ORIGINAL ARTICLE

A two-dimensional pedestrian navigation solution aided with a visual gyroscope and a visual odometer

Laura Ruotsalainen · Heidi Kuusniemi · Mohammad Zahidul H. Bhuiyan · Liang Chen · Ruizhi Chen

Abstract Integration of different positioning systems such as wireless local area networks (WLAN) and INS has proven to be the most promising approach for navigation in global navigation satellite system challenging environments. However, the integrated solution suffers from the errors induced by the sensors in the INS and the low availability and infrastructure dependency of the WLAN. Visual aiding is a complementary method for augmenting these systems, because it suffers from errors and availability problems of a different nature. We introduce the concept of a "visual gyroscope" and a "visual odometer," based on recovering user information by tracking the feature motion between consecutive images. All calculations are of sufficiently low complexity to be adoptable for navigation with current smartphones. The camera orientation is observed using the vanishing point locations, and this information is transformed into user heading change information and also used for visual odometer calculations. The visual odometer retrieves the camera translation information based on a special camera configuration and the motion of the matched scale-invariant feature transform features. The performance of both of these tools is evaluated. Experiments in selected environments have proven that visual aiding with a visual gyroscope and visual odometer improves the two-dimensional navigation solution significantly.

Introduction

The positioning accuracy of consumer-grade global navigation satellite system (GNSS) receivers is around 1–10 m when there is an unobstructed line-of-sight to the satellites but decreases tremendously, or the positioning becomes even impossible, when the user is in an urban canyon or indoors. In these challenged GNSS environments, the user absolute position may be obtained with other radio navigation systems like wireless local area networks (WLAN), Bluetooth, or radio frequency identification (RFID). These systems need a priori prepared infrastructure and are therefore restricted to certain places, and their availability is too low in some environments to serve the needs of pedestrian navigation, depending on the number of access points available. When the user initial absolute position is known, the position may be propagated using relative positioning approaches, like self-contained sensors. The propagated position may then be used to augment the position measurements obtained with GNSS or other radio sensors for more accurate and available positions, or even for short-time stand-alone navigation. Honeywell's GLANSER is an example of systems for first responders utilizing the technology (Hawkinson et al. 2012). Fusion of different positioning systems and related error detection is an active research area (Moafipoor 2009; Moafipoor et al. 2012).

The most commonly used self-contained sensors in pedestrian navigation are digital compasses for measuring the user heading, gyroscopes for heading changes, and

L. Ruotsalainen (✉) · H. Kuusniemi ·
M. Z. H. Bhuiyan · L. Chen · R. Chen
Department of Navigation and Positioning,
Finnish Geodetic Institute, Masala, Finland
e-mail: laura.ruotsalainen@fgi.fi

accelerometers for speed. When these measurements are used as inputs to pedestrian dead-reckoning (PDR) algorithms or integrated with absolute position measurements using a Kalman filter, the position is obtained continuously despite the degradation of GNSS. However, self-contained sensors suffer from errors that may decrease the position accuracy substantially, especially when consumer-grade micro-electro-mechanical (MEMS) sensors are used. Gyroscopes provide the user heading rate, but they suffer from drift which result in heading errors that increase with time. The accelerometers in a smartphone are too erroneous to be used to obtain the user speed without augmentation with, for example, GNSS or using calibration and special algorithms (Susi et al. 2011; Pei et al. 2011).

The errors in indoor positioning may be mitigated using information obtained from consecutive images (Veth and Raquet 2007, Fletcher et al. 2007). When a pedestrian is capturing images with a smartphone in hand, the motion of the features in the images may be transformed into information about the user motion. The visual motion information is not affected by the same error sources as GNSS and self-contained sensors, and is therefore a complementary information source for augmenting the position measurements. Visual aiding increases the accuracy, availability, continuity, and integrity of the navigation solution.

The challenges of visual-aiding indoors are the varying lighting conditions of the environment and the low amount of distinctive features. Besides, when the camera motion is observed by following the motion of the features in the consecutive images, the image processing method has to be able to exclude the dynamic objects in the scene from disturbing the motion perception.

We introduce a twofold system where the consecutive images are used to derive user heading change information and the camera pitch, namely a visual gyroscope, and the user speed information, a visual odometer. The visual gyroscope is based on following the vanishing point location change in the image. The vanishing point is a point in the image where parallel lines in the scene seem to intersect. This phenomena are introduced by the perspective projection mapping of the three-dimensional objects in the scene into two-dimensional features in the image (Hartley and Zisserman 2003). Indoor and urban environments are constructed in a way that their structures constitute a three-dimensional grid (Coughlan and Yuille 1999) containing straight parallel lines and therefore the method based on the vanishing points is suitable for the environment which otherwise may be poor in features (Kessler et al. 2010). The method overcomes also the two other problems that disturb visual aiding: the lines are found from images taken in low lighting conditions and the straight lines are seldom found from dynamic objects in the scene, reducing the

possibility of mistakenly following a moving object. The vanishing point location is related to the camera orientation, arising only from the rotation of the camera and not affected by the translation. It may therefore be used for calculating the heading change and pitch of the camera. The deficiency of the visual gyroscope is its inability to observe the heading change in sharp turns and, therefore, it has to be used in parallel with other systems providing the direction.

The visual odometer is based on finding the image points corresponding to the same object in consecutive images. These matched points may then be used to find the camera rotation and translation between the images. Because the camera rotation has already been observed using the vanishing points, the number of matching image points required for obtaining the camera translation information, and therefore the number of features in the scene is decreased making the method feasible for indoor environments. The two most challenging aspects in resolving the translation from the consecutive images are the depth of the objects followed and the ambiguity problem, which are solved with a special camera configuration (Campbell et al. 2005; Ruotsalainen 2012). The method presented needs information only about the camera height. All calculations are sufficiently simple to be adopted for navigation with a smartphone.

We introduce first the methods used for obtaining the visual gyroscope and odometer measurements, presented earlier in (Ruotsalainen 2012). Next, the performance of the visual gyroscope and the visual odometer is examined. Then, the navigation solution formation by integrating visual gyroscope and odometer measurements with those of GNSS, WLAN, and self-contained sensors using a Kalman filter is explained. Finally, test results with the integrated and stand-alone visual navigation systems are presented and examined.

Visual gyroscope and odometer

The visual gyroscope and the visual odometer are used to obtain the user heading change and the translation. The visual gyroscope functions autonomously, and no calibration is needed besides the optional camera calibration. When the camera is assumed static in the hand of the user navigating, facing forwards, and at a known height, the heading change and translation of the camera may be transformed into the user heading change and translation.

