Multi-Sensor SLAM for Tactical Situational Awareness

Laura Ruotsalainen, Martti Kirkko-Jaakkola, Liang Chen, Simo Gröhn, Robert Guinness, Heidi Kuusniemi, *Department of Navigation and Positioning, Finnish Geospatial Research Institute (FGI), National Land Survey of Finland, Finland*

BIOGRAPHIES

Dr. Laura Ruotsalainen is a Research Manager and Deputy Director of the Department of Navigation and Positioning at the Finnish Geospatial Research Institute (FGI), Finland, where she leads the research group on Sensors and indoor navigation. She received her doctoral degree in 2013 from the Department of Pervasive Computing, Tampere University of Technology (TUT), Finland. Her doctoral studies were partly conducted at the Department of Geomatics Engineering at the University of Calgary, Canada. Her doctoral research was focused on vision-aided seamless indoor/outdoor pedestrian navigation. Her current research interests cover adaptive integration of sensors and radio positioning means for robust navigation, situational awareness and Global Navigation Satellite systems (GNSS) interference mitigation.

Dr. Martti Kirkko-Jaakkola is a Senior Research Scientist at the Finnish Geospatial Research Institute. He received his M.Sc. and D.Sc. (Tech.) degrees from Tampere University of Technology, Finland, in 2008 and 2013, respectively. His research interests include various precise GNSS positioning and inertial navigation applications using low-cost equipment.

Dr. Liang Chen received the M.Sc. degree in control theory and control engineering from Jiangsu University, China, in 2004, and the Ph.D. degree in signal and information processing from Southeast University, China, in 2009. He was a Post-Doctoral Research Scientist with the Department of Mathematics, Tampere University of Technology, Finland, from 2009 to 2011. He is a Senior Research Scientist with the Department of Navigation and Positioning, Finnish Geospatial Research Institute, Finland. He is an Adjunct Professor with Jiangsu University. His research interests include statistical signal processing for positioning, wireless positioning using signals of opportunity, and sensor fusion algorithms for indoor positioning.

Simo Gröhn is a research scientist and a PhD student at the Department of Navigation and Positioning at FGI since 2015. The topic of his thesis is the integration of vision based positioning and sensor data for indoor positioning. His research focus consists of vision-aiding in positioning, sensor error modelling and integration algorithms. Mr. Gröhn holds a Master of Science (tech) degree and before joining FGI has worked in research and development projects related to optical sensors in industry.

Robert Guinness is a Senior Research Scientist at the Finnish Geospatial Research Institute, where he leads the research group on Intelligent Mobility and Geospatial Computing. He received his M.Sc. in 2006 from the International Space University, and he is currently a doctoral candidate at Tampere University of Technology, Finland. His research interests include context awareness for navigation applications, machine learning, and privacy-preserving location technologies.

Prof. Heidi Kuusniemi is the Director at the Department of Navigation and Positioning at FGI. She is also an Adjunct Professor at TUT, Finland, and Aalto University, Finland, where she also is a lecturer on GNSS technologies. She is the President of the Nordic Institute of Navigation. She received her M.Sc. and Doctor of Technology degrees from TUT in 2002 and 2005, respectively. Her doctoral studies on reliability monitoring in personal satellite navigation were partly conducted at the Department of Geomatics Engineering at the University of Calgary, Canada. Kuusniemi's research interests cover various aspects of GNSS and sensor fusion.

ABSTRACT

Tactical and rescue operations need infrastructure-free accurate and reliable localization, information about the possibly unknown environment and knowledge of the status of each individual. Simultaneous Localization and Mapping (SLAM) is a key technology providing means for the first two mentioned needs.

In this paper we discuss a particle filter fusion for obtaining accurate localization for use in the SLAM algorithm. At present, there is no single infrastructure-free method providing accurate and reliable positioning indoors, but the key is to adaptively integrate measurements from various sensors having different sources of error and different operational restrictions. Here we discuss the principles of obtaining motion measurements using a monocular camera, a foot-mounted Inertial Measurement Unit (IMU), sonar, and a barometer.

Then, the implementation of a particle filter integrating all measurements for an accurate and reliable localization solution for use in SLAM will be described. Finally, experimental setup and results will be discussed. The developed method is shown to provide beyond the state-of-the-art performance and is anticipated to result in a SLAM solution addressing the demanding requirements set for a tactical or rescue system providing also situational awareness.

INTRODUCTION

Rescue and military applications require rapid, accurate and reliable information about unknown environments. In addition to the user position and navigation information, knowledge of the environment and user's motion is needed. Requirements for the system are stringent; it should function also in indoor environments, be infrastructure-free, lightweight and inexpensive. The infrastructure-free requirement is motivated by the fact that rescue and military personnel must be able to operate reliably in any environment, regardless of the available infrastructure. Simultaneous Localization and Mapping (SLAM) is a key technology for providing an accurate and reliable infrastructure-free solution for indoor situational awareness (Davison et al. 2007). However, indoor environments and the requirements for the system make the implementation of SLAM using existing algorithms challenging. Most existing algorithms were developed for use in robotics where size and weight requirements are not as stringent. Due to size limitations, we will implement SLAM using a monocular camera as an input. However, existing algorithms for monocular SLAM do not provide reliable enough results for rescue or military applications.

The long-term objective of our research is to develop a method for infrastructure-free tactical situational awareness. At present, most functioning indoor localization systems are based on processing short- range radio signals from pre-installed networks and therefore cannot be considered as infrastructure-free. Advances in sensor technology have been rapid during the last several decades. Self-contained Micro-Electro-Mechanical (MEMS) sensors fulfill the size and cost requirements set for an infrastructure-free military and rescue system (Rantakokko et al. 2011). Use of, e.g., inertial sensors provides enough information for propagating a known initial position for the purposes of forming a SLAM solution. However, the MEMS sensors suffer from biases and drift errors that may decrease the position accuracy substantially (Collin 2006). Therefore, sophisticated error modelling and implementation of integration algorithms are key for providing a viable final result. Our approach is to integrate a monocular camera, multiple Inertial Measurement Units (IMUs), a barometer and a ranging sensor to obtain a solution for SLAM, as well as tactical motion information, e.g. detecting whether a rescue person or a soldier is running or crawling, etc. In this paper we discuss a particle filter implementation for integrating measurements from visual perception, a foot-mounted IMU, a barometer and sonar.

