Toward the use of smartphones for mobile mapping

Andrea Masiero, Francesca Fissore, Francesco Pirotti, Alberto Guarnieri & Antonio Vettore

Published online: 08 Oct 2016.

Submit your article to this journal ⬀

View related articles ⬀

View Crossmark data ⬀

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS

Toward the use of smartphones for mobile mapping

Andrea Masiero [iD], Francesca Fissore [iD], Francesco Pirotti [iD], Alberto Guarnieri [iD] and Antonio Vettore [iD]

Interdepartmental Research Center of Geomatics (CIRGEO), University of Padova, Padova, Italy

ABSTRACT

This paper considers the use of a low cost mobile device in order to develop a mobile mapping system (MMS), which exploits only sensors embedded in the device. The goal is to make this MMS usable and reliable even in difficult environments (e.g. emergency conditions, when also WiFi connection might not work). For this aim, a navigation system able to deal with the unavailability of the GNSS (e.g. indoors) is proposed first. Positioning is achieved by a pedestrian dead reckoning approach, i.e. a specific particle filter has been designed to enable good position estimations by a small number of particles (e.g. 100). This specific characteristic enables its real time use on the standard mobile devices. Then, 3D reconstruction of the scene can be achieved by processing multiple images acquired with the standard camera embedded in the device. As most of the vision-based 3D reconstruction systems are recently proposed in the literature, this work considers the use of structure from motion to estimate the geometrical structure of the scene. The detail level of the reconstructed scene is clearly related to the number of images processed by the reconstruction system. However, the execution of a 3D reconstruction algorithm on a mobile device imposes several restrictions due to the limited amount of available energy and computing power. This consideration motivates the search for new methods to obtain similar results with less computational cost. This paper proposes a novel method for feature matching, which allows increasing the number of correctly matched features between two images according to our simulations and can make the matching process more robust.

1. Introduction

Thanks to the continuous increase of applications using geo-spatial data (Guarnieri et al. 2015; Habib et al. 2005; Pirotti et al. 2015; Remondino, Guarnieri, and Vettore 2005), in the last decades several mobile mapping systems (MMSs) have been developed, mostly based on the use of terrestrial or airborne vehicles (Chiang, Noureldin, and El-Sheimy 2003; Kraus and Pfeifer 1998; Pirotti et al. 2014; Remondino et al. 2011; Toth 2001; Toth and Grejner-Brzezinska 1997), equipped with remote sensing instruments such as laser scanners and cameras.

MMSs have become quite popular even among the general public due to the success of web tools which allow street view navigation. Despite the popularity of such applications, it is worth to notice that the acquired georeferenced spatial data can be used in definitely wider range of applications, in the real time case (e.g. location-based services) and the post-processed one (e.g. by a Geographic Information System (GIS) (El-Sheimy and Schwarz 1998; Hadeel, Jabbar, and Chen 2011; Piragnolo et al. 2015; Tao 2013), and for recognition purposes (Facco, Masiero, and Beghi 2013; Facco et al. 2011; Jaakkola et al. 2010; Pfeifer, Gorte, and Winterhalder 2004)).

The diffusion of MMSs has been limited until now by their quite high cost, mostly due to the need for quite expensive sensors (e.g. terrestrial laser scanners) and vehicles (e.g. cars). However, motivated by the world-wide capillary diffusion of mobile devices (e.g. smartphones, tablets) embedded with both positioning (e.g. GNSS, inertial sensors) and remote sensing instruments (e.g. camera), several efforts have been recently spent in order to develop mobile MMSs based on smartphones.

Two great advantages are showcased with the development of an MMS based on the use of a smartphone – the much lower cost comparing with other MMSs, and the much wider diffusion of these devices, which represents a potentially very large customer base. However, several challenging issues are related to the realization of such a system: first, the limited amount of available energy imposes stringent requirements on the system power consumption (and hence a restriction to the available computational resources) and/or on its battery autonomy. Furthermore, the current generation of smartphones is typically embedded with several MEMS sensors (Schiavone, Desmulliez, and Walton 2014), which, however, typically provide quite noisy measurements (e.g. positioning obtained by means of only

CONTACT Andrea Masiero ✉ masiero@dei.unipd.it

inertial sensors is unreliable). For instance, although several works in the literature recently have tried to tackle the positioning problem by MEMS sensor measurements when the GNSS signal is not available (or not reliable, as in certain city centers, or indoors), it is still a challenging problem (Chen, Meng, et al. 2015; Chen, Zou, et al. 2015; Bahl and Padmanabhan 2000; Huang and Gao 2013; Lukianto and Sternberg 2011; Saeedi, Moussa, and El-Sheimy 2014).

The aim of this paper is twofold. First, it aims at providing a new solution for the indoor positioning problem when the GNSS signal is not reliable (Piras, Marucco, and Charqane 2010). More specifically, this paper improves the positioning approach proposed in Masiero, Pirotti, et al. (2014). Here the positioning problem is tackled by means of the use of a pedestrian dead reckoning approach, referring to the study of Widyawan et al. (2012), and, similarly to the previous study (Masiero, Pirotti, et al. 2014), a computationally efficient particle filter is used (it requires a quite low number of particles, e.g. ~100). The proposed improvement is to make it more effective in a wide range of conditions of interest (e.g. in emergency conditions, during firefighter intervention), in particular, a new movement mode detection method is proposed, and an altitude estimation based on the use of the barometer.

The standard camera embedded in the smartphone is used as a remote sensing device providing images to be processed in order to obtain a 3D reconstruction of the scene (or of the object of interest). Thus, the goal is to allow 3D reconstruction based on the solution of the Structure from Motion (SfM) problem directly on the mobile device. The overall reconstruction algorithm can be summarized with the following steps: feature extraction and matching, reconstruction of the geometry of the scene, and computation of a dense 3D point cloud. Since the execution of such computations on the device can drastically reduce the battery life, reducing the power consumption needed for computing the 3D reconstruction is of fundamental importance. Several approaches have been recently considered in order to efficiently compute the solution of the SfM problem given a set of images (Agarwal et al. 2010; Brand 2002; Byröd and Åström 2010).

Independently of the specific adopted algorithm, it is clear that the accuracy and the computational complexity of the 3D reconstruction is closely related to the quality of the candidate matching features. In order to improve the results of the feature matching steps, this paper considers the use of an alternative feature description, similar to affine scale-invariant feature transform (affine SIFT, or ASIFT) (Morel and Yu 2011), that takes advantage of the information provided by the navigation system in order to improve the feature matching ability of the system, while simultaneously reducing the computational burden required in the ASIFT approach. According to the results shown in Section 5, the proposed method allows to increase the number of correctly matched features with respect to the standard SIFT (i.e. with respect to the state of art).

2. System description

Most of the navigation systems on the market exploit the use of the GPS/GNSS signal, however, since this solution is not reliable in indoor environments in this paper an alternative navigation procedure will be considered: the proposed solution (that integrates information provided by a three-axis accelerometer, a three-axis magnetometer, and a barometer) is specifically designed to be usable in indoor environments (i.e. when GNSS is not reliable), however it can be used outdoors as well (as a stand-alone navigation system or integrated with the GNSS positioning and/or with WiFi information). Despite not considered in the minimum requirements, a three-axis gyroscope can be considered as well, in order to provide more reliable estimations of the change of the device orientation and of the heading direction. The geometrical characteristics of the building are assumed to be pre-loaded on the navigation device before starting the navigation algorithm.