The method presented is advantageous for smartphone navigation because of its ability to monitor the pitch of the camera continuously. No assumption about the camera being fixed is needed. The method does not need any preliminary information of the environment, and because

the camera rotation is calculated independently from the translation, the required number of features in the scene is reduced.

The principle of the visual gyroscope

The urban scenes consist mostly of straight lines in three orthogonal directions. The process of mapping the three-dimensional scene into the two-dimensional image is called projective transformation. The transformation preserves the straight lines, but parallel lines in the scene do not stay parallel, but seem to intersect. The lines in three orthogonal directions form three intersection points are called vanishing points. The vanishing point in the propagation direction (z-axis) is called the central vanishing point. The three vanishing point locations are defined by the camera orientation with respect to the scene and the camera intrinsic parameters. This relation is described with $\mathbf{V} = \mathbf{KR}$, where $\mathbf{V}$ is the vanishing point location matrix $[\mathbf{v}_x\ \mathbf{v}_y\ \mathbf{v}_z]$, $\mathbf{K}$ is the calibration matrix containing the camera intrinsic parameters, namely the focal length, principal point, and skew, and $\mathbf{R}$ is the camera rotation matrix (Gallagher 2005).

The intrinsic parameters may be defined by calibrating the camera. It is adequate for a pedestrian navigation system to calibrate the camera once and assume the parameters do not change. When the navigation solution accuracy may be compromised for the sake of adaptability, the focal length may be found from the image's exchangeable image file (EXIF) data, and the principal point may be assumed to be the image central point. The calibration matrix with zero skew is described as

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & u \\ 0 & f_y & v \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

The rotation matrix $\mathbf{R}$ is the identity when all axes of the camera are aligned with the three-dimensional grid of the scene, namely the walls, floor, and ceiling. In this configuration, the central vanishing point $\mathbf{v}_z$ lies at the principal point and the other two vanishing points in infinity on the x and y image axes. When the camera is rotated by changing the heading θ degrees and the pitch toward the floor plane ϕ degrees, the rotation matrix has the form

$$\mathbf{R} = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ \sin\phi\sin\theta & \cos\phi & -\sin\phi\cos\theta \\ -\cos\phi\sin\theta & \sin\phi & \cos\phi\cos\theta \end{bmatrix} \tag{2}$$

Roll is assumed to be zero which is justified by restricting at this point the camera rotation to heading and pitch. Extension of the method to account for roll requires the calculation of either vertical or horizontal vanishing point

and is straightforward with incremental effect on the accuracy (Ruotsalainen et al. 2012). When the calibration and rotation matrices are as explained above, the heading change (θ) and pitch (ϕ) may be obtained from the central vanishing location point as

$$\mathbf{v}_z = \begin{bmatrix} f_x\sin\theta + u\cos\phi\cos\theta \\ -f_y\sin\phi\cos\theta + v\cos\phi\cos\theta \\ \cos\phi\cos\theta \end{bmatrix} \tag{3}$$

In order to find the central vanishing point, it is necessary to identify the image of the straight lines in the propagation direction. This is done using the Hough Lines algorithm (Hough 1962). Because the heading change and pitch calculations are done using the central vanishing point, vertical and horizontal lines are excluded from the calculations. The central vanishing point is found by using a voting scheme, namely each vanishing point candidate is voted for by all the lines found, and the point that gets most of the votes, in this case, the intersection point of most of the lines is selected as the correct one.

The principle of the visual odometer

The principle of obtaining the camera translation between two images is based on looking at the image point motion of a static object. The relation between two points found in consecutive images is called homography and can be written as

$$\mathbf{x}' = \mathbf{Rx} + \mathbf{t}/Z \tag{4}$$

where $\mathbf{x}' = (x',\ y',\ 1)$ is the image point in the second image and $\mathbf{x} = (x,\ y,\ 1)$ in the first. The points are in a normalized form, meaning that the camera intrinsic parameters are considered (Hartley and Zisserman 2003). $\mathbf{R}$ is a matrix expressing the camera rotation between the images, accommodating also the pitch, $\mathbf{t} = [t_x\ t_y\ t_z]^{\mathrm{T}}$ is the camera translation between the images, and Z is the object's distance from the camera. The rotation matrix $\mathbf{R}$ may be formulated with the visual gyroscope measurements, but the camera translation is also dependent of the distance of the object, whose image points are observed.

The distance to the object may be resolved by using objects with known sizes (Santos et al. 2009) or by using a stereo camera and triangulation (Jirawimut et al. 2002). The first of the mentioned approaches is restricted to an area that is prepared beforehand, and the latter needs special equipment. Campbell et al. (2005) developed an outdoor robot navigation system using a special camera configuration to resolve the distance problem. They used optical flow calculations for finding the camera rotation and translation. The method presented follows theirs, but is further developed for indoor use and for a smartphone.

Measuring the distance of an object from the camera

The distance to the object from the camera is calculated using information of the camera height (h), the focal length in units of vertical pixels (f_y), and the image height in pixels (H). The focal length and image height may be read from the image's EXIF file or from the calibration results. With these parameters and the configuration shown in Fig. 1, the distance to the object may be obtained.

The vertical field-of-view (*vfov*) may be calculated using the known image height (H) and the vertical focal length (f_y) component with

$$vfov = \arctan\left(\frac{H}{f_y}\right) \qquad (5)$$

Thereby, the angle (β) between the principal ray of the camera and the ray from camera to the object may be calculated with

$$\beta = \arctan\left(\left(\frac{2y}{H} - 1\right)\tan\left(\frac{vfov}{2}\right)\right) \qquad (6)$$

Finally, using β and the camera pitch ϕ, the distance Z is obtained as

$$Z = \frac{h\cos(\beta)}{\sin(\phi + \beta)} \qquad (7)$$

This configuration requires the object to lie on the floor and in close vicinity of the camera, in the region between the camera and the intersection point of the principal ray and the floor plane. The vicinity requirement is rational also in the sense that the motion of the far objects is very small in terms of pixels and may therefore be overwhelmed by noise. The floor plane is recovered using the information obtained from the visual gyroscope calculations. The lines found for the vanishing point calculations are usually found from the floor, especially with a camera with a pitch angle larger than zero toward the floor plane (Ruotsalainen 2012). The image points employed for the visual odometer calculations are required to lie close to the lines used for successful vanishing point derivation. If no such points are
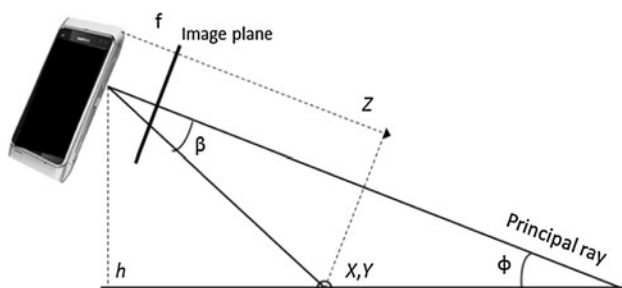
found, a coarser method is introduced, considering all points found below the vanishing point, or if the vanishing point is not found, the principal point. When the image points followed are projections of the objects lying on the floor, the translation matrix y component shows the translation in the propagation direction and the x component the sideway translation. Because the points lying on the same plane are used only to resolve these two components, the degeneracy problem arising when trying to recover all motion parameters using planar image points is avoided (Torr et al. 1999).