The selection of the sensors is justified by their complementary nature, resulting in improved accuracy and reliability when correctly integrated. The SLAM solution will be based on an innovative monocular visual localization method, providing heading change and translation information, which is discussed in more detail in (Ruotsalainen et al 2015). It was developed to overcome two major obstacles complicating visual perception: the unknown scale of translation observed using a monocular camera and the shortage of features indoors. It provides promising results. However, the deficiency of the method is its inability to observe heading change when the scene incurs sudden changes, such as when the heading change exceeds 90 degrees. A foot-mounted IMU provides also heading change and translation (Skog et al. 2010), and when integrated with the visual measurements, a complete localization solution may be obtained also indoors.

Although the integration provides a two-dimensional solution with sufficient accuracy, tactical applications require also accurate vertical location information; in fact, a vertical position error of 3 meters may be unacceptable in a multi-storey building. Therefore measurements from a barometer, providing air pressure information transformed into the user height (Sabatini and Genovese 2013), are added to the integration. However, using a barometer for height information is challenging due to dynamic atmospheric conditions. For example, a change in pressure due to a window being opened may degrade the height observation significantly. The barometer is therefore supplemented by a sonar sensor providing ranging information (Leonard and Durrant-Whyte 1992). When the sonar is directed down towards the floor, the change in the range may be observed accurately. The height change observed using the sonar may be integrated with the barometer height measurements, as well as the two-dimensional localization solution, and as a result, an accurate and reliable 3-dimensional solution is achieved.

Particle filtering is a sophisticated option for integrating measurements emerging from pedestrian motion having non-Gaussian error characteristics (Arulampalam et al. 2002). However, the measurements come from sensors observing very different phenomena and therefore have different error sources. Although the sensors complement each other due to their different observations and error sources, special care has to be taken in error observation and modelling. Here we discuss the independent measurements obtained using the sensors specified above, the modelling of their errors, and implementation of the particle filter integrating all measurements for an accurate and reliable SLAM solution.

The method developed is tested via experiments conducted in an office environment. Test setup and results will be discussed further below. The results obtained

using the developed method show improvement on the accuracy and reliability for localization of monocular SLAM compared to previous methods. The proposed data fusion approach yields a vertical accuracy sufficient for floor identification in demonstrated test environment without utilizing WiFi or other local infrastructure. The method also advances the state-of-the-art in infrastructure-free SLAM solutions based on a monocular camera. Lastly, the research makes significant progress towards a functioning infrastructure-free situational awareness system, which is desperately needed in the application areas in question.

MEASUREMENT MODELS

This section describes the measurement models for all sensors used in the fusion, namely the mechanization for obtaining translation and heading information from a camera, a foot-mounted IMU, a barometer and sonar.

Visual Gyroscope and Visual Odometer

A camera can be considered as an additional self-contained sensor when integrated to a navigation system via specific mechanization. When the camera is attached to the body, motion of features in consecutive images provides enough information for observing translation and heading of the user. Motion of the features may be transformed into heading information in a straight forward manner under favorable conditions, as discussed below. However, observation of the absolute translation information is more challenging when using a monocular camera.

Due to the equipment size requirements, a stereo camera (with sufficient baseline) is not an option in pedestrian applications for rescue personnel or unmounted soldiers, so a monocular camera has to be used instead. Monocular cameras have advantages for visual perception compared to stereo cameras; they have wider field-of-view and provide faster image processing capabilities (Zhuang and Roth 1996). However, the difficulty in resolving the distance between the camera and a landmark (also called depth information) induces an ambiguous scale into the translation, and this issue is one of the most hindering challenges when using monocular visual perception for localization purposes. This section describes the concepts of visual gyroscope and visual odometer (Ruotsalainen 2013). A visual gyroscope measures the user heading change and a visual odometer provides absolute translation information, overcoming the scale problem by using information obtained from the visual gyroscope and a special configuration described below.

Visual Gyroscope

Indoor environments are often poor in features, complicating the visual perception. However, most human-made environments, like indoors and urban areas consist of straight lines in three orthogonal directions. Lines are suitable features to be used for visual perception because they are not easily disturbed by dynamic objects, like other humans in the scene.

Lines in three orthogonal directions seem to intersect in a point called a vanishing point. The locations of the three vanishing points V are defined by the orientation of the camera with respect to the scene and the calibration matrix K comprised of the intrinsic parameters of the camera. This relation is described by $\mathbf{V}=\mathbf{KR}$ (Gallagher 2005), where $\mathbf{R}$ is the rotation matrix of the camera, describing the attitude of the camera with respect to the scene. The vanishing point matrix $\mathbf{V}$ encompasses the homogenous representations of the horizontal, vertical, and central (at the direction where the camera lens is pointing) vanishing points $\mathbf{v}_x$, $\mathbf{v}_y$, $\mathbf{v}_z$, respectively. The x and y coordinates of the vanishing point in i direction may be represented by $\mathbf{v}_i = (x_{vi}, y_{vi}, 1)^T$.

The attitude of the camera ($\mathbf{R}$) with respect to the scene may be computed, if the locations of at least two vanishing points are determined and the camera's calibration matrix $\mathbf{K}$ is known. When the camera is mounted to the user's body, the camera heading obtained from $\mathbf{R}$ may be further interpreted as the user heading and therefore used for localization.