In order to make the use of the system as simple (and comfortable) as possible to the user, the device is assumed to be hand held, and differently from most of the previously proposed systems for pedestrian navigation, the use of external sensors (e.g. Foxlin 2005; Ruiz et al. 2012; Widyawan et al. 2012) is not considered.

Remote sensing ability is achieved by using the standard camera sensor embedded in the smartphone. In this work, a SfM approach is used in order to provide 3D reconstruction of the environment. The accuracy and robustness of the obtained results is usually related to the ability of the 3D reconstruction algorithm to provide a (possibly) large number of good matches between features in different images: this typically ease both the estimation of the scene structure and the dense reconstruction. Motivated by these considerations, in Section 4 a novel method is proposed in order to improve SIFT (scale-invariant feature transform) feature matching and, consequently, the overall 3D reconstruction.

The proposed method is actually based on a similar rational of the ASIFT method (Morel and Yu 2009, 2011), however, it improves the computational efficiency of the ASIFT method taking into account of the information provided by the inertial measurement unit (IMU) and/or by the navigation system. Furthermore, the method requires an approximate knowledge of the value of the intrinsic camera calibration parameters (this information is usually available from the operative system of the device).

The system has been developed in the Android environment and the results provided in this paper have been obtained by a Huawei Sonic U8650 (Figure 1) and an

Figure 1. Smartphone (Huawei Sonic U8650) coordinate system (u_s, v_s, w_s).

LG Google Nexus 5. Notice that, despite (since now) the system has been developed only for Android, this is not a limitation of the proposed approach, which actually can be considered for other operative systems (and devices) as well.

3. Navigation

Since in terrestrial applications often the height with respect to the floor is of minor interest, in this section the navigation problem will be separated in estimating movements on a planar map and on the vertical direction (Section 3.1). The positioning system that will be described in the following of this section is an evolution of that presented in Masiero, Pirotti, et al. (2014). The most significant differences are related to the estimation of the movements along the vertical direction (Section 3.1), and to the detection of different movement modes (Section 3.2).

The rationale of the positioning system is that of using a dead reckoning-like approach (Foxlin 2005; Ruiz et al. 2012): detect the human steps by means of a proper analysis of the accelerometer measurements (Jahn et al. 2010), then the combined use of magnetometer and accelerometer (and gyroscope, if available) measurements allows to estimate the movement direction with respect to the North (Bonnet et al. 2009).

Let (u_t, v_t, w_t) be the 3D position of the device (e.g. smartphone), expressed with respect to the North, East, and vertical directions (i.e. global reference system), before the t-th step, then

$$\begin{bmatrix} u_{t+1} \\ v_{t+1} \end{bmatrix} = \begin{bmatrix} u_t \\ v_t \end{bmatrix} + s_t \begin{bmatrix} \sin \alpha_t \\ \cos \alpha_t \end{bmatrix} \qquad (1)$$

where s_t is the length of the t-th step, and α_t is the corresponding heading direction. Notice that an estimation of the initial 3D position (u_0, v_0, w_0) is assumed to be a priori available.

The step length s_t is estimated by properly combining (in a linear estimation fashion) the current values of the following variables: the acceleration peak difference, the average of the acceleration absolute values in the time interval related to the considered step, the time duration of the step, and their inverse values. The weighting parameters in the considered estimator are computed on a learning data-set. More details on the considered variables can be found in Jahn et al. (2010). Alternatively, s_t can be fixed to a constant value (an approximation of the mean step length): the tracking algorithm proposed in Masiero, Pirotti, et al. (2014) and summarized in this section is designed to compensate (relatively small) step length errors.

The mobile device is supposed to be carried by the user's hand and the heading direction is assumed to be *approximately* fixed with respect to the local coordinate system (u_s, v_s, w_s), i.e. the user does not drastically change the device orientation during the navigation. The tracking system is designed to estimate and correct device attitude changes, with absolute value lower than 36 degrees, with respect to the conventional orientation. Allowing free changes of the device orientation can be achieved by generalizing the initial heading estimation procedure proposed by Masiero, Pirotti, et al. (2014), or as done by Deng et al. (2015).

Let y_t be the vector of measurements corresponding to the t-th step and Y_t be the collection of measurements y_τ from $\tau = 0$ to $t-1$. Furthermore, let $q_t = [u_t \ v_t]^T$, then the probability distribution of the estimated position q_t after the $(t-1)$-th step is expressed as follows:

$$p(q_t|Y_t) = \sum_{i=1}^{n} w_{i,t}\delta(q_t - q_{i,t}) \qquad (2)$$

where $q_{i,t}$ and $w_{i,t} = 1/n$ are the position and weight of the i-th particle at t, while $\delta(\cdot)$ is the Dirac delta function.

Then, at the next user's step the above probability distribution is updated as follows:

- for each particle i:
- draw a sample $q_{i,t+1}$ from the proposal distribution:

$$q_{i,t+1} = q_{i,t} + s_{i,t} \begin{bmatrix} \sin \alpha_{i,t} \\ \cos \alpha_{i,t} \end{bmatrix} \qquad (3)$$

where the heading direction $\alpha_{i,t}$ and the step length $s_{i,t}$ are sampled from Gaussian distributions centered in α_t and $ks_t + b_t$, respectively. k and b_t are scalar variables that aim at reducing the effect of measurement errors, as shown in Masiero, Pirotti, et al. (2014).

(2) if $q_{i,t+1}$ violates the geometrical constraints of the building, then the last part of the particle trajectory is rotated of a random angle $\alpha_{b,t}$ ($|\alpha_{b,t}| \leq \pi/5$). Since most of the times, the violation is due to small deviations of the heading direction from its true value (e.g. measurement errors due to small calibration errors),

small rotations of the last part of the trajectory can often tackle this issue.

(3) if $q_{i,t+1}$ still violates building geometrical constraints then set $w_{i,t+1} = 0$, otherwise set $w_{i,t+1} = 1/n$. When the WiFi connection is available, the weights can be computed taking into account also of the WiFi radio signal strength, as shown in Widyawan et al. (2012).

- Scale the particle weights in order to normalize their sum to 1.
- Resample n particles from the following

$$p\left(q_{t+1}|Y_{t+1}\right) \approx \sum_{i=1}^{n} w_{i,t+1}\delta(q_{t+1} - q_{i,t+1}) \quad (4)$$

and set $w_{i,t+1} = 1/n$.

Notice that the computational complexity of this particle filter is linear with respect to the number of particles n, and, more interestingly, it allows to achieve good positioning performance while using a small number of particles (e.g. $n \approx 100$).

The reader can refer to (Masiero, Pirotti, et al. 2014) for more details on the original version of the particle filter summarized above. The estimates obtained by this filter are integrated with those of the altitude described in Section 3.1 in order to obtain 3D estimations of the device position. 3D orientations can be obtained as well, by processing the IMU measurements with a Kalman filter.

3.1. Estimating variations of altitude

The variations of the altitude of the device are estimated in this subsection by measuring the variations of atmospheric pressure by means of a barometer embedded in the mobile device. Since atmospheric pressure changes with time and space, a fast calibration procedure is assumed to be performed at the beginning of the navigation. Furthermore, the working conditions are assumed to be invariant during the navigation, e.g. constant atmospheric pressure (and temperature): for instance, in ideal conditions (null measurement error) the same pressure value can be measured in the same spatial position at the beginning and at the end of the navigation. This assumption is clearly an approximation of the reality; however, it is reasonably good for the typical extent (in the spatial dimension and in the time duration) of pedestrian navigation.