The two image points in consecutive images representing the same object are found with a procedure called matching. The method presented matches the SIFT descriptors (scale-invariant feature transform) of the points (Lowe 1999). Due to the low amount of features in indoor environments, even less certain matches are accepted. The loose matching criteria, as well as the occasional use of the coarse floor plane recovery, necessitate careful error handling. Another issue that the homography-based motion deprivation brings about is the ambiguity problem related to the magnitude of the translation.

Error detection and ambiguity resolving for the visual odometer

The image point **x**, presented with homogenous coordinates (x, y, 1), is related to the object coordinates $\mathbf{X} = (X, Y, Z, 1)$ as

$$\mathbf{x} = \mathbf{K}[\mathbf{R}|\mathbf{t}_{cw}]\mathbf{X} \qquad (8)$$

where **R** is the camera rotation matrix and $\mathbf{t}_{cw} = -\mathbf{RC}$ is the camera center transformation into world coordinates. When two image points of the same object are obtained, the object coordinates may be calculated with the Iterative-Eigen method using singular value decomposition (SVD) (Hartley and Sturn 1997). In this case, $\mathbf{t}_{cw}$ incorporates only camera height for the first image and translation between images for the second. The translation ambiguity may be resolved by comparing the known camera height to the object Y coordinate (Kitt et al. 2011). Points with translation or Y coordinate values deviating more than the threshold from the mean values of all observations are considered erroneous and discarded.

Performance of the visual gyroscope and the visual odometer

In this section, the performance and major error causes of the visual gyroscope and odometer as well as methods to avoid propagating the errors to the navigation solution are discussed.

Fig. 1 The special camera configuration for resolving the object distance (Z), using the camera height (h) and pitch (ϕ)

Visual gyroscope performance

The visual gyroscope does not provide any absolute heading value and must therefore be integrated with measurements from other sources. It cannot be used during sharp turns, when the visibility to building boundaries forming lines is lost. Therefore, the visual heading needs to be augmented with heading measurements obtained using another system, for example, a gyroscope, a magnetometer, or a floor plan.

Low lighting of the navigation environment reduces the number of lines found from the image, possibly resulting in erroneous vanishing point location. If all lines are found from the same image side and their slopes have the same sign, their intersection points do not usually fall to the correct vanishing point as shown in Fig. 2. The Hough Line algorithm parameters are adjusted so that lines shorter than a threshold are left out of the computation to reduce the number of non-parallel lines disturbing the calculation. An optimal threshold was found by experimenting. When the scene consists of a plane, there are no lines in the image and the vanishing point cannot be calculated.

Rotation measurements obtained with erroneous vanishing point locations distort the navigation solution. The success probability in calculating the vanishing point correctly, that is, reliability, may be estimated based on monitoring the line geometry. The calculated visual measurements are tagged with an ordinal value {0, 0.1, 0.5, 1}. Value 0 is assigned, when there are no lines, only one line, or no intersecting lines and thus no vanishing point is found. A value of 0.1 is assigned in the situation presented in Fig. 2, where all lines are on the same image side and have similar slopes. When the heading change angle is larger than a threshold, the user is either turning or the calculation is erroneous, in both cases the measurement

Fig. 3 An image from the validation process showing how a false vanishing point location (*red dot*) was obtained

should not be wholly trusted. Measurements assigned with a value of 0.5 are calculated with lines on only one image side and with similar slopes. When the lines used for calculations are either on different image sides or their slopes have different signs, the measurement is trusted, and a value of 1 is assigned. These values are accommodated in the navigation solution formation; the smaller the value, the less weight is assigned to the heading change and speed measurements.

The error detection feasibility was verified using approximately 1,000 images. The vanishing point locations assigned with the value 1 were verified manually. In challenging visual environments, as for example in shopping malls and outdoor environments with many dynamic objects and restricted view to line features, the obtained success rate was 75 %. In an office environment, the success rate obtained was 93 %. Figure 3 shows an example of obtaining an incorrect vanishing point. In the image, the building edges and the road are blurred by snow and sand, and heavy shadows resulted from the glaring sunlight introduce many lines that are not parallel to the propagation direction.

The accuracy of the heading change and the pitch were evaluated with a static camera in an office environment. The mean error obtained for the heading change was 0.8° and for the pitch 0.3° in an environment with changing light as well as dynamic objects in the scene sometimes encompassing the view totally. Table 1 shows the statistics for the 7,555 images taken in a 2.5-h time span.

Noise in the visual gyroscope

The most significant source affecting the gyroscope accuracy is the drift. The Allan variance analysis method (Allan 1966) was originally developed for the oscillator stability study, but because it is suitable for the study of any

Fig. 2 Erroneous vanishing point location (shown with *red dot*) due to poor line geometry

Table 1 Statistics for heading change measurement performance with 7,555 images and a static camera

| Statistics	Min error	Max error	Mean error	SD of error	CEP
Heading change	0	18.4	0.8	0.6	0.8
Pitch	0.0	10.7	0.3	0.3	0.3

All values are in degrees

instrument, it is applied here to evaluate the camera gyroscope noise level. The Allan variance $\sigma_C^2(t_A)$ equation (Kirkko-Jaakkola et al. 2012), modified here for the camera gyroscope is

$$\sigma_C^2(t_A) = \frac{1}{2(N-1)} \sum (\tilde{y}(t_A)_{k+1} - \tilde{y}(t_A)_k)^2 \qquad (9)$$

where $\tilde{y}(t_A)_k$ is the average value of a bin containing the heading change and pitch values for an integration time t_A. N is the number of bins for the integration time at issue. The variation in heading change and pitch measurements was calculated from the 7,555 images taken. The associated Allan deviation plot is shown in Fig. 4. The figure shows the uncorrelated noise affecting the visual gyroscope stability for the short integration times. After the deviation has reached a minimum value, the rate random walk starts to increase the deviation again. The bias instability measure may be found from the minimum value, and is 0.058 degrees/s for the heading, and 0.045 degrees/s for the pitch, at the integration time of 245 s.