Visual Odometer

The relation between two points representing the same feature in separate images is called a homography and can be written as

$$\mathbf{x}' = \mathbf{Rx} + \mathbf{t}/Z \qquad (1)$$

where $\mathbf{x}$ and $\mathbf{x}'$ are normalized homogenous points (Hartley and Zissermann 2003) in the first and the second image, respectively, $\mathbf{R}$ is the camera's rotation matrix observed using the visual gyroscope presented above, $\mathbf{t} = [t_x, t_y, t_z]$ is the translation vector between points $\mathbf{x}$ and $\mathbf{x}'$, and Z is an unknown scale factor, corresponding to the depth of the objects in the first image.

The image point locations are resolved by looking at feasible features in consecutive images and matching the ones representing the same object using a SIFT-algorithm (Lowe 1999). Using this information and (1) the translation $\mathbf{t} = [t_x, t_y, t_z]$ with the unknown scale Z, i.e. $\mathbf{t}/Z$, may be computed.

The visual odometer discussed below resolves the value of the parameter Z using a special configuration of the camera shown in Figure 1 (Ruotsalainen 2013). When the camera is calibrated and its height (h) above the ground is known, the depth of the object projected into image points $\mathbf{x}$ and $\mathbf{x}'$, namely Z, may be resolved using the set of formulae (2)-(4) provided below.

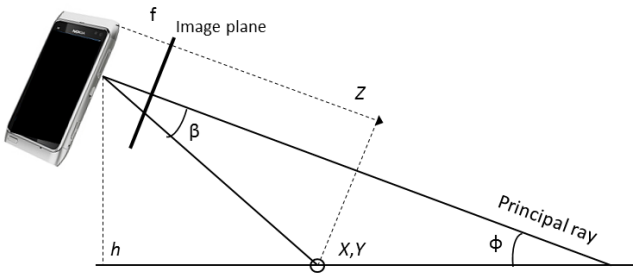

Figure 1. Special configuration of the camera for resolving the depth of the objects in images.

An object point, represented with coordinates (X, Y, Z), is projected onto a point (x, y) in an image. The height (H) of the image in pixels and the vertical field of view (vfov) are known from the camera calibration. Using these, the angle (β) between the camera's principal ray and the ray from the camera to the object may be resolved using

$$\beta = \arctan\left(\left(\frac{2y}{H} - 1\right) \tan\left(\frac{vfov}{2}\right) \right) \qquad (2)$$

Because the camera's pitch ϕ is obtained from the visual gyroscope, the Y coordinate of the object is obtained using trigonometry from

$$Y = \frac{h}{\sin(\phi + \beta)}, \qquad (3)$$

and finally the depth, namely the object's Z coordinate, may be calculated as

$$Z = Y\cos(\beta) = \frac{h\cos(\beta)}{\sin(\phi + \beta)}. \qquad (4)$$

After obtaining Z, the so-called scale problem is solved, and absolute translation can be obtained.

Error detection

Visual perception is prone to errors. For the methods discussed here, the most drastic errors develop both from the failure of finding parallel lines in the environment and failure in matching the correct image points in consecutive images. The performance of both the visual gyroscope and indirectly the visual odometer rely on finding the parallel lines for vanishing point computation. An error measure named LDOP is assigned to express the probability of success in line detection. However, even if line detection succeeds, the visual odometer may fail due to mismatching the SIFT-descriptors. Therefore an algorithm called 1-Point RANSAC is used for discarding the mismatched points (Civera et al. 2010). Although the 1-Point RANSAC algorithm is a sophisticated and fast method for discarding the outliers, some mismatched image points are still left in the data set. Therefore the visual odometer's translation measurements are still filtered. Both detection methods will be discussed briefly below.

(Ruotsalainen et al. 2012) discussed error detection based on detecting the geometry of lines found from the scene and intersecting at the central vanishing point. The evaluation is based on dividing the image into four sections around the estimated central vanishing point. If the lines intersecting at the vanishing point are found from four or three sections, the geometry is good for the calculations, and a low LDOP (line dilution of precision) value is assigned.

If lines are present in less than three sections, the LDOP value is evaluated based on the angle difference between each line pair and the image x-axis, given by

$$LDOP = \sqrt{\frac{2}{|\mathbf{H}|}} \qquad (5)$$

where (α_1) is the angle between the x-axis and the first line and (α_2) between the x-axis and the second line and $\mathbf{H}^{-1}$ is defined as

$$\mathbf{H}^{-1} = \frac{1}{|\mathbf{H}|} \begin{bmatrix} \sin^2(\alpha_1) + \sin^2(\alpha_2) & -\cos(\alpha_1)\sin(\alpha_1) + \cos(\alpha_2)\sin(\alpha_2) \\ \cos(\alpha_1)\sin(\alpha_1) + \cos(\alpha_2)\sin(\alpha_2) & \cos^2(\alpha_1) + \cos^2(\alpha_2) \end{bmatrix} \qquad (6)$$

where $|\mathbf{H}| = \sin^2(\alpha_1 - \alpha_2)$ is the determinant of $\mathbf{H}$. The smallest possible value for LDOP is $\sqrt{2}$, occurring when the two lines are perpendicular.

As mentioned above, the mismatched image points are discarded using the 1-point RANSAC, a modification of the traditional RANSAC algorithm (Fischler and Bolles 1981), algorithm. However, some outliers are left to the data set, mainly due to including some points into visual odometer calculations that are not close to the floor and therefore violating the configuration requirements. The outlier points distort the translation measurements obtained using the visual odometer algorithm. Therefore, the speed measurements obtained from the translation between two epochs are processed with a simple 1D median filter using a window size of three epochs. When a speed value differs dramatically from its preceding epochs, the median of three recent values is used.