According to the above assumptions (i.e. considering the atmosphere as an ideal gas at constant temperature), from Boyle and Stevino's laws the pressure p_w at an altitude w can be expressed as follows:

$$p_w = p_0 e^{-(w-w_0)/a} \quad (5)$$

where p_0 is the pressure at altitude w_0 (notice that in certain cases, for instance when changes of temperature

occur in the considered environment, a linear model between p_w and w can be considered as well).

At the beginning of the navigation procedure, p_0 is estimated as (the average value of) the pressure measurement(s) at the known altitude w_0.

If certain environmental variables are known (e.g. temperature, georeferenced position, and corresponding value of the gravitational acceleration), then the value of the parameter a can be analytically computed. However, in order to reduce the measurement errors due to sensor calibration, the following simple procedure is adopted: since w_0 is often expressed with respect to the ground and the initial position is assumed to be known, then a can be computed as the best fitting value in Equation (5) by measuring p_w with w corresponding to the ground altitude.

The goal of the above procedure for the estimation of p_0 and a is to be as simple (and fast) as possible for the user. However, it is worth to notice that a more robust estimation of such parameters should be adopted if possible (i.e. varying the altitude during the calibration procedure on all the range of the expected values of interest during the navigation).

Once the parameters p_0 and a have been computed, the variation of altitude with respect to w_0 can be estimated as follows:

$$\left(w - w_0\right) = a \log \frac{p_0}{p_w} \quad (6)$$

When an accurate estimation of w is required, a Kalman filter can be implemented as well to exploit the temporal smoothness of the device movements in order to reduce the influence of measurement errors.

3.2. Detection of movement mode

The rationale of this subsection is that of improving the positioning estimation by providing information about the current action of the user. For instance, step lengths are typically different when walking on the stairs with respect to walking on a corridor: hence with such information the positioning algorithm can easily adapt the step length according to the zone where the user is currently moving. Taking into account of the above considerations, in the first part of this subsection a method for detecting several user's movement modes will be presented. Then, Equation (4) will be updated in order to exploit the movement mode information.

To be more specific, the aim of this section is that of presenting a method for detecting five actions typically related to moving inside of a building: going up or down on stairs, and going in the up or down direction with a lift, walking on a floor.

A support vector machine (SVM) approach has been used in order to detect and correctly classify such actions. Measurements provided by the barometer and by the accelerometer are used as input for the classifying

machines: indeed, 5 SVMs are used, where each SVM is used to recognize one specific action.

All the SVMs use the same data input. A linear conversion is used to properly relate the change of pressure measured by the barometer with the change of the device altitude. Then, a Δt interval of measurements is considered, and the following variables are provided to the SVM classifiers as input:

- The mean altitude variation in the Δt interval (the value of this variable is obtained by a linear fit of the data in the considered interval).
- The standard deviation (in the Δt interval) of the absolute value of the measured acceleration vector.

The mean altitude variation in Δt often allows to discriminate each of the cases from a device moving on the same floor. Indeed, a significant (positive/negative) change on its value corresponds to a device movement (also) in the vertical (up/down) direction. Furthermore, its value is usually (approximately) fixed to a constant value when the user is not moving (e.g. as quite usual in a lift), whereas it can assume quite different values for different humans walking on stairs.

Despite such variable can often be successfully used to detect the actions of interest here, its reliability strongly depends on the length of the time interval Δt: computing the mean over a longer time interval allows to significantly reduce the influence of the measurement noise and of the human movements, whereas for short time intervals such factors can lead to wrong classification results.

The standard deviation (in the Δt interval) of the absolute value of the measured acceleration vector has been considered as well in order to improve the classification results: in particular it can significantly improve the discrimination between going up/down with a lift or on the stairs: indeed, human (and consequently device) movements are usually quite limited while being in a lift, hence (excluding the initial lift acceleration and the final deceleration) the acceleration measured by the sensor is mostly similar to the gravitational acceleration. Instead, while walking the acceleration measured by the device is subject to significant changes due to the human steps: its mean value might be not so different from the gravitational acceleration, but its standard deviation (due to the acceleration changes due to the steps) results to be usually much larger than in the lift case.

Thanks to the use of the movement mode detector presented above, and by exploiting the law of total probability, the position estimation Equation (4) can be updated as follows:

$$p\left(q_{t+1}|Y_{t+1}\right) = \sum_{j} p\left(q_{t+1}|Y_{t+1}, \theta_{t+1,j}\right) p\left(\theta_{t+1,j}|Y_{t+1}\right)$$
(7)

where $\theta_{t+1,j}$ indicates the detection of the j-th mode at time $t+1$, whereas $p\left(\theta_{t+1,j}|Y_{t+1}\right)$ is the probability of

mode j given the current measurements. As shown in Section 5, the movement mode detector has low probability error, hence the detected mode is often the correct one, i.e. $p\left(\theta_{t+1,j}|Y_{t+1}\right)$ is close to 1 for the correct value of j. Furthermore, $p\left(q_{t+1}|Y_{t+1}\right)$ is formally defined similarly to Equation (4). The main difference with respect to (4) is that in $p\left(q_{t+1}|Y_{t+1}\right)$ the parameter values can be particularized with respect to the specific case of interest, e.g. the mean step length while walking on a corridor is typically quite different with respect to that while going up/down on the stairs.

4. 3D reconstruction

In the proposed system, 3D reconstruction of the scene is obtained by means of SfM approach (Hartley and Zisserman 2003). To be more specific, the reconstruction procedure can be summarized as follows:

- Computation and matching of feature points in the acquired images.
- Solution of the SfM problem on the matched features (i.e. bundle adjustment (Agarwal et al. 2010), or incremental SVD method (Brand 2002; Masiero, Vettore, et al. 2014)).
- Dense point cloud computation (this is based on dense pixel matching in different images (Furukawa and Ponce 2010) and triangulation for computing the corresponding 3D positions (Hartley and Zisserman 2003; Masiero and Cenedese 2012)).

In particular, this section deals with the improvement of the first part of the reconstruction procedure, the feature matching step, while the other steps are performed with standard algorithms.

Feature matching is based on the appearance of the 2D image regions in the neighborhood of the considered feature locations. Since images are taken from different point of views, the same feature can undergo certain appearance changes, then the goal of several matching techniques recently proposed in the literature is that of extracting features invariant to such deformations. As widely known, SIFT descriptors (Lowe 1999) allow to obtain reliable matching between features in two images when the images are related by a different scaling, illumination and rotation on the image plane (see Figure 2(b)). However, matching issues can occur when in presence of different changes of the point of view between the two images (e.g. rotations along the other two axes as in Figure 2(c) and (d)). As shown in Morel and Yu (2009), (2011), in this case local changes between features in the two images can be usually well represented by means of affine transformations.