The test showed also the method's tolerance to dynamic objects, also seen from Fig. 5. The maximum errors in the heading angle and pitch were, however, substantial due to dynamic objects obscuring the scene almost totally.

The visual gyroscope errors are time invariant. Hence, one erroneous measurement does not necessarily introduce
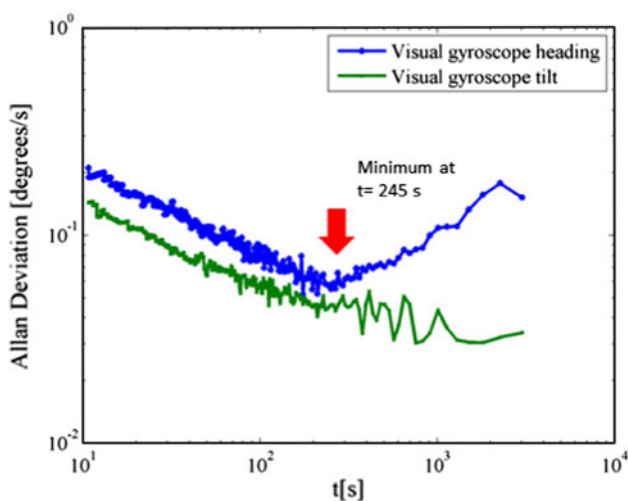
Fig. 4 Allan deviation of the visual gyroscope

Fig. 5 Calculation of the central vanishing point (*red dot*) is largely tolerant to dynamic objects in the scene

drift in the propagated heading value if it is identified by error detection.

Performance of the visual odometer

After resolving the camera intrinsic parameters, the visual odometer does not need calibration before or during navigation. It does not depend on any knowledge of the environment, but only the camera height must be estimated. The most drastic errors may be avoided by monitoring the changes in pitch; if the change is considerable it is most likely due to an error in vanishing point location and in this case the previous pitch and heading values are used. The method of the visual odometer is not as tolerant to dynamic objects as the visual gyroscope. The mean error of the user speed obtained in different navigation environments is approximately 0.25 m/s.

Visually aided two-dimensional navigation solution

Information obtained from consecutive images is always relative, namely speed and heading changes between the two images. For navigation purposes, this relative information has to be integrated with at least the initial position and heading, preferably also with measurements from other sensors. Integration of several radio positioning and self-contained sensors augmented with visual measurements has shown improvement in the navigation solution accuracy, availability, and continuity (Kuusniemi et al. 2011).

Kalman filtering is a tool for integration and propagating the position result from a stand-alone system. This section introduces two Kalman filters used: the Kalman filter integrating visual measurements with measurements from different sensors and radio positioning sources (Kuusniemi et al. (2011), and the Kalman filter providing a navigation solution by propagating the visual measurements when no other positioning systems are available.

Kalman filter for integration

The pedestrian positioning model used herein is a constant speed model defined as

$$
\begin{aligned}
X_{k+1} &= X_k + \dot{X}_k \Delta t + w_1 \\
Y_{k+1} &= Y_k + \dot{Y}_k \Delta t + w_2 \\
\dot{X}_{k+1} &= \dot{X}_k + w_3 \\
\dot{Y}_{k+1} &= \dot{Y}_k + w_4
\end{aligned}
\tag{10}
$$

where X and Y are the latitude and longitude, respectively, transformed into the metric ENU (East, North, Up) coordinate frame, $\dot{X}$ and $\dot{Y}$ are their time derivatives, k denotes the current epoch, Δt is the time interval between two epochs, and w_i is the state uncertainty component of the element i. The state vector for the model is $\mathbf{x}_k = \begin{bmatrix} X & Y & \dot{X} & \dot{Y} \end{bmatrix}_k^T$, the state model in the Kalman filter is defined as $\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_k$, and the measurement model $\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k$. The state model propagates the state using a transition matrix $\mathbf{F}$, and the measurement model relates the measurement with the state using a measurement matrix $\mathbf{H}$. The process noise $\mathbf{w}_k$ is distributed as $\mathbf{w}_k \sim N(0, \mathbf{Q}_k)$ and the measurement noise $\mathbf{v}_k \sim N(0, \mathbf{R}_k)$, where $\mathbf{Q}_k$ is the process noise and $\mathbf{R}_k$ is the measurement noise matrix. The $\mathbf{F}$ and the $\mathbf{Q}$ matrices are

$$
\mathbf{F}_k =
\begin{bmatrix}
1 & 0 & \Delta t & 0 \\
0 & 1 & 0 & \Delta t \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\tag{11}
$$

$$
\mathbf{Q}_k =
\begin{bmatrix}
\tilde{q}_1 \frac{\Delta t^4}{4} & 0 & \tilde{q}_1 \frac{\Delta t^3}{2} & 0 \\
0 & \tilde{q}_2 \frac{\Delta t^4}{4} & 0 & \tilde{q}_2 \frac{\Delta t^3}{2} \\
\tilde{q}_1 \frac{\Delta t^3}{2} & 0 & \tilde{q}_1 \Delta t^2 & 0 \\
0 & \tilde{q}_2 \frac{\Delta t^3}{2} & 0 & \tilde{q}_2 \Delta t^2
\end{bmatrix}
\tag{12}
$$

where $\tilde{q}_i$ are the spectral density vector elements: the $\tilde{q}_1$ is a spectral density value for the North component and $\tilde{q}_2$ is the spectral density for the East component, chosen based on empirical assessment. The measurement vector $\mathbf{z}$ is

$$
\mathbf{z}_k =
\begin{bmatrix}
X_{\mathrm{GPS}} \\
Y_{\mathrm{GPS}} \\
X_{\mathrm{WLAN}} \\
Y_{\mathrm{WLAN}} \\
S_{\mathrm{ACC1}} \cos \theta_{\mathrm{DC1}} \\
S_{\mathrm{ACC1}} \sin \theta_{\mathrm{DC1}} \\
S_{\mathrm{ACCV}} \cos \theta_V \\
S_{\mathrm{ACCV}} \sin \theta_V \\
S_{\mathrm{ACC2}} \cos \theta_{\mathrm{DC2}} \\
S_{\mathrm{ACC2}} \sin \theta_{\mathrm{DC2}}
\end{bmatrix}
\tag{13}
$$

where the subscript *GPS* stands for the GPS measurements, *WLAN* for the WLAN measurements, *ACC1* and *DC1* are

the speed and heading measurements presented in the experiments section, *ACCV* and *V* relate to the speed and heading obtained with visual gyroscope and odometer, and *ACC2* and *DC2* are the speed and heading obtained with a Nokia N8 mobile phone.