Foot-mounted PDR

The outputs of a conventional strapdown IMU are a specific force measurement $\mathbf{f}^b$ and an angular rate measurement ω^b, both corresponding to the IMU's body

coordinate frame b . Neglecting factors such as the Coriolis effect, the full six-degrees-of-freedom strapdown inertial navigation solution can be obtained by solving the system of differential equations (Titterton and Weston 2004)

$$\begin{cases} \dot{C} &= C(\omega^b \times) \\ \dot{\mathbf{v}}^n &= C\mathbf{f}^b + \mathbf{g}^n \\ \dot{\mathbf{p}}^n &= \mathbf{v}^n \end{cases}, \quad (7)$$

where the matrix C denotes the direction cosine matrix between the body frame b and the navigation coordinate frame n; × denotes the 3×3 skew-symmetric cross product matrix; $\mathbf{v}^n$ and $\mathbf{p}^n$ are the velocity and position in the navigation frame, respectively; and $\mathbf{g}^n$ denotes the local gravitational acceleration vector. Using this method, the initial position, velocity, and attitude must be known, and the rates can be used to propagate from the initial position using the standard dead reckoning approach.

In general, the quality of low-cost MEMS inertial sensors is inadequate for use in the above mechanization except for very short periods of time. This is why pedestrian dead reckoning (PDR) mechanizations are typically used which resort to detecting steps from the inertial sensors' output waveform and applying an external model for the stride length (Beauregard and Haas 2006). Such mechanization avoids the error-prone double integration of inertial measurements. However, human gait differs from person to person, which makes the stride length difficult to predict (Leppäkoski 2015).

However, mounting the IMU to the foot of a pedestrian constitutes a special case: Unless the shoe slips, the IMU remains stationary for a short period of time between steps (Foxlin 2005). In this case, the strapdown mechanization can essentially be interpreted as a boundary value problem instead of an initial value problem; this makes it possible to compute the step displacement using (7). Not only does this approach avoid the need for an external step length model, but it also works when stepping sidewards (or backwards) and when walking in stairs. However, the vertical performance depends on the accuracy of the local gravitational acceleration value used in the strapdown computations.

Since the strapdown mechanization is traditionally implemented with the help of an error-state Kalman filter modeling the propagation of position, velocity, and attitude errors, it is straightforward to implement foot-mounted PDR by applying a zero-velocity update (ZUPT) to the error-state filter whenever the IMU is detected to be at rest (Titterton and Weston 2004): The velocity error therefore becomes directly observable, and other error states can be estimated based on their correlation with the velocity error. Typically, a ZUPT can be applied roughly every second, which slows down the error growth significantly. Various algorithms for stationarity detection exist (Skog et al. 2010).

However, the MEMS gyros are quite sensitive to temperature changes and need minutes to stabilize after e.g. entering indoors from cool outdoor conditions before they can provide accurate heading information (Leland 2005). Also, high linear accelerations induce large G-sensitivity errors (Bancroft and Lachapelle 2012). Tactical operations often include both large changes in the temperature and large accelerations. Therefore, foot-mounted IMUs alone do not provide sufficient performance.

Barometer and Sonar Measurements for Accurate Vertical Positioning

Barometer

Determining the altitude based on inertial measurements only is prone to drift due to measurement errors and inaccurate knowledge of the local gravitational acceleration. On the other hand, the problem of unknown scale makes it difficult to track vertical motion using cameras. Therefore, we employ a barometer and sonar for estimating the altitude.

A barometer observes the ambient pressure p which depends on the altitude h according to the International Standard Atmosphere model as (Parviainen et al. 2008)

$$h(p) = \frac{T_0}{k}\left(\left(\frac{p}{p_0}\right)^{\frac{kR}{g}} - 1\right) \quad (8)$$

where T_0 and p_0 are the sea level temperature and pressure, respectively; $k = -6.5 \cdot 10^{-3}$ K/m is the temperature lapse rate; $R = 287.04\ \mathrm{m^2/(Ks^2)}$ is the gas constant for air; and g is the magnitude of the gravitational acceleration. Unfortunately, the sea level temperature and pressure are unknown and vary in time. Therefore, unless these values are updated frequently, barometric measurements are useful for a short period only. Moreover, barometers are susceptible to abrupt changes in the ambient pressure caused by, e.g., opening a door or firing a gun (Parviainen et al. 2008). Also, drastic changes in temperature, for example in the case of fire, disturb the height solution obtained using a barometer. Therefore, for an accurate and reliable solution, complementary source of height measurements has to be used. In this research, sonar ranging measurements from the user body down to the floor are fused with the barometer height. Usually the range between the user and the floor doesn't change fast and therefore the outliers in

the barometer height solution, caused by the environmental changes, may be discarded.

Sonar

Ultrasonic sensors, i.e. sonars, produce high, centimeter level accuracy and millimeter level resolution range measurements. Sensors measure the time-of-flight of transmitted and again received sound signals reflecting from structures. This measurement is converted to ranging information by knowing that the signals traverse at the speed of sound.

Ultrasonic ranging sensors are used in indoor navigation as they are low cost, compact, reliable and scalable (Ijaz et al. 2013).

PARTICLE FILTER FUSION

Development of a sophisticated algorithm for data fusion is the key for obtaining an accurate and reliable localization solution. The solution is represented at each epoch t_k by the state ($\mathbf{x}_k$) of the system and estimated based on measurements $\mathbf{z}_k$, collected since the start of the localization. The estimation is usually done by using recursive Bayesian estimation algorithms (Gelman et al. 2000, Thrun et al. 2005). The Bayesian algorithms contain two phases; prediction and update.