In order to make the SIFT descriptor robust even to these kind of transformations, Morel and Yu (2009), (2011) proposed to simulate $N = 32$ versions of each of the original images, where each version simulates the

Figure 2. Examples of image views obtained by rotating the camera around the object of interest. (a) original camera pose. (b) rotating along the optical axis of the original camera pose. (c) rotating the camera toward the right of the object. (d) rotating the camera toward the top of the object.

appearance of the corresponding image after applying a specific affine transformation. The N affine transformations can be associated to different values of the rotation angles, i.e. they represent a discretization of the set of possible rotations that provide variant results to the SIFT descriptors (combinations of rotations as in Figure 2(c) and (d)). Then, feature points are extracted in all the N versions of the images. Thus, when searching for feature matching between image I_1 and I_2, all the corresponding SIFT descriptors are matched: all the possible combinations of matchings are checked, i.e. the features extracted in each of the N versions of I_1 are matched (when possible, according to the standard SIFT matching procedure, (Lowe 1999) with the features extracted in each version of I_2. In accordance with the type of transformations applied to each image, this procedure is named affine SIFT. Applying the feature matching procedure (Lowe 1999) to each couple of transformed images the computational time for feature matching is proportional to N^2 times that of the original SIFT.

Taking into account of the above considerations, the goal of this section is that of exploiting the information provided by the navigation system in order to improve the performance of the SIFT-based feature matching but at a lower computational cost with respect to the affine SIFT.

First, in order to make the use of system as simple as possible for the user, the camera embedded in the device is assumed to be uncalibrated (see (Karel and Pfeifer 2009; Ma et al. 2004; Remondino and Fraser 2006) for camera calibration and its advantages). However, in most of the cases it can be approximated by a pinhole camera. Then, notice that despite the real value of the interior parameter matrix K is unknown, it can be roughly approximated as follows:

$$K = \begin{bmatrix} a & 0 & u_0 \\ 0 & a & v_0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (8)$$

where pixels are assumed to be approximately squares, sensor axes are assumed to be orthogonal, and the displacement of the sensor center with respect to the optical

axis is approximated with $(u_0, v_0) \approx (c/2, r/2)$, where r and c are the number of image rows and of columns, respectively. The parameter a is related to the focal length and to the pixel size. When the characteristics of the device are available, the value of a can be quite easily approximated. When no information on such characteristics is available, the procedure described in the following section shall be repeated for different values of a ranging in the following interval $(1/3(r + c), 3(r + c))$ (Fusiello and Irsara 2010; Heyden and Pollefeys 2005), where the real value of a is supposed to be the value of a that allows to obtain the largest number of matching features.

Furthermore, let R_t be the matrix provided by the navigation system describing the device orientation during the acquisition of the image at time t. Since the approximate orientation matrix R_t is assumed to be available for all the acquired images, then, for each couple of acquired images I_1 and I_2, the corresponding orientation matrices and interior parameter matrix K are used in order to compute the *approximate rectification* of I_1 with respect to I_2, i.e. I_1 is transformed in order to simulate the view from the same orientation of I_2. After such transformation two corresponding features in the two images should have approximately the same orientation, hence the SIFT descriptor is modified in order to use absolute orientation angles instead of relative ones, i.e. invariance with respect to (large) rotations is now an undesired condition.

Since the navigation system provides also an estimation of the device position at each time interval, for each couple of matched features it is possible to compute also an approximation of the corresponding scales: however, doing this for all the couples of candidate matched features might be time consuming, hence it might be optionally considered just for an (almost) final list of matched features.

Finally, exploiting the information provided by the navigation system and the approximate interior parameter matrix K, an approximate fundamental matrix can be computed, and an approximate epipolar constraint can be imposed in order to reduce the number of candidate feature matchings. However, it is worth to notice that the fundamental matrix computed in this way is typically a quite rough approximation of the correct one, hence the

approximate epipolar constraint should be applied with a threshold significantly larger than 0.

Interestingly, this approach allows to perform the feature matching procedure with a constant computational burden (with respect to N), thus significantly reducing the computational complexity with respect to the affine SIFT. The performance (in terms of correctly matched features) of the proposed matching feature procedure will be shown in the next section.

5. Results

The positioning approach presented in this work is an evolution of that in Masiero, Pirotti, et al. (2014). In this section, the functionality of the main changes with respect to (Masiero, Pirotti, et al. 2014) is tested, i.e. the altitude estimation and the detection of different movement modes (and their use during navigation). Experiments have been conducted on three floors of a university building, and in order to make the results statistically more robust, experimental data have been collected by three volunteers, two men and one woman, with heights from 1.65 to 1.85 m.

First, the (variation of) altitude estimation has been tested on 21 check points distributed on two buildings, where the range of considered altitudes ranges from the ground to 5 m, approximately. The average altitude estimation is of 0.6 m, approximately. However, this has been obtained learning the parameters of the estimation model varying the device altitude of only 1 m, approximately, whereas the computed estimation model was used mostly outside of the altitude range used in the calibration. Instead, the altitude estimation error obtained restricting the considered altitudes only to those on

the same floor of that of the calibration procedure is of 0.2 m. Hence, the simple parameter learning procedure previously proposed can be useful when the goal is that of detecting changes of floors, however when the required estimation error has to be relatively small it has to be used only in the range of altitudes considered during the learning of the model parameters (hence in order to obtain reliable estimations on a larger range of altitudes it is necessary to increase the range of altitudes also during the learning of the model parameters).

The detection of movement modes has been validated in the same university building previously considered. In order to make the system as simple as possible to use for an unqualified user, the results reported in the rest of this section have been obtained with uncalibrated sensors (however, the use of calibrated sensors allows to significantly improve the results).

Measurements involved the use of two stairs and two lifts, going in both up and down directions. Figure 3 shows the training results for the SVM classifier of the action of going down on the stairs ($\Delta t = 3$ s).