The measurement matrix $\mathbf{H}$ is

$$
\mathbf{H}_k =
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\tag{14}
$$

and finally, the measurement covariance matrix $\mathbf{R}$ is

$$
\begin{aligned}
\mathbf{R}_k = \mathrm{diag}(&\sigma^2_{X_{\mathrm{GPS}}}, \sigma^2_{Y_{\mathrm{GPS}}}, \sigma^2_{X_{\mathrm{WLAN}}}, \sigma^2_{Y_{\mathrm{WLAN}}}, \sigma^2_{S_{\mathrm{ACC1}} \cos \theta_{\mathrm{DC1}}}, \sigma^2_{S_{\mathrm{ACC1}} \sin \theta_{\mathrm{DC1}}}, \\
&\sigma^2_{S_{\mathrm{ACCV}} \cos \theta_V}, \sigma^2_{S_{\mathrm{ACCV}} \sin \theta_V}, \sigma^2_{S_{\mathrm{ACC2}} \cos \theta_{\mathrm{DC2}}}, \sigma^2_{S_{\mathrm{ACC2}} \cos \theta_{\mathrm{DC2}}})
\end{aligned}
\tag{15}
$$

The values applied in the covariance matrix were chosen by empirical assessment and depend on the performance level of the GPS receiver, WLAN infrastructure and fingerprinting density, and the quality of the self-contained sensors.

Not all measurements are available at all time instants, epochs, and thus the measurement vector and matrix dimensions vary. The WLAN has a much lower update rate than GPS and the other sensors. Time synchronization is crucial in real-time applications, but since this analysis was conducted in post-processing, by suitable interpolation, the time discrepancy can be omitted. The magnetometers and accelerometers are utilized in a pedestrian dead-reckoning manner: the accelerometers (ACC1 and ACC2) were utilized for obtaining pedestrian speed by assessing the acceleration patterns (step detection) and the magnetometers (DC1 and DC2) to obtain the heading with respect to the true North (magnetic declination accounted for). The visual gyro and odometer were not used to calibrate the self-contained sensors at this stage—that is part of future work.

Kalman filter for visual position propagation

A Kalman filter was used for propagating the visual heading change and speed measurements accommodating simultaneously the measurement credibility. The filter in this case is very simple, and its purpose is to robustly propagate the visual measurements in the absence of other position measurements. The pedestrian positioning model used is

$$X_{k+1} = X_k + S_{k+1}\Delta t \sin\theta_{k+1} + w_1$$
$$Y_{k+1} = Y_k + S_{k+1}\Delta t \cos\theta_{k+1} + w_2 \qquad (16)$$

where X and Y are easting and northing, respectively, scaled into meters, S is the user speed, θ is the heading change, k denotes the current epoch, Δt is the time interval between two epochs, and w_i is the state uncertainty component of the element i. The state vector for the model is $\mathbf{x}_k = [XY]_k^T$. The $\mathbf{F}$, $\mathbf{H}$, and $\mathbf{Q}$ matrices for the filter are

$$\mathbf{F} = \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{Q} = \begin{bmatrix} 2^2 & 0 \\ 0 & 2^2 \end{bmatrix} \qquad (17)$$

The variances in $\mathbf{R}_k = \mathrm{diag}\left(\sigma_X^2, \sigma_Y^2\right)$ are set based on the trustworthiness values assigned for the visual gyroscope measurements.

Experimental results

The performance of the visual augmented multi-sensor multi-network integrated solution in a conventional office building was introduced with detailed test results in Ruotsalainen (2012). We present further test results in more challenging environments, namely an office building with an outdated WLAN radio map, a shopping center, and an outdoor environment. All tests use a NovAtel SPAN (synchronized position attitude navigation) GPS/INS high-accuracy positioning system as reference. The images were taken with a Nokia N8 smartphone camera and have approximately a 1.2-s time interval between images. The visual odometer-derived speed measurements were filtered by cutting the measurements larger than 1.5 m/s, which is considered a reasonable assumption for normal pedestrian navigation.

Test in indoor office environment with outdated WLAN radio map

Performance of the navigation solution using Kalman filtering to integrate GPS, WLAN, and sensor measurements with visual measurements was tested in an office environment with an outdated WLAN radio map. The motivation for the test was to see the improvement on the navigation solution provided by visual aiding in case of a vulnerable setup receiving only rarely absolute position measurements for calibrating the position. Due to the inability of the visual gyroscope to evaluate the heading change angle magnitude during sharp turns, these were detected from the visual odometer measurements. The only situation when the visual odometer did not find matching image points in consecutive images is when the magnitude

Fig. 6 Test equipment. The Nokia N8 phone acquiring the images was attached to the holder in the front of the cart

of a turn is almost 90° or larger. The number of images with no matching points was monitored and the degree of the turn was thereby evaluated based on the monitoring result. In future work, these situations will be managed with the use of a gyroscope.

Test setup

The test setup consisted of a Fastrax IT500 high-sensitivity GPS receiver, a Nokia N8 mobile phone for WLAN positioning, and a multi-sensor positioning (MSP) device comprising of a 3-axis VTI accelerometer and 2-axis Honeywell compass. The MSP device is explained in (Kuusniemi et al. 2012). The equipment was placed in a cart and the results were post-processed using Matlab. The test equipment is shown in Fig. 6.

A major weakness of the WLAN fingerprinting procedure is its vulnerability to the environment changes. In the test discussed, two access points in the office were out of order, and two had changed locations. Also, some new electrical equipment was placed to the vicinity of one access point. Thus, this altered setup caused the average WLAN positioning accuracy to be reduced to 11 m from the previous 6 m.

Test results

Stand-alone performance of the visual odometer was evaluated. The cumulative distance obtained with the visual odometer is presented in Fig. 7, being 173 m with the ground truth distance of 162 m. The speed measurements are presented in Fig. 8. The speed mean error
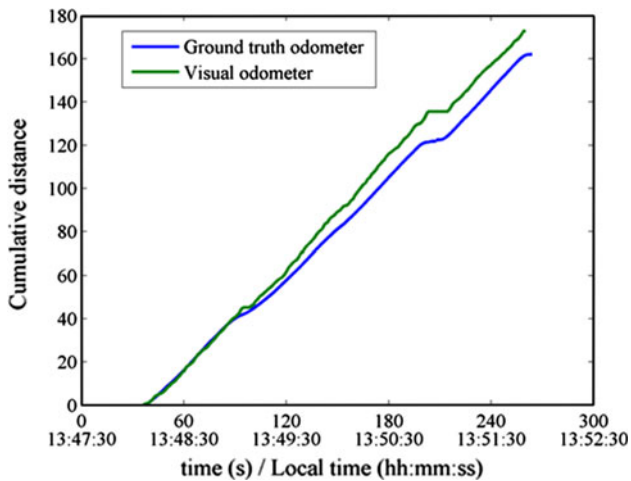
Fig. 7 Cumulative distance travelled obtained with visual odometer (*green*) and reference (*blue*)

Fig. 8 User speed obtained with visual odometer (*blue*) compared with the reference speed (*red*)

obtained with the 183 images used is 0.26 m/s and the statistics are presented in Table 2.