In the prediction step, the *a priori* probability $p(\mathbf{x}_k|\mathbf{z}_{0:k-1})$ is calculated from the *a posteriori* probability $p(\mathbf{x}_k|\mathbf{z}_{0:k})$, representing the last state, and the process model $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, describing the system dynamics (Klingbeil et al. 2010), by:

$$p(\mathbf{x}_k|\mathbf{z}_{0:k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1})\,p(\mathbf{x}_{k-1}|\mathbf{z}_{0:k-1}) \tag{9}$$

In the update step, when new measurements are obtained, the *a posteriori* probability is updated using:

$$p(\mathbf{x}_k|\mathbf{z}_{0:k}) \propto p(\mathbf{z}_k|\mathbf{x}_k,\mathbf{z}_{0:k-1})\,p(\mathbf{x}_k|\mathbf{z}_{0:k-1}) \tag{10}$$

where $p(\mathbf{z}_k|\mathbf{x}_k,\mathbf{z}_{0:k-1})$ is the likelihood function, representing the measurement model, and $p(\mathbf{x}_k|\mathbf{z}_{0:k-1})$ is the *a priori* probability calculated in the prediction step.

Traditionally, different types of Kalman filters (Anderson and Moore 1979, Julier et al. 2000) have been used for Bayesian estimation in localization. However, the performance of the solution is degraded in certain cases due to their assumptions regarding linearity of the system and Gaussian modelling of the uncertainty. Especially for pedestrian localization, particle filters provide an improved solution (Arulampalan et al. 2002).

Particle Filter Based Navigation Method

In a particle filter (PF), the *a posteriori* probability is approximated by a set of N weighted samples at time k-1, i.e., discrete random samples $(w_{k-1}^{(i)}, \mathbf{x}_{k-1}^{(i)})$, i=1,…,N (Kuusniemi et al. 2012), as:

$$p(\mathbf{x}_{k-1}|\mathbf{z}_{0:k-1}) \approx \sum_i w^{(i)} \delta(\mathbf{x}_{k-1} - \mathbf{x}_{k-1}^{(i)}) \tag{11}$$

where $\delta(\cdot)$ denotes the Dirac delta function. On the update step, new samples are generated from the current state based on the *a posteriori* distribution and the new measurements. New weights, constrained to sum up to unity, are computed as:

$$w_k^{(i)} \propto w_{k-1}^{(i)} \cdot p(\mathbf{z}_k|\mathbf{x}_k^{(i)}) \tag{12}$$

Thereby, the new set of samples is approximately distributed according to $p(\mathbf{x}_k|\mathbf{z}_{0:k})$ at time t_k. The effective number of particles in the set is evaluated at each epoch; if the number is below a predefined threshold, the set of particles is considered to be degenerate and needs to be resampled. Resampling means drawing a new set of particles from the distribution constituted by the old particles and their respective weights. In other words, low-weight particles are likely to be discarded and high-weight particles duplicated (Hol et al. 2006).

Process and Measurement Models

This section describes the process and measurement models used in this research.

Process models

Pedestrian positioning equations use a navigation model, which can be described as follows:

$$\begin{aligned}
E_{k+1} &= E_k + S_k \cos(\psi_k)\Delta t + n_{1,k} \\
N_{k+1} &= N_k + S_k \sin(\psi_k)\Delta t + n_2,k \\
H_{k+1} &= H_k + n_{3,k} \\
\dot{\psi}_{k+1} &= \dot{\psi}_k + n_4,k \\
\psi_{k+1} &= \dot{\psi}_k \Delta t + \psi_k + n_{5,k} \\
S_{k+1} &= S_k + n_{6,k}
\end{aligned} \tag{13}$$

where k indexes the epoch, i.e. at time t_k, E is the coordinate in East direction (longitude scaled into meters), N is the coordinate in North direction (latitude scaled into meters), H is the relative height value above the floor (in meters), S is speed (in m/s), ψ is the

heading defined with the origin East and counter-clockwise positive (in degrees), $\dot{\psi}$ is the heading rate (in degree/s), Δt is the time between epochs k and $k+1$, and $n_{i,k}$ is the state uncertainty component of the state vector for element i at time t_k. In summary, the state vector for estimation is $\mathbf{x}_k = [E \;\; N \;\; H \;\; \dot{\psi} \;\; \psi \;\; S]_k^T$.

Measurement model

The measurement model is defined as:

$$\begin{aligned} \dot{\psi}_{footIMU,k} &= \dot{\psi}_k + v_{1,k} \\ \dot{\psi}_{visual,k} &= \dot{\psi}_k + v_{2,k} \\ S_{footIMU,k} &= S_k + v_{3,k} \\ S_{visual,k} &= S_k + v_{4,k} \\ H_{baro,k} &= H_k + v_{5,k} \\ H_{ultrasonar,k} &= H_k + v_{6,k} \end{aligned} \tag{14}$$

The measurement vector is:

$$z_k = [\dot{\psi}_{footIMU} \quad \dot{\psi}_{visual} \quad S_{footIMU} \quad S_{visual}, H_{baro}, H_{ultrasonar}]_k^T \tag{15}$$

Depending on how many measurements exist at epoch *k*, the size of $\mathbf{z}_k$ vary. The measurement uncertainties of v_1, v_2, v_3, v_4 are based on the error model derived from statistical analysis. In the algorithm presented in this paper, all measurement noises are modelled as zero mean Gaussian noise, namely:

$$\begin{aligned} v_{1,k} &\sim N(0, \sigma^2_{ft,heading}) \\ v_{2,k} &\sim N(0, \sigma^2_{vs,heading}) \\ v_{3,k} &\sim N(0, \sigma^2_{ft,speed}) \\ v_{4,k} &\sim N(0, \sigma^2_{vs,speed}) \\ v_{5,k} &\sim N(0, \sigma^2_{baro,height}) \\ v_{6,k} &\sim N(0, \sigma^2_{us,height}) \end{aligned} \tag{16}$$

where *vs* denotes visual, *ft* foot-mounted, *baro* a barometer and *us* ultrasonar. Since the process model (13) is nonlinear, the PF is applied for state estimation, which follows the steps introduced in the previous section.