The mean altitude variation during the interval Δt long typically allows to properly discriminate each of the cases of interest. On the one hand, going up/down leads to positive/negative mean altitude variation, while, on the other hand, being in a lift or on the stairs can usually be determined by comparing the absolute value of the mean altitude variation (that is usually (approximately) fixed to a constant value for a lift). The latter case can be also distinguished by considering the standard deviation (during the Δt interval) of the absolute value of the measured acceleration vector: human (and device) movements are usually quite limited while being in a lift, leading typically to a small standard deviation. Instead,
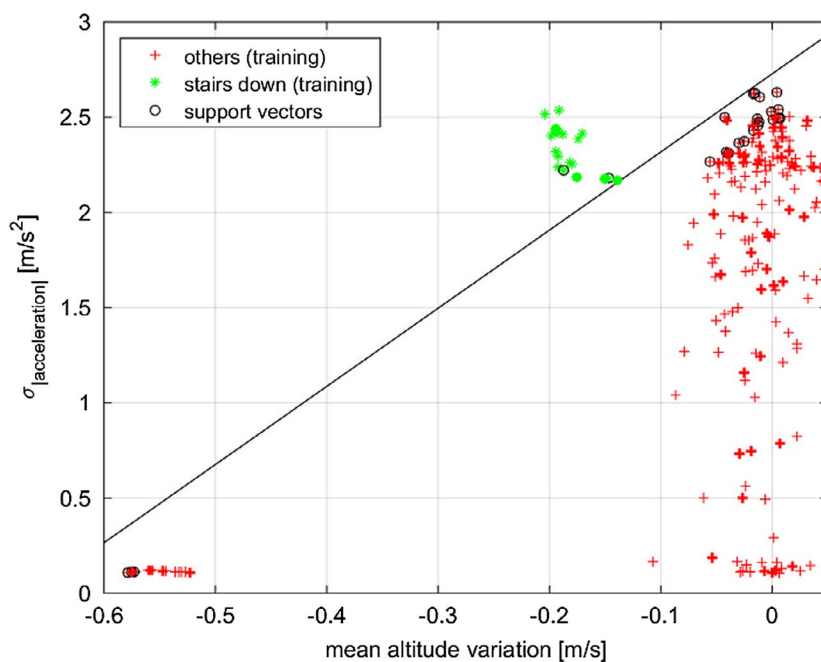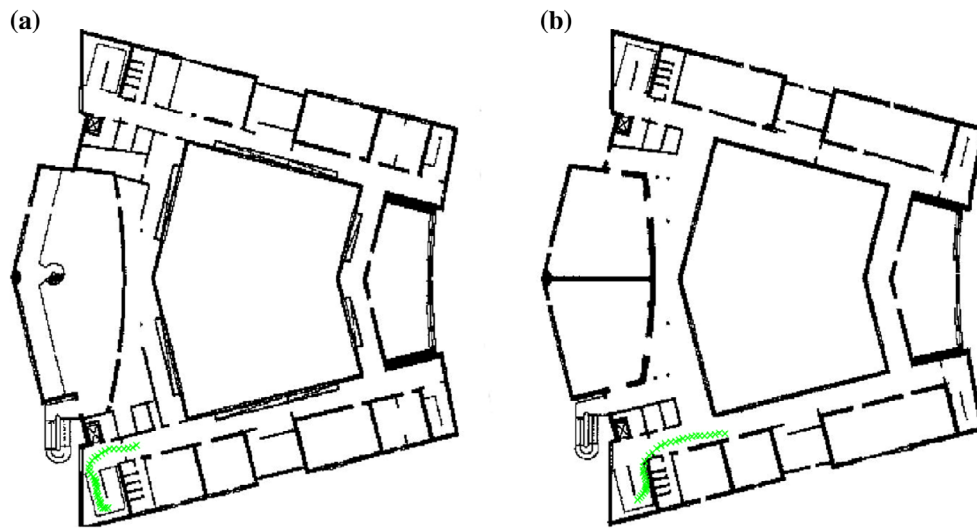

Figure 3. Training results of the SVM aiming at classifying the action of going down on the stairs. The time interval Δt has been set to 3 s.
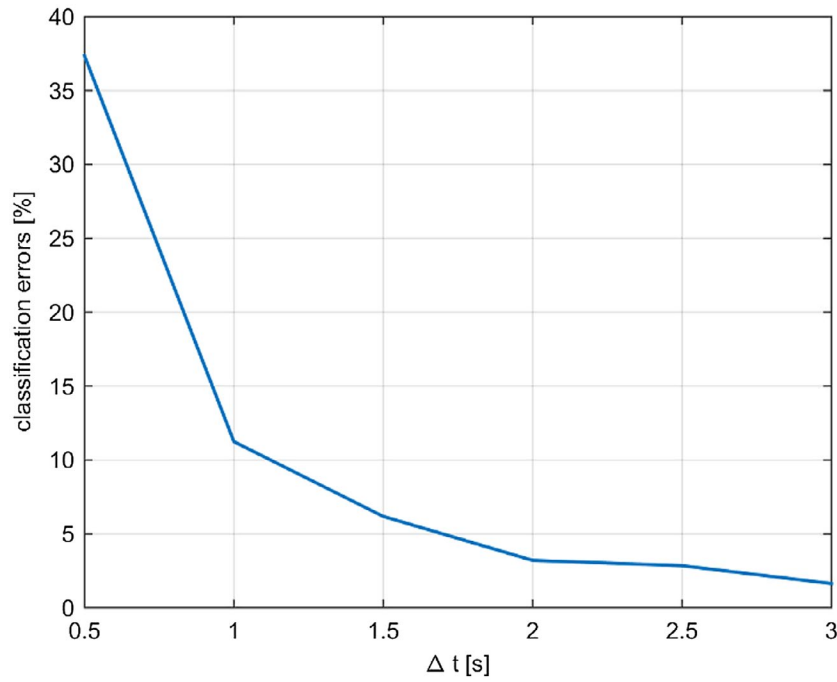
Figure 4. Comparison of the classification error obtained with the SVM classifiers by varying the value of the time interval Δ*t*. The reported results are the mean of 100 independent Monte Carlo simulations.


Figure 5. Portion of estimated trajectory distributed on two floors, (a) and (b).

during a walk the acceleration measured by the device can change because of the human steps, leading to a much larger value of the standard deviation.

The two considered variables most of the times can be successfully used to detect the four actions of interest here. However, the reliability of the computed classifiers strongly depends on the length of the time interval Δ*t*. The longer is the interval, the smaller is the influence of measurement noise (and of the human movements).

Figure 4 compares the classification error obtained with the SVM classifiers by varying the value of the time interval Δ*t* (from 0.5 to 3 s). The reported results are the mean of 100 independent Monte Carlo simulations. As
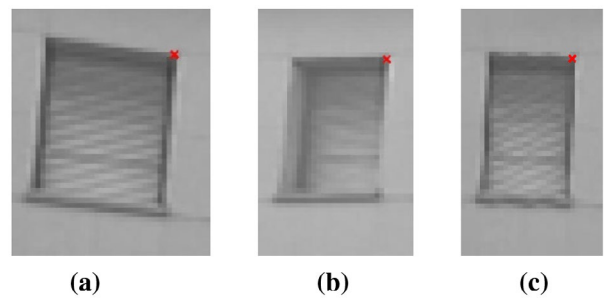
Figure 6. Details of an image region close to a feature point in the first (a) and in the second camera view (b). Image region taken from (a) remapped accordingly to the (approximate) orientation of the second camera view (c).
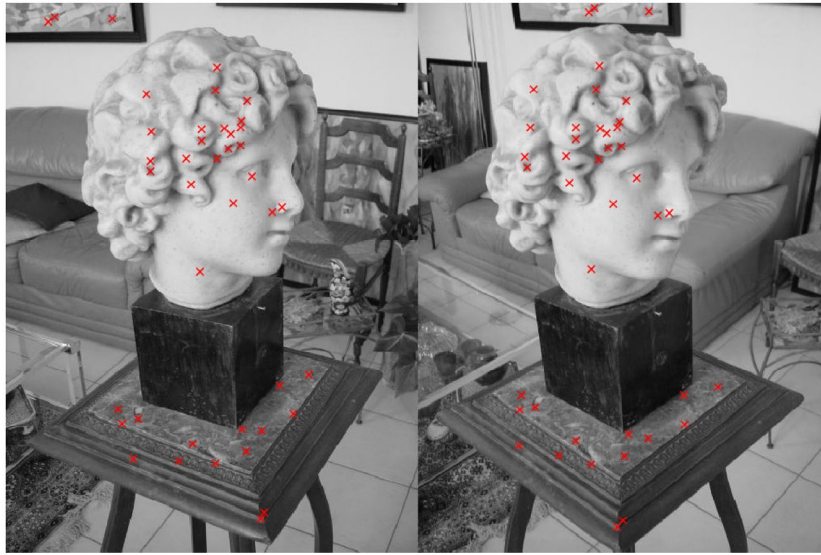
Figure 7. Example of images used for testing the feature matching procedure. Feature matchings shown in the figure have been obtained by means of the SIFT.

shown in the figure, the number of classification errors becomes less than 5% when considering $\Delta t \geq 2$ s.