Figure 9 shows the positioning results obtained with different systems; the fused solution does not have visual aiding. Figure 10 shows the same combinations with the difference that the fused solution is visual aided. Visual aiding improves the position solution significantly, decreasing the fused solution mean error to 5.8 m from the 7.8 m, as may be seen from Table 3 showing the positioning error statistics.

The experiments conducted with the method using a fully functional, updated WLAN fingerprinting database were presented in (Ruotsalainen 2012). The experiments showed a mean error of 0.3 m/s for the visual odometer speed, and a standard deviation of 0.3 m/s. The reason for larger speed error compared with the experiments presented here is that the errors in measurements were not filtered as explained above. The visual measurements were fused with measurements from other positioning systems similarly as in the experiment presented herein. The mean error for WLAN positioning alone was 5.9 m, GPS alone 17.8 m, fused without visual aiding 6.7 m, and visual aided fused 5.3 m.

Test in a shopping mall environment

The method was tested in a shopping mall environment with many dynamic objects (shoppers), degraded geometry due to wide corridors restricting in most images the view of the corridor sides, varying lighting conditions, and various objects forming many non-parallel lines. Due to the absence of an absolute positioning system in the environment, the visual gyroscope and odometer were tested as a stand-alone visual system in the *Iso Omena* shopping center in Espoo, Finland. The accuracy of the position
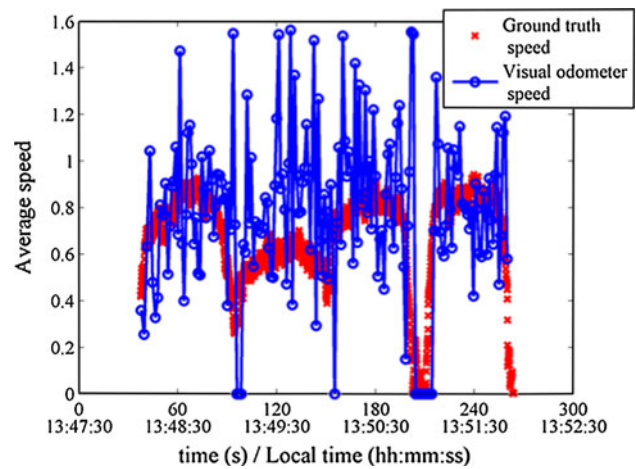
Table 2 Statistics of the visual odometer speed

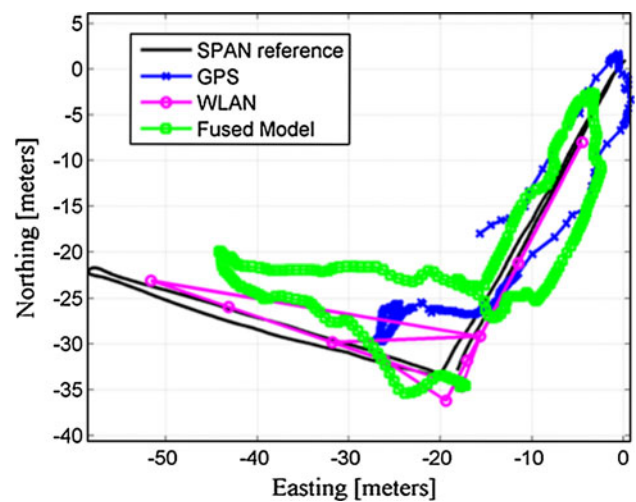| Statistics	Min error	Max error	Mean error	SD of error
	0.002	1.48	0.26	0.24

Units are in m/s

Fig. 9 Example 1 Indoor navigation solutions obtained with different positioning systems {reference (*black*), GPS only (*blue*), WLAN only (*purple*), and fused solution without visual aiding (*green*)}

solution in the test environment is expected to increase substantially when the visual measurements are integrated with other radio positioning and sensor measurements.

Due to the challenges set by the environment, the visual gyroscope performance decreased significantly: the heading mean error was 4.4°, with a 0.2° standard deviation, and the maximum errors reached 32°. The visual odometer performance did not suffer as heavily, the mean error in speed being 0.25 m/s, the standard deviation 0.2 m/s, and the maximum error 0.9 m/s. The cumulative distance
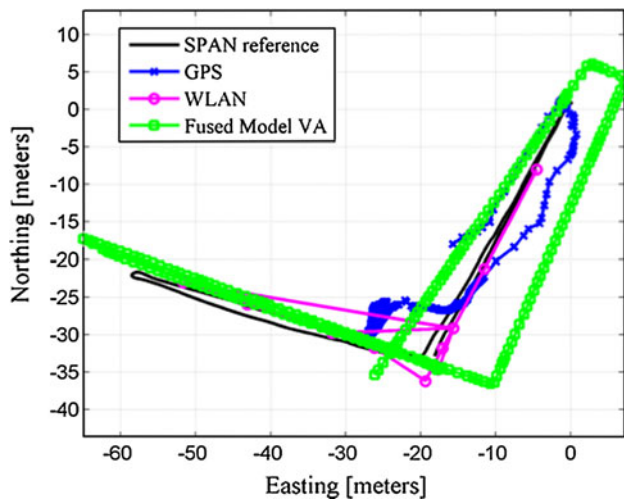
Fig. 10 Example 2: Indoor navigation solution obtained with different positioning systems {reference (*black*), GPS only (*blue*), WLAN only (*purple*), and visual-aided fused solution (*green*)}
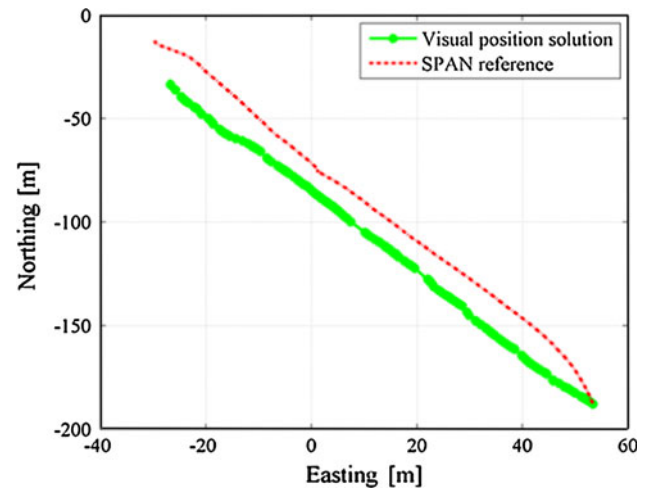

Fig. 11 The two-dimensional position solution in the shopping center with visual stand-alone solution (*green*) and SPAN reference (*red*)

Table 3 Positioning error statistics in meters

Statistics	Min error	Max error	Mean error	SD of error	Timely availability (%)
WLAN	0.5	36.9	11.0	10.9	11
GPS	0.3	32.6	14.9	8.4	100
Fused	0.5	16.1	7.8	4.8	100
Fused with visual aiding	0.4	13.1	5.8	3.7	100

obtained in the route was 179 m in the 198-m long true path, yielding an agreement of 90 %. The position obtained with the visual heading change and speed measurements was propagated using the Kalman filter presented in (16) and (17). Figure 11 shows the two-dimensional position obtained. The position mean error was 14 m in the 198 m route inside the shopping mall (Fig. 2), the standard deviation 6.2 m, and the maximum error 23.2 m.

Test in an outdoor environment

Outdoor performance was tested in the close vicinity of the shopping center wall, as shown in Fig. 3. The test is presented as a stand-alone visual solution, because of large errors in the initial GPS position when exiting the mall and the lack of other positioning systems. The results are compared with the position solution obtained with the Fastrax IT500 high-sensitivity GPS receiver. Figure 12 shows the two-dimensional navigation solution. The visual solution is initialized with the true initial position and heading obtained with the reference system. The position is immediately in the right track, but drifts slowly due to erroneous measurements. The GPS solution shown in Fig. 13 takes tens of seconds to attain the correct position, inducing a maximum error of 51 m, but is accurate thereafter.
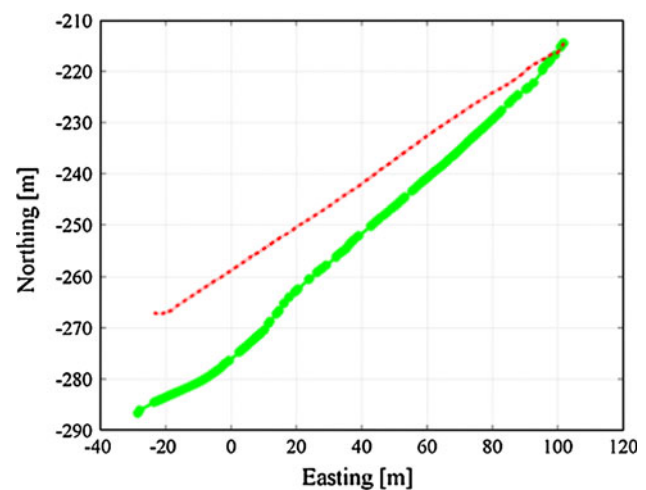

Fig. 12 Outdoor position solutions obtained with visual stand-alone system (*green*) and SPAN reference (*red*)

The most significant errors were caused by the reduced view of lines due to sand and snow on the road, and non-parallel lines present in some images. The mean error of the visual gyroscope's heading change was 3.3° and the mean error in the visual odometer speed 0.2 m/s. The cumulative distance obtained was 131 m in the 146-m long true path, yielding an agreement of 90 %. The visual measurements were propagated using the Kalman filter (16
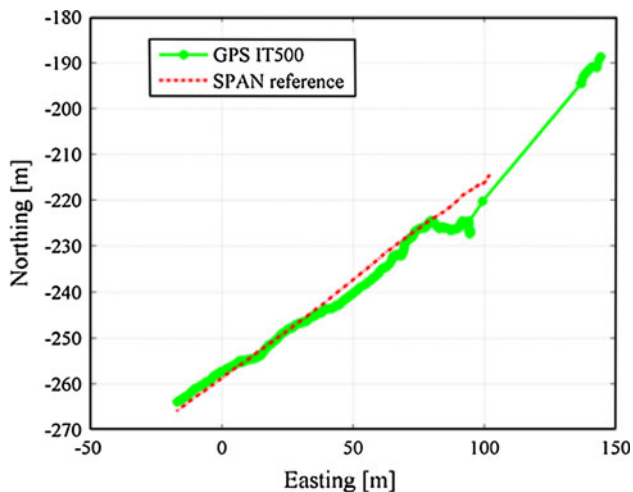
Fig. 13 Outdoor position solutions obtained with GPS (*green*) and SPAN reference (*red*)

Table 4 Error statistics for outdoor positioning. Units are in meters

Statistics	Min error	Max error	Mean error	SD of error
Visual stand-alone system	0.3	22.5	10.3	5.8
GPS	0.4	51.0	16.7	10.1

and 17). The mean errors of the visual stand-alone position and GPS solutions were 10.3 and 16.7 m, respectively. The position error statistics are presented for both systems in Table 4.

Conclusions

A visual gyroscope and visual odometer intended for augmenting other navigation sensors in GNSS challenging environments to improve pedestrian navigation accuracy, availability, continuity, and integrity were introduced. The performance analysis and experiments in different environments showed promising results for the visual gyroscope and odometer accuracy, as well as improvement of the integrated navigation solution. The average mean heading change error was 0.8° with a static test, though it arose to 4° in the most challenging environment, a shopping mall with wide corridors and many dynamic objects. The error detection algorithm of the vanishing point calculations increases the system robustness. The mean error in the visual odometer speed measurements was under 0.3 m/s for all test environments. While the accuracy of the visual gyroscope and odometer presented is not sufficient for positioning alone, integration of the visual measurements with observations from other positioning systems

resulted in a significant improvement in the navigation solution performance. Future work includes development of better estimation methods for visual measurements as well as fault detection. The accuracy of the integrated solution will be improved using map matching and more suitable integration algorithms, for example, particle filters. All algorithms were designed to be implementable in current smartphones; however, the results presented were computed by post-processing the data in Matlab. Present visual gyroscope implementation in a smartphone environment performs the calculations for one image with 0.5 Hz rate. Future work includes visual odometer smartphone implementation. The method developed is advantageous for visual aiding the navigation solution in all demanding GNSS environments.