EXPERIMENTAL RESULTS

The performance of the multi-sensor integration for the localization part of SLAM was tested in an office environment. This section discusses the test setup and results.

Equipment Setup

All equipment used for localization was attached to the test person, one IMU to the foot and all other equipment to a test rack carried on the back as shown in Figure 2. The sensors and systems used for the experiment discussed in this paper included an Osmium MIMU22BT IMU, GoPro HD2 camera, XSENS MTI-G-700 barometer, and HRUSB-MaxSonar sonar. The Novatel SPAN system, including a GPS receiver with a Honeywell HG1700 AG58 tactical grade IMU, was used to determine the reference trajectory for assessing the performance of the algorithm. At the start of the experiment the system was initialized outdoors for ten minutes to obtain an accurate reference solution. Data collection started immediately when entering the building. The data collected in an office corridor was post-processed, including a path of 160 meters, with a duration of 150 seconds.

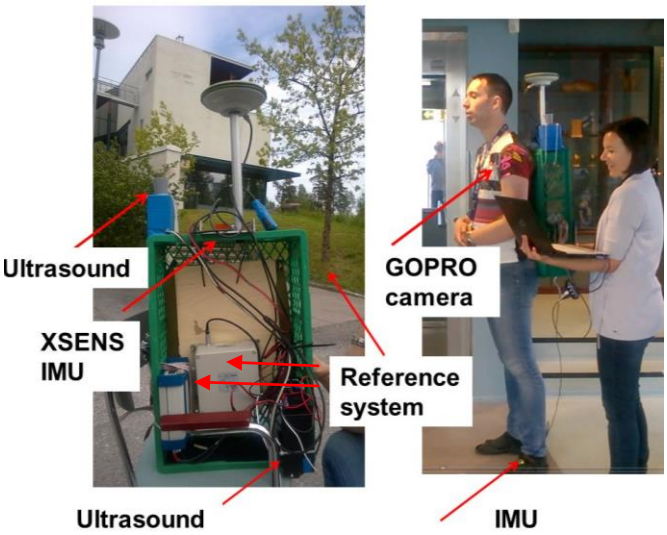

Figure 2: Test setup and equipment

The GoPro images were used for computing the visual gyroscope and odometer measurements with 6 Hz rate. Errors in observations were detected using the error measures discussed above. Figure 3 shows the LDOP values obtained from the experiment as a result of visual gyroscope processing. The LDOP values reveal those turns in the corridor where the visual gyroscope fails (Ruotsalainen 2013). The corridor has three turns larger than 90 degrees, and at those turning epochs LDOP values exceed the threshold set at 15. Based on our evaluation of the data and previous experience, the visual algorithms provide more accurate speed and heading measurements than the foot-mounted IMU when the LDOP values are small. One reason for the much poorer performance of the foot-mounted IMU data was the sudden change in the ambient temperature when entering the indoor space.

Keeping this in mind, the measurement errors of the particle filter are modelled as follows. When the LDOP value is below the threshold, the heading errors are

modelled as $\sigma^{corrid}_{vs,heading} < \sigma^{corrid}_{ft,heading}$ and the speed errors as $\sigma^{corrid}_{vs,speed} < \sigma^{corrid}_{ft,speed}$, *corrid* denotes the situation when going straight through corridor. Similarly, when the LDOP value exceeds the threshold the foot-mounted result is given more weight, and the heading errors are modelled as $\sigma^{turn}_{vs,heading} > \sigma^{turn}_{ft,heading}$ and the speed errors as $\sigma^{turn}_{vs,speed} > \sigma^{turn}_{ft,speed}$, *turn* denotes the situation when turning.

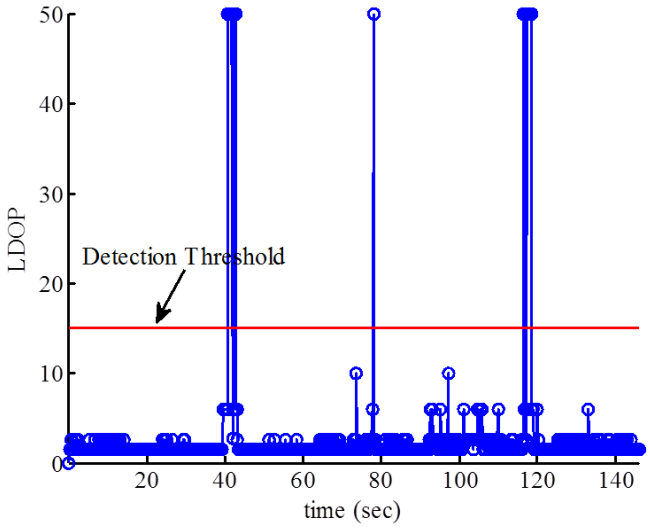

Figure 3: Visual gyroscope's LDOP values in the experiment and the threshold for rejecting the visual measurements

Due to the barometer dependency on the environmental changes, more confidence was put in the sonar height measurements in the fusion, i.e. the height measurement errors were modelled as $\sigma_{baro,height} > \sigma_{us,height}$. Figure 4 shows the ultrasonic ranging sensor, i.e. sonar, used in the experiments. Only one sonar was used in the experiments, and it was oriented facing down towards the floor.

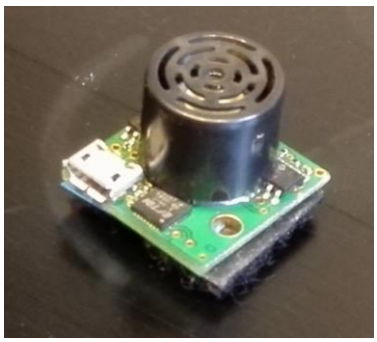

Figure 4: Ultrasonic ranging sensor HRUSB-MaxSonar used in the experiments

Results

All measurements obtained using the sensors presented above were fused using the particle filter and error models also discussed above. The number of particles in the filter was set to 1000. The mean error in the horizontal position was 3.14m with a standard deviation of 2.82 m. Figure 5 shows the path resulting from the fusion (red) compared to the reference trajectory (blue).

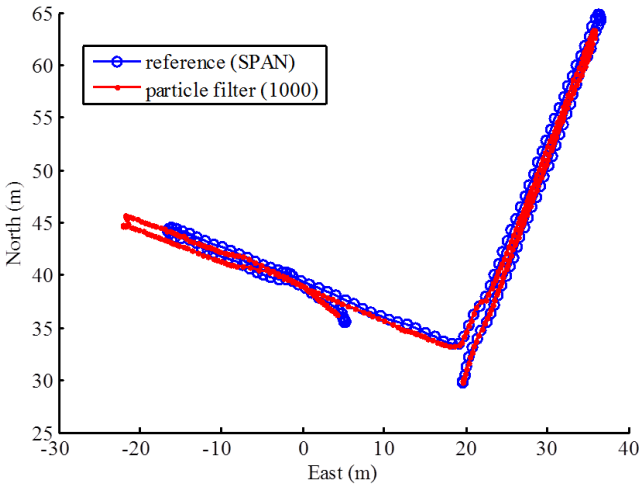

Figure 5: Particle filter fusion results compared with the SPAN reference trajectory in horizontal plane.

The standard deviation of the height observed using only the barometer was 0.15m, whereas using the sonar it was 0.01m and using the particle filter it was 0.03m. This result, represented also in Figure 6, show that the integration of absolute height measurements from a barometer and relative ones obtained using sonar improves the precision and therefore the stability of the height solution. However, the improvement in the precision is incremental and the barometer would be sufficient in a favorable environment and situation, when there are no changes in the temperature or pressure. The most important benefit from fusing the barometer height and sonar range measurements is that it will provide improved vertical accuracy and reliability in the case of unexpected changes in the environment's conditions.

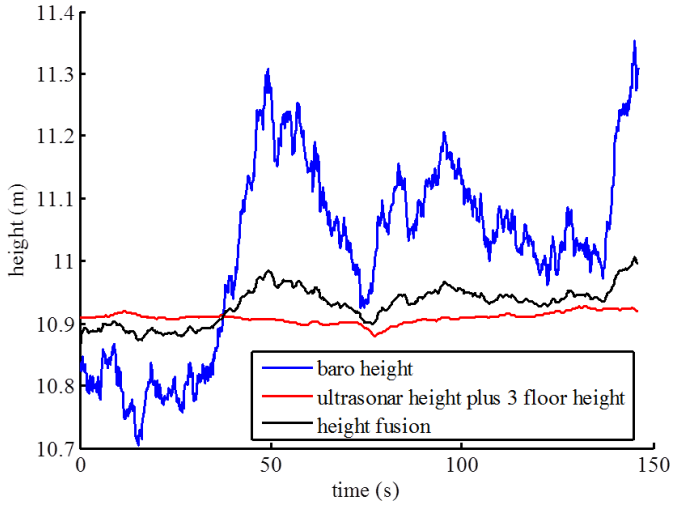

Figure 6: Barometer and sonar height and the fusion results

CONCLUSIONS

This paper discussed a multisensor integration method for obtaining accurate and reliable localization for a Simultaneous Localization and Mapping (SLAM) solution. Due to the nonlinear nature of pedestrian navigation, e.g. a soldier's or rescue person's motion, a particle filter is a suitable algorithm for fusing measurements from different sensors. Different types of measurements and error sources of visual measurements obtained using a monocular camera, a foot-mounted IMU, a barometer and a sonar were discussed. Better performance would be obtained by accurately modelling the error for each sensor, but in this paper zero mean Gaussian distributions were used for all measurement errors.

Finally, the implemented particle filter algorithm fusing all measurements was evaluated by collecting data in an indoor corridor environment. The horizontal accuracy of the localization solution was 3.14 m with standard deviation of 2.82 m. The result is comparable with state-of-the-art solutions but not adequate for the applications for which we are aiming. Therefore, the next step in our research is to carefully model the errors of each sensor and include the specific models into the fusion.

The height solution was also improved as a result of the fusion. Although a barometer would provide sufficient accuracy for most applications when the environment and situation are favorable, the advantage of using sonar measurements in the fusion is in the case of unexpected events, namely noticeable changes in pressure and temperature. When a barometer provides distorted height measurements, fusion of measurements from sonar, which is not affected by the aforementioned events, will provide the accuracy and reliability required.

Future work includes development of the SLAM algorithm to address also the mapping issues, namely the tracking of environmental features, loop closure, and map formation. Also, multiple IMUs will be used in future research, which is anticipated to result in a more accurate position solution as well as to provide information for context recognition purposes. When the results obtained in experiments of moderate duration in office corridors are satisfactory, the duration will be increased, and we will also tackle more challenging indoor environments, including multi-storey environments. Lastly, we aim to simulate various operational scenarios in order to ensure the results hold up well for the intended application of tactical situational awareness.

ACKNOWLEDGMENTS

This research has been conducted within the project INTACT (INfrastructure-free TACTical situational awareness), funded by the Scientific Advisory Board for Defence of the Finnish Ministry of Defence and the Finnish Geospatial Research Institute (FGI), Finland.

REFERENCES

Anderson, B.D.O. and Moore, J.B (1979) *Optimal Filtering*. Englewood Cliffs, New Jersey: Prentice–Hall.

Arulampalam, M. Maskell, S., Gordon, N. and T. Clapp (2002) "A tutorial on particle fiters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE trans. on Signal Processing*, Vol.50, Issue 2 , pp.174-188.

Bancroft, J. and G. Lachapelle (2012), "Estimating MEMS Gyroscope G-Sensitivity Errors in Foot Mounted Navigation", in *Proceedings of Ubiquitous Positioning, Indoor Navigation and Location Based Service*, 2-5 Oct, Helsinki, Finland.

Beauregard, S., and Haas, H. (2006) "Pedestrian dead reckoning: A basis for personal positioning," in *Proceedings of the 3rd Workshop on Positioning, Navigation and Communication*, Hannover, Germany, pp. 27–35.

Civera, J., Grasa, O., Davison A. and J. Montiel (2010) "1-Point RANSAC for EKF Filtering: Application to Real-Time Structure from Motion and Visual Odometry," in *Journal of Field Robotics - Visual Mapping and Navigation Outdoors*, Vol. 27, Issue 5, pp. 609-631.

Collin, J. (2006) *Investigations of self-contained sensors for personal navigation*, Ph.D. dissertation, Tampere University of Technology, Finland.

Davison, A., Reid, I., Molton, N. and O. Stasse (2007) "MonoSLAM: Real-Time Single Camera SLAM," in *Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, Issue 6, pp. 1052-1067.

Fischler, M. and R. Bolles (1981) “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography”, in *Communications of the ACM*, Vol. 6, Issue 24, pp. 381–395.

Foxlin, E. (2005) “Pedestrian tracking with shoe-mounted inertial sensors,” in *IEEE Computer Graphics and Applications*, Vol 25, Issue 6, pp. 38–46.

Gallagher, A. (2005) "Using vanishing points to correct camera rotation in images," in *Proceedings of IEEE 2nd Canadian Conference on Computer and Robot Vision*, Victoria, BC, Canada, May 9-11, pp.460-467.

Gelman, A.B., Carlin, J. S., Stern, H. S. and D. B. Dubin (2000) *Bayesian Data Analysis*, 2nd ed. Chapman & Hall.

Hol, J. D., T. B. Schön, and F. Gustafsson (2006), “On resampling algorithms for particle filters,” in *IEEE Nonlinear Statistical Signal Processing Workshop*, Cambridge, UK, Sep. 2006, pp. 79–82

Ijaz, F., Yang, H., Ahmad, A. and C. Lee (2013) "Indoor positioning: a review of indoor ultrasonic positioning systems," in *Advanced Communication Technology*.

Julier, S. Uhlmann, J. and H. Durrant-Whyte (2000) “A new method for the nonlinear transformation of means and covariances in filters and estimators,” in *IEEE Transactions on Automatic Control*, Vol. 45, Issue 3, pp. 477–482.

Klingbeil L., Reiner, R., Romanovas, M., Traechtler M. and Manoli Y. (2010) “Multi-modal Sensor Data and Information Fusion for Localization in Indoor Environments”, *in Proceedings of Workshop on Positioning Navigation and Communication*, Dresden, Germany, 11-12 March, pp. 187-192

Kuusniemi, H., Bhuiyan, M.Z.H., Ström, M., Söderholm, S., Jokitalo, T., Chen, L. and R. Chen (2012), “Utilizing pulsed pseudolites and high-sensitivity GPS for ubiquitous outdoor/indoor satellite navigation”, *in Proceedings of* IPIN,November, 13-15, Sydney, Australia.

Leland, R. (2005), “Mechanical-thermal noise in MEMS gyroscopes”, in *IEEE Sensors Journal*, Vol 5, Issue 3, pp. 493-500.

Leonard, J. and H. Durrant-Whyte (1992) *Directed Sonar Sensing for Mobile Robot Navigation*, Springer Science+Business Media, LLC.

Leppäkoski, H. (2015) *Novel Methods for Personal Indoor Positioning*. D.Sc. Thesis, Tampere University of Technology, Tampere, Finland.

Lowe, D. (1999) "Object recognition from local scale-invariant features," in *Proceedings of International Conference on Computer Vision*, Corfu, Greece, Sep. 20-25, pp. 1150-1157.

Parviainen, J., Kantola, J., & Collin, J. (2008) “Differential barometry in personal navigation”. In *Proceedings of IEEE/ION Position, Location and Navigation Symposium*, pp. 148–152, Monterey, CA, USA.

Skog, I., Handel, P., Nilsson, J.O., and J. Rantakokko (2010) “Zero-Velocity Detection—an Algorithm Evaluation”, in *IEEE Transactions on Biomedical Engineering*, Vol 57, Issue 11, pp. 2657 – 2666.

Rantakokko J., Rydell, J., Stromback, P., Handel, P., Callmer, J., Tornqvist, D., Gustafsson, F., Jobs, M. and Gruden, M. (2011), “Accurate and reliable soldier and first responder indoor positioning: multisensor systems and cooperative localization”, in *IEEE Wireless Communications*, Vol. 18, Issue 2, pp. 10 – 18.

Ruotsalainen L., Gröhn, S., Kirkko-Jaakkola M., Chen L., Guinness, R. and H. Kuusniemi (2015) “Monocular Visual SLAM for Tactical Situational Awareness”, In *Proceedings of the IPIN*, 13-16 October, Banff, Canada, in press.

Ruotsalainen, L. (2013) *Vision-Aided Pedestrian Navigation for Challenging GNSS Environments*, vol. 151, Doctoral Dissertation. Publications of the Finnish Geodetic Institute.

Ruotsalainen L., Bancroft J., Kuusniemi H., Lachapelle G. and R. Chen (2012) “Utilizing Visual Measurements for Obtaining Robust Attitude and Positioning for Pedestrians”, in *Proceedings of ION GNSS*, September 17-21, Nashville, TN.

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

Titterton, D. H., and Weston, J. L. (2004) *Strapdown Inertial Navigation Technology*, second edition, American Institute of Aeronautics and Astronautics, Reston, VA, USA.

Zhuang and H. and Z. Roth (1996) *Camera-aided robot calibration*, CRC Press Inc.