Then, the use of the movement mode detector has been tested during navigation: a slightly adapted version of the Widyawan's particle filter (Widyawan et al. 2012) has been applied in order to track the device while the user is going on the stairs (Figure 5(a) and (b) show an estimated trajectory sample (distributed among two floors); 50 trajectories have been considered). WiFi has been deactivated to validate the functionality of the positioning approach based on the use of the inertial sensor measurements. The use of the movement mode detector during the navigation (i.e. by means of (7)) allowed a

20% reduction of the estimation error with respect to the standard case (3.2–4.0 m error, respectively).

For what concerns the feature matching approach presented in Section 4, first the effect of (approximate) image rectification is shown in Figure 6. Figure 6(a) and (b) show the details of the same window (with a corresponding feature) given in two different images (Figure 6(a) and (b) have been obtained by cropping two much larger images of the façade of a university building), whereas Figure 6(c) shows the approximately rectified version of Figure 6(a) to be more easily comparable to Figure 6(b). It is clear that the use of uncalibrated sensors ensures lower quality results with respect to those
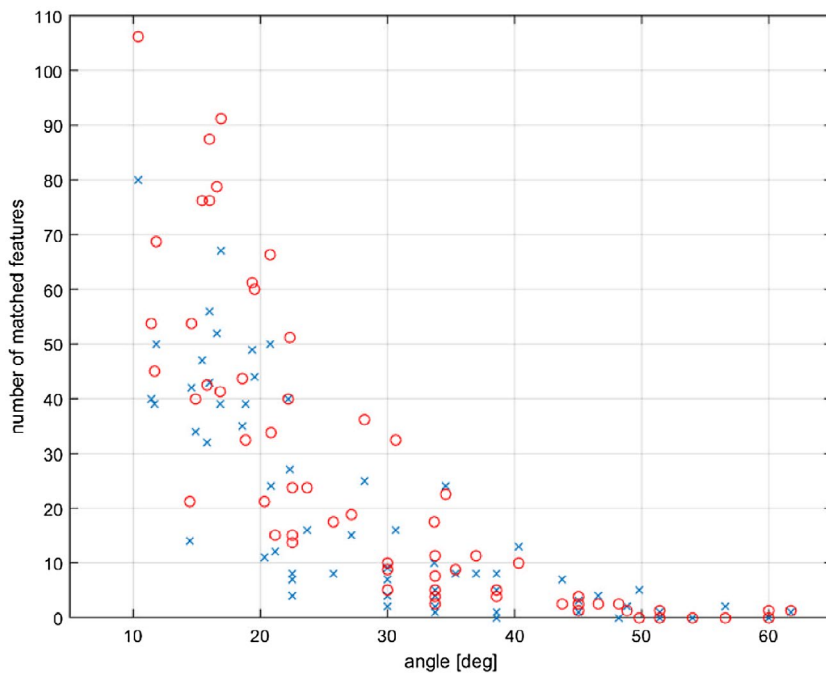
Figure 8. Number of correctly matched features varying the angle between the camera poses. Comparison of number of matched features with standard SIFT (blue *x*-marks), and with the method proposed in Section 4 (red circles).

expected in a calibrated case. Nevertheless, the rectification procedure has partially succeeded in producing an image more similar to the middle one with respect to the left one, in particular close to the feature position. Interestingly, by comparing the left border of the synthetic window (Figure 6(c)) with that in Figure 6(a) and (b), it is possible to notice that obviously the system cannot properly estimate image parts that are not visible in the original image (e.g. the internal left border of the window is not visible in the Figure 6(a), and, consequently, it cannot be presented in Figure 6(c) as well).

Then, the feature matching approach has been tested on a set of images downloadable on the Internet from the website of (Lhuillier and Quan 2005) (Figure 7 shows two of them, with the corresponding SIFT feature points; the size of all the considered images is 640×480 pixels). Since orientation information is not available for these images, approximate orientations have been computed after matching the features in the images and adding to the computed orientation angles a Gaussian random noise with standard deviation 0.15 radiant (100 independent Monte Carlo simulations have been considered in order to provide statistically reliable results). Figure 8 compares the number of correctly matched features by the proposed method (red circles) with that obtained by means of the standard SIFT (blue x-marks) varying the value of the rotation angle between the two camera poses (feature locations and SIFT descriptors have been computed with VLFeat (Vedaldi and Fulkerson 2010) for both the considered methods). The proposed method allowed to detect approximately 23% more correct feature matchings.

6. Conclusions

This paper presented recent improvements on an indoor positioning approach and a new strategy to improve feature matching results. The proposed navigation approach has been designed to work even in particularly difficult working conditions, e.g. with uncalibrated sensors, when WiFi connection is not available. The method presented here allows to estimate altitude variations and to exploit a movement mode detector in order to improve positioning estimation (20% of positioning error reduction, approximately, with respect to the version not using the movement mode detector).

Furthermore, the presented method for feature matching has reduced the computational burden required by the ASIFT, while ensuring a significant improvement in the number of correctly matched features with respect to the standard SIFT.

Notes on contributors

Andrea Masiero received the MSc in Computer Engineering and the PhD degree in Automatic Control and Operational Research from the University of Padova (Italy). He currently holds a Post-doc position at CIRGEO. His research interests range from Geomatics, to Computer Vision, Smart Camera Networks, modeling, and control of adaptive optics systems. His research mainly focused on statistical and mathematical modeling, identification and control of spatio-temporal processes, signal processing, statistical filtering, optimization, and information fusion. He is currently working on low-cost positioning and mobile MMSs.

Francesca Fissore received MSc in Environmental Physics at University of Turin. In April 2014, she received her PhD in "Monitoring of Systems and Environmental Risk Management" at the University of Genoa, during which she dealt with low-cost sensors for the detection and monitoring of environmental parameters and developed a real time web-based monitoring tool able to acquire, manage, smooth, aggregate, historicize, and display environmental parameters. Currently, she holds a Post-doc position at CIRGEO where she is working to realize a mobile MMS using low cost sensors. The project aims at developing approaches for the statistical processing of mapping data and the monitoring and collection of information from multiple sensors loaded on mobile devices.

Francesco Pirotti is an assistant professor at the Department of Land, Environment, Agriculture and Forestry of University of Padova. His research interests are in remote sensing applications for forestry and the environmental sciences, natural hazards, and risk, in particular using laser scanning (LiDAR) data. He uses collaborative web solutions for applying new algorithms to large remote sensing data-sets. He has also investigated full-waveform LiDAR data for forestry applications. He is in several national and international scientific committees and research groups related to geomatics.

Alberto Guarnieri received the degree in Electrical Engineering at the University of Padova (Italy) in 1998, then in 2004 he got the PhD in Geodetic and Topographic Sciences at the University of Bologna (Italy). Since November 2008, he is assistant professor by the Department of Land, Environment, Agriculture and Forestry (TESAF) of the University of Padova. His research activity is mainly focused on 3D modeling supported by terrestrial laser scanning and digital photogrammetry for cultural heritage and hydrogeological risk mapping. He is also involved in the development of land-based and UAV-based low-cost mobile MMSs. He currently teaches higher education courses on Global Navigation Systems (GNSS) and GIS at the School of Agricultural Sciences and Veterinary Medicine of the University of Padova. He is active in the International Society of Photogrammetry and Remote Sensing (ISPRS), serving from 2012 to 2016 as secretary of Inter-Commission Working Group ICWG I/Va – Mobile Scanning and Imaging Systems for 3D Surveying and Mapping.

Antonio Vettore is a full professor of geomatics at the University of Padova. Author of more than 200 publications in the fields of Surveying, Photogrammetry, Cartography (Geomatics). His research topics are: (1) Mobile Mapping Systems, Inertial Navigation Systems, Road Tracking, Image Processing. (2) Laser Scanner applications and GPS/INS techniques for Surveying and 3D modeling of cultural heritage sites. (3) Processing of LiDAR data from airborne or ground sensors (ALS and TLS), both discrete return and full waveform. He is the head of research units in several national (Italian Ministry of Education, Italian Civil Protection Agency) and international projects. He is actively involved in the International Society of Photogrammetry and Remote

Sensing and served in 2000–2004 as chair of the Working Group "Kinematic and Integrated Positioning System." He is the member of the Scientific Committee of MMS (Mobile Mapping Systems) at ASPRS (American Society of Photogrammetry and Remote Sensing).

ORCiD

Andrea Masiero 🆔 http://orcid.org/0000-0002-8028-2815
Francesca Fissore 🆔 http://orcid.org/0000-0003-0672-9313
Francesco Pirotti 🆔 http://orcid.org/0000-0002-4796-6406
Alberto Guarnieri 🆔 http://orcid.org/0000-0002-9483-289X
Antonio Vettore 🆔 http://orcid.org/0000-0002-5394-6700

References

Agarwal, S., N. Snavely, S. M. Seitz, and R. Szeliski. 2010. "Bundle Adjustment in the Large." In *Computer Vision – ECCV 2010*, edited by K. Daniilidis, P. Maragos, and N. Paragios, 29–42. Lecture Notes in Computer Science 6312. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-642-15552-9_3.

Bahl, P., and V. N. Padmanabhan. 2000. "RADAR: An in-Building RF-Based User Location and Tracking System." In *IEEE INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, Tel Aviv, 2, 775–784. doi:10.1109/INFCOM.2000.832252.

Bonnet, S., C. Bassompierre, C. Godin, S. Lesecq, and A. Barraud. 2009. "Calibration Methods for Inertial and Magnetic Sensors." *Sensors and Actuators A: Physical* 156 (2): 302–311. doi:10.1016/j.sna.2009.10.008.

Brand, M. 2002. "Incremental Singular Value Decomposition of Uncertain Data with Missing Values." In *Computer Vision – ECCV 2002*, edited by Anders Heyden, Gunnar Sparr, Mads Nielsen, and Peter Johansen, 707–720. Berlin: Lecture Notes in Computer Science 2350. Springer. http://link.springer.com/chapter/10.1007/3-540-47969-4_47.

Byröd, M., and K. Åström. 2010. "Conjugate Gradient Bundle Adjustment." In *Computer Vision – ECCV 2010*, edited by Kostas Daniilidis, Petros Maragos, and Nikos Paragios, 114–127. Berlin: Lecture Notes in Computer Science 6312. Springer. http://link.springer.com/chapter/10.1007/978-3-642-15552-9_9.

Chen, G., X. Meng, Y. Wang, Y. Zhang, P. Tian, and H. Yang. 2015a. "Integrated WiFi/PDR/Smartphone Using an Unscented Kalman Filter Algorithm for 3D Indoor Localization." *Sensors* 15 (9): 24595–24614. doi:10.3390/s150924595.

Chen, Z., H. Zou, H. Jiang, Q. Zhu, Y. C. Soh, and L. Xie. 2015b. "Fusion of WiFi, Smartphone Sensors and Landmarks Using the Kalman Filter for Indoor Localization." *Sensors* 15 (1): 715–732. doi:10.3390/s150100715.

Chiang, K.-W., A. Noureldin, and N. El-Sheimy. 2003. "Multisensor Integration Using Neuron Computing for Land-Vehicle navigation." *GPS Solutions* 6 (4): 209–218. doi:10.1007/s10291-002-0024-4.

Deng, Z.-A., G. Wang, Y. Hu, and D. Wu. 2015. "Heading Estimation for Indoor Pedestrian Navigation Using a Smartphone in the Pocket." *Sensors* 15 (9): 21518–21536. doi:10.3390/s150921518.

El-Sheimy, N., and K. P. Schwarz. 1998. "Navigating Urban Areas by VISAT – A Mobile Mapping System Integrating GPS/INS/Digital Cameras for GIS Applications." *Navigation* 45 (4): 275–285. doi:10.1002/j.2161-4296.1998.tb02387.x.

Facco, P., A. Masiero, and A. Beghi. 2013. "Advances on Multivariate Image Analysis for Product Quality Monitoring." *Journal of Process Control* 23 (1): 89–98. doi:10.1016/j.jprocont.2012.08.017.

Facco, P., A. Masiero, F. Bezzo, M. Barolo, and A. Beghi. 2011. "Improved Multivariate Image Analysis for Product Quality Monitoring." *Chemometrics and Intelligent Laboratory Systems* 109 (1): 42–50. doi:10.1016/j.chemolab.2011.07.006.

Foxlin, E. 2005. "Pedestrian Tracking with Shoe-Mounted Inertial Sensors." *IEEE Computer Graphics and Applications* 25 (6): 38–46. doi:10.1109/MCG.2005.140.

Furukawa, Y., and J. Ponce. 2010. "Accurate, Dense, and Robust Multiview Stereopsis." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (8): 1362–1376. doi:10.1109/TPAMI.2009.161.

Fusiello, A., and L. Irsara. 2010. "Quasi-Euclidean Epipolar Rectification of Uncalibrated Images." *Machine Vision and Applications* 22 (4): 663–670. doi:10.1007/s00138-010-0270-3.

Guarnieri, A., A. Masiero, A. Vettore, and F. Pirotti. 2015. "Evaluation of the Dynamic Processes of a Landslide With Laser Scanners and Bayesian Methods." *Geomatics, Natural Hazards and Risk* 6 (5–7): 614–634. doi:10.1080/19475705.2014.983553.

Habib, A., M. Ghanma, M. Morgan, and R. Al-Ruzouq. 2005. "Photogrammetric and Lidar Data Registration Using Linear Features." *Photogrammetric Engineering & Remote Sensing* 71 (6): 699–707. doi:10.14358/PERS.71.6.699.

Hadeel, A., M. Jabbar, and X. Chen. 2011. "Remote Sensing and GIS Application in the Detection of Environmental Degradation Indicators." *Geo-spatial Information Science* 14 (1): 39–47. doi:10.1007/s11806-011-0441-z.

Hartley, R., and A. Zisserman. 2003. *Multiple View Geometry in Computer Vision*. Cambridge: Cambridge University Press.

Heyden, A., and M. Pollefeys. 2005. "Multiple View Geometry." In *Emerging Topics in Computer Vision*, edited by Gerard Medioni and Sing Bing Kang, 45–107. Upper Saddle River, NJ: Prentice Hall.

Huang, B., and Y. Gao. 2013. "Ubiquitous Indoor Vision Navigation Using a Smart Device." *Geo-spatial Information Science* 16 (3): 177–185. doi:10.1080/10095020.2013.817110.