References

Allan DW (1966) Statistics of atomic frequency standards. Proceedings of the IEEE, vol 54, no. 2, pp 221–230

Campbell J, Sukthankar R, Nourbakhsh I, Pahwa A (2005) A robust visual odometry and precipice detection system using consumer-grade monocular vision. Proceedings of the IEEE international conference on robotics and automation (ICRA), pp 3421–3427

Coughlan J, Yuille AL (1999) Manhattan world: compass direction from a single image by Bayesian inference. Proceedings of the international conference on computer vision, Corfu, Greece, September, pp 941–947

Fletcher J, Veth M, Raquet J (2007) Real-time fusion of image and inertial sensors for navigation, Proceedings of the 63rd annual meeting of the Institute of Navigation, Cambridge, MA, April, pp 534–544

Gallagher AC (2005) Using vanishing points to correct camera rotation in images. Proceedings of the 2nd Canadian conference on computer and robot vision, May 9–11, IEEE Xplore, pp 460–467. doi:10.1109/CRV.2005.84

Hartley RI, Sturn P (1997) Triangulation. Comput Vis Image Underst 68(2):146–157

Hartley R, Zisserman A (2003) Multiple view geometry in computer vision, 2nd edn. Cambridge University Press, Cambridge

Hawkinson W, Samanant, P et al (2012) GLANSER: geospatial location, accountability, and navigation system for emergency responders—system concept and performance assessment. Proceedings of the IEEE/ION position location and navigation

symposium (PLANS), Minneapolis, MN, USA, April 23–26, pp 98–105

Hough PVC (1962) Method and means for recognizing complex patterns. US Patent 3069654

Jirawimut R, Prakoonwit S, Cecelja F, Balachandran W (2002) Visual odometer for pedestrian navigation. Proceedings of the IEEE instrumentation and measurement technology conference, Anchorage, AK, USA, May 21–23, pp 43–48

Kessler C, Ascher C, Frietsch N, Weinmann M, Trommer G (2010) Vision-based attitude estimation for indoor navigation using vanishing points and lines. Proceedings of the IEEE/ION PLANS 2010, pp 310–318

Kirkko-Jaakkola M, Collin J, Takala J (2012) Bias prediction for MEMS gyroscopes. IEEE Sens J 12(6):2157–2163

Kitt B, Rehder J, Chambers A, Schönbein M, Lategahn H, Singh S (2011) Monocular visual odometry using a planar road model to solve scale ambiguity. Proceedings of the 5th European conference on mobile robots (ECMR2011), pp 43–48

Kuusniemi H, Chen L, Ruotsalainen L, Pei L, Chen Y, Chen R (2011) Multi-sensor multi-network seamless positioning with visual aiding. Proceedings of the ICL-GNSS 2011, IEEE Xplore, June 29–30, Tampere, Finland, pp 146–151. doi:10.1109/ICL-GNSS.2011.5955257

Kuusniemi H, Liu J, Pei L, Chen Y, Chen L, Chen R (2012) Reliability considerations of multi-sensor multi-network pedestrian navigation. IET Radar Sonar Navig 6(3):157–164

Lowe DG (1999) Object recognition from local scale-invariant features. Proceedings of the international conference on computer vision, Corfu, Greece, September, pp 1150–1157

Moafipoor S (2009), Intelligent personal navigation supported by knowledge-based systems for estimating dead reckoning navigation parameters. Dissertation, The Ohio State University

Moafipoor S, Grejner-Brzezinska DA, Toth CK (2012) Real-time integrity monitoring of a dead reckoning personal navigator using a two-stage neural Kalman filter. J Navig 65(4):635–649. doi:10.1017/S0373463312000240

Pei L, Chen R, Liu J, Kuusniemi H, Chen Y, Tenhunen T (2011) Using motion-awareness for the 3D indoor personal navigation on a Smartphone. Proceedings of the ION GNSS 2011, Portland, OR, September 19–23, pp 2906–2913

Ruotsalainen L (2012) Visual Gyroscope and Odometer for Pedestrian Indoor Navigation with a Smartphone. Proceedings of the ION GNSS 2012, Nashville, TN, USA, September 17–21 (in press)

Ruotsalainen L, Bancroft J, Lachapelle G (2012) Mitigation of attitude and gyro errors through vision aiding. Proceedings IPIN 2012, Sydney, Australia, November 13–15, IEEE (in press)

Santos A, Tarrataca L, Cardoso JMP (2009) An analysis of navigation algorithms for smartphones using J2ME. Proceedings of the MOBILWARE, Berlin, Germany, April 28–29, Springer, pp 266–279

Susi M, Renaudin V, Lachapelle G (2011). Detection of quasi-static instants from handheld MEMS devices. Proceedings of the international conference on indoor positioning and indoor navigation IEEE/IPIN 2011, Guimaraes, Portugal, September 21–23

Torr P, Fitzgibbon A, Zisserman A (1999) The problem of degeneracy in structure and motion recovery from uncalibrated image sequences. Int J Comput Vis 32(1):27–44

Veth M, Raquet J (2007) Fusion of low-cost imaging and inertial sensors for navigation, technical report, 0704-0188, Air Force Institute of Technology, OH, 11 pp

Author Biographies

Laura Ruotsalainen is a Research Scientist at the Department of Navigation and Positioning at the Finnish Geodetic Institute and a PhD candidate in the Department of Computer Systems, Tampere University of Technology, Finland. Her doctoral research is focused on visual-aided seamless indoor/outdoor navigation.

Heidi Kuusniemi is a Research Manager at the Department of Navigation and Positioning at the Finnish Geodetic Institute. She received her DSc(Tech.) degree in 2005 from the Tampere University of Technology, Finland. Now, her research interests cover various aspects of GNSS navigation. She is the President of the Nordic Institute of Navigation since 2011.

Mohammad Zahidul H. Bhuiyan received his PhD degree in 2011 from the Tampere University of Technology, Finland. He joined the Department of Navigation and Positioning at the Finnish Geodetic Institute in 2011 as a Senior Research Scientist with research interests covering various aspects of GNSS receiver design and sensor fusion algorithms for seamless outdoor/indoor positioning.

Liang Chen is a Senior Research Scientist at the Department of Navigation and Positioning at the Finnish Geodetic Institute. He holds a postdoctoral researcher position at the Academy of Finland from 2011 to 2014. He received his PhD from Southeast University, China, in 2009. His research interests include statistical signal processing for positioning, wireless positioning using signals of opportunity and sensor fusion.

Ruizhi Chen is the Head of the Department of Navigation and Positioning at the Finnish Geodetic Institute. He is an editor of one book, author of 4 book chapters and more than 130 scientific articles. Mobile geospatial computing is his current research interest, especially in the areas of ubiquitous positioning in smartphones, mobile location-based services, and smartphone-based context awareness.