

VISION-AIDED CONTEXT-AWARE FRAMEWORK FOR PERSONAL NAVIGATION SERVICES

S. Saeedi*, A. Moussa, Dr. N. El-Sheimy

Dept. of Geomatic Engineering, The University of Calgary, 2500, University Dr, NW, Calgary, AB, T2N 1N4 Canada
– (ssaeedi, amelsaye, elsheimy) @ucalgary.ca

Commission IV, WG IV/5

KEY WORDS: Navigation, Vision, Data mining, Recognition, Fusion, Video, IMU

ABSTRACT:

The ubiquity of mobile devices (such as smartphones and tablet-PCs) has encouraged the use of location-based services (LBS) that are relevant to the current location and context of a mobile user. The main challenge of LBS is to find a pervasive and accurate personal navigation system (PNS) in different situations of a mobile user. In this paper, we propose a method of personal navigation for pedestrians that allows a user to freely move in outdoor environments. This system aims at detection of the context information which is useful for improving personal navigation. The context information for a PNS consists of user activity modes (e.g. walking, stationary, driving, and etc.) and the mobile device orientation and placement with respect to the user. After detecting the context information, a low-cost integrated positioning algorithm has been employed to estimate pedestrian navigation parameters. The method is based on the integration of the relative user's motion (changes of velocity and heading angle) estimation based on the video image matching and absolute position information provided by GPS. A Kalman filter (KF) has been used to improve the navigation solution when the user is walking and the phone is in his/her hand. The Experimental results demonstrate the capabilities of this method for outdoor personal navigation systems.

1. INTRODUCTION

Due to the rapid developments in mobile computing, wireless communications and positioning technologies, using smartphones as a PNS is getting popular. This evolution has facilitated the development of applications that use the position of the user, often known as LBS. Using various sensors on smartphones provides a vast amount of information; however, finding a ubiquitous and accurate pedestrian navigation solution is a very challenging topic in ubiquitous positioning (Lee & Gerla, 2010; Mokbel & Levandoski, 2009). Position estimation in outdoor environments is mainly based on the global positioning systems (GPS) or assisted GPS (AGPS); however, it is a challenging task in indoor or urban canyon, especially when GPS signals are unavailable or degraded due to the multipath effect. In such cases, usually other navigation sensors and solutions are applied for pedestrians. The first alternative is wireless radio sensors, such as Bluetooth, RFID (Radio Frequency IDentification) or WLAN (Wireless Local Area Network). These systems have limited availability and need a pre-installed infrastructure that restricts their applicability. The second navigation system is the IMU (Inertial Measurement Unit) sensors that provide a relative position based on the distance travelled and device's orientation. The distance and orientation information can be measured with a gyroscope and an accelerometer sensor. The main drawback of the IMU is that they are based on the relative position estimation techniques and use the previous states of the system; therefore, after a short period of time low cost MEMS (Micro Electro-Mechanical Systems) sensors measurements typically result in large cumulative drift errors unless the error are bounded by measurements from other systems (Aggarwal et al., 2010). Another solution is the vision-based navigation using video camera sensors. These systems are based on two main strategies: estimation of absolute position information using a priori formed databases which highly depends on the

availability of image database for that area (Zhang and Kosecka, 2006) and estimating relative position information using the motion of the camera calculated from consecutive images which suffers from cumulative drift errors (Ruotsalainen et al., 2011; Hide et al., 2011). Since there is not a single comprehensive sensor for indoor navigation, it is necessary to integrate the measurements from different sensors to improve the position information.

Modern smartphones contain a number of Low cost MEMS sensors (e.g. magnetometer, accelerometer, and gyroscope) that can be used for integrated ubiquitous navigation even if GPS signals are unavailable. Vision sensors are ideal for PNS since they are available in good resolution on almost all smartphones. Therefore, in this research a vision sensor is used to capture the user's motion parameters using consecutive image frames and to provide navigation aid when measurements from other systems such as GPS are not available. This system doesn't need any special infrastructure and makes use of camera as an ideal aiding system. Since mobile users carry the device with different orientation and placement, in almost everywhere (indoor and outdoor environments) while doing various activities (such as walking, running and driving), using specific customized and context-aware algorithms are necessary for different users' modes. Therefore, a mobile navigation application must be aware of user and device context to use appropriate algorithