again reached. In order to work, we expect FootSLAM to need a certain minimum average restrictedness of motion (related to entropy) but it can survive some open areas given sufficient particles. To achieve accurate mapping, especially in the rooms, we need a large number of particles ($>$ $10,000$). This needed 200 MB and took about 30 minutes on a 2,4GHz Pentium core. In our experiments we applied two independent additive random walk processes to $\mathbf{E}_k$ comprising heading bias and heading temporal drift error states affecting $\mathbf{Z}_k^U$. The additive error in 2D space between $\mathbf{Z}_k^U$ rotated according to $\mathbf{E}_k$, and $\mathbf{U}_k$, was white and Gaussian.

In many SLAM approaches without location sensors there is a map scaling error in addition to rotation and translation. In FootSLAM, this error is due to biases in the worn sensors, occasional erroneous ZUPTs and subsampling/clipping of the IMU signals that affect the step length estimation, and also a result of particles exploring hypotheses of different lengths. In our quantitative evaluation and videos [1] we have *not* used an individual length correction factor. This was only done using a constant factor of 1.15 which we have established quite reliably for that sensor setup. Only for the illustrations in the known floorplans did we adjust the scaling so that the resulting map fits the known floor plan and this additional scaling factor was less than 10%. Scaling can possibly be automated once tracks start to be grounded to outdoors GPS measurements or mobile radio positioning.

In this paper we have introduced FootSLAM, a novel scheme for simultaneous localization and mapping for pedestrians using only foot mounted inertial sensors. We have used a probabilistic map representation using a hexagon grid of a 2D area that is learnt per particle in a FastSLAM like approach. Our results are very encouraging: We have shown using real world data that accurate maps and stable positioning can in fact be achieved. Since our maps are probabilistic, estimation of pedestrians' future paths could also be performed - similar to work for driver intent estimation [7]. Further work should also integrate more sensors, address 3D, as well as collective mapping where users collect data during their daily lives and maps are combined and improved.

Acknowledgements

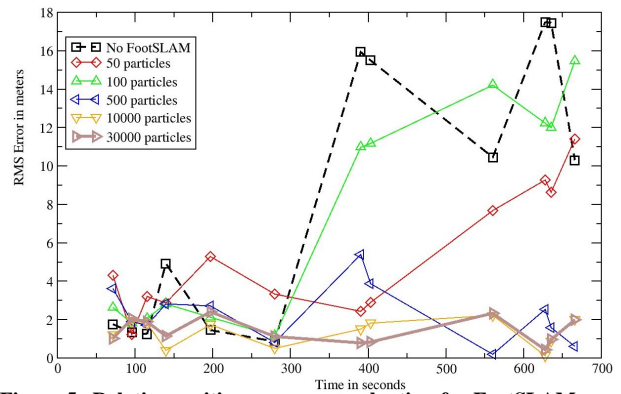**Figure 5. Relative position accuracy evaluation for FootSLAM - each curve is a *single* run of our particle filter on one data set.**

REFERENCES

1. FootSLAM videos. http://www.kn-s.dlr.de/indoornav.

2. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear / non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, Feb. 2002.

3. S. Beauregard, Widyawan, and M. Klepal. Indoor PDR performance enhancement using minimal map information and particle filters. In *Proc. of the IEEE/ION PLANS 2008*, Monterey, USA, May 2008.

4. A. Elfes. Using occupancy grids for mobile robot perception and navigation. *IEEE Computer Magazine*, 22(6), 1989.

5. E. Foxlin. Pedestrian tracking with shoe-mounted inertial sensors. *IEEE Computer Graphics and Applications*, 25(6):38–46, Nov. 2005.

6. B. Krach and P. Robertson. Cascaded estimation architecture for integration of foot-mounted inertial sensors. In *Proc. of the IEEE/ION PLANS 2008*, Monterey, USA, May 2008.

7. J. Krumm. A markov model for driver turn prediction. In *SAE 2008 World Congress, Detroit, MI USA*. Springer-Verlag New York, Inc., 2008.

8. M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM: A factored solution to the simultaneous localization and mapping problem. In *Proc. AAAI National Conference on Artificial Intelligence*, Edmonton, Canada, 2002.

9. R. Smith, M. Self, and P. Cheeseman. Estimating uncertain spatial relationships in robotics. In I. J. Cox and G. T. Wilfong, editors, *Autonomous robot vehicles*, pages 167–193. Springer-Verlag New York, Inc., 1990.

10. O. Woodman and R. Harle. Pedestrian localisation for indoor environments. In *Proc. of the UbiComp 2008*, Seoul, South Korea, Sept. 2008.