Jaakkola, A., J. Hyyppä, A. Kukko, X. Yu, H. Kaartinen, M. Lehtomäki, and Y. Lin. 2010. "A Low-cost Multi-sensoral Mobile Mapping System and Its Feasibility for Tree Measurements." *ISPRS Journal of Photogrammetry and Remote Sensing* ISPRS Centenary Celebration Issue 65 (6): 514–522. doi:10.1016/j.isprsjprs.2010.08.002.

Jahn, J., U. Batzer, J. Seitz, L. Patino-Studencka, and J. Gutierrez Boronat. 2010. "Comparison and Evaluation of Acceleration Based Step Length Estimators for Handheld Devices." In *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 1–6. Zurich: IEEE. doi:10.1109/IPIN.2010.5646888.

Karel, W., and N. Pfeifer. 2009. "Range Camera Calibration Based on Image Sequences and Dense Comprehensive Error Statistics." In *IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics*, 7239:72390D–72390D–12. doi:10.1117/12.807785.

Kraus, K., and N. Pfeifer. 1998. "Determination of Terrain Models in Wooded Areas with Airborne Laser Scanner Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 53 (4): 193–203. doi:10.1016/S0924-2716(98)00009-4.

Lhuillier, M., and L. Quan. 2005. "A Quasi-dense Approach to Surface Reconstruction From Uncalibrated Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (3): 418–433. doi:10.1109/TPAMI.2005.44.

Lowe, D.G. 1999. "Object Recognition from Local Scale-Invariant Features." In *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1150–1157, Vol. 2. New York: IEEE. doi:10.1109/ICCV.1999.790410.

Lukianto, C., and H. Sternberg. 2011. "STEPPING – Smartphone-Based Portable Pedestrian Indoor Navigation." *Archiwum Fotogrametrii, Kartografii I Teledetekcji* 22: 311–323.

Ma, Y., S. Soatto, J. Kosecka, and S. S. Sastry. 2004. *An Invitation to 3-D Vision: From Images to Geometric Models*. New York: Springer Science & Business Media.

Masiero, A., and A. Cenedese. 2012. "On Triangulation Algorithms in Large Scale Camera Network Systems." In *2012 American Control Conference (ACC)*, 4096–4101. Montréal: IEEE. doi:10.1109/ACC.2012.6315278.

Masiero, A., A. Guarnieri, F. Pirotti, and A. Vettore. 2014. "A Particle Filter for Smartphone-Based Indoor Pedestrian Navigation." *Micromachines* 5 (4): 1012–1033. doi:10.3390/mi5041012.

Masiero, A., A. Guarnieri, A. Vettore, and F. Pirotti. 2014. "An ISVD-based Euclidian Structure from Motion for Smartphones." *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-5: 401–406.

Morel, J., and G. Yu. 2009. "ASIFT: A New Framework for Fully Affine Invariant Image Comparison." *SIAM Journal on Imaging Sciences* 2 (2): 438–469. doi:10.1137/080732730.

Morel, J., and G. Yu. 2011. "Is SIFT Scale Invariant?" *Inverse Problems and Imaging* 5 (1): 115–136. doi:10.3934/ipi.2011.5.115.

Pfeifer, N., B. Gorte, and D. Winterhalder. 2004. "Automatic Reconstruction of Single Trees from Terrestrial Laser Scanner Data." *Proceedings of 20th ISPRS Congress*, 114–119. Istanbul: ISPRS.

Piragnolo, M., A. Masiero, F. Fissore, and F. Pirotti. 2015. "Solar Irradiance Modelling with NASA WW GIS Environment." *ISPRS International Journal of Geo-Information* 4 (2): 711–724. doi:10.3390/ijgi4020711.

Piras, M., G. Marucco, and K. Charqane. 2010. "Statistical Analysis of Different Low Cost GPS Receivers for Indoor and Outdoor Positioning." In *Position Location and Navigation Symposium (PLANS), 2010*. Indian Wells USA: IEEE/ION. 838–849. doi:10.1109/PLANS.2010.5507325.

Pirotti, F., A. Guarnieri, A. Masiero, and A. Vettore. 2015. "Preface to the Special Issue: The Role of Geomatics in Hydrogeological Risk." *Geomatics, Natural Hazards and Risk* 6 (5–7): 357–361. doi:10.1080/19475705.2014.984248.

Pirotti, F., G. V. Laurin, A. Vettore, A. Masiero, and R. Valentini. 2014. "Small Footprint Full-Waveform Metrics Contribution to the Prediction of Biomass in Tropical Forests." *Remote Sensing* 6 (10): 9576–9599. doi:10.3390/rs6109576.

Remondino, F., L. Barazzetti, F. Nex, M. Scaioni, and D. Sarazzi. 2011. "UAV Photogrammetry for Mapping and 3D Modeling-current Status and Future Perspectives." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (1): C22.

Remondino, F., and C. Fraser. 2006. "Digital Camera Calibration Methods: Considerations and Comparisons." *International Archives of Photogrammetry and Remote Sensing*. http://masters.dgtu.donetsk.ua/2007/ggeo/sagaidak/library/book2.htm.

Remondino, F., A. Guarnieri, and A. Vettore. 2005. "3D Modeling of Close-Range Objects: Photogrammetry or Laser Scanning?" In *Proceedings SPIE 5665, Videometrics VIII* 5665: 216–225. doi:10.1117/12.586294.

Ruiz, A. R. J., F. S. Granja, J. C. P. Honorato, and J. I. G. Rosas. 2012. "Accurate Pedestrian Indoor Navigation by Tightly Coupling Foot-Mounted IMU and RFID Measurements." *IEEE Transactions on Instrumentation and Measurement* 61 (1): 178–189. doi:10.1109/TIM.2011.2159317.

Saeedi, S., A. Moussa, and N. El-Sheimy. 2014. "Context-aware Personal Navigation Using Embedded Sensor Fusion in Smartphones." *Sensors* 14 (4): 5742–5767. doi:10.3390/s140405742.

Schiavone, G., M. P. Y. Desmulliez, and A. J. Walton. 2014. "Integrated Magnetic MEMS Relays: Status of the Technology." *Micromachines* 5 (3): 622–653. doi:10.3390/mi5030622.

Tao, W. 2013. "Interdisciplinary Urban GIS for Smart Cities: Advancements and Opportunities." *Geo-spatial Information Science* 16 (1): 25–34. doi:10.1080/10095020.2013.774108.

Toth, C. K. 2001. "Sensor Integration in Airborne Mapping." In *Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference, 2001. IMTC 2001*, vol. 3: 2000–2005. Budapest, Hungary: IEEE. doi:10.1109/IMTC.2001.929550.

Toth, C. K., and D. A. Grejner-Brzezinska. 1997. "Performance Analysis of the Airborne Integrated Mapping System (AIMS^TM)." *International Archives of Photogrammetry and Remote Sensing* 32: 320–326.

Vedaldi, A., and B. Fulkerson. 2010. "Vlfeat: An Open and Portable Library of Computer Vision Algorithms." In *Proceedings of the 18th ACM International Conference on Multimedia*, 1469–1472. MM '10. New York: ACM. doi:10.1145/1873951.1874249.

Widyawan, G. Pirkl, D. Munaretto, C. Fischer, C. An, P. Lukowicz, M. Klepal, et al. 2012. "Virtual Lifeline: Multimodal Sensor Data Fusion for Robust Navigation in Unknown Environments." *Pervasive and Mobile Computing* 8 (3): 388–401. doi:10.1016/j.pmcj.2011.04.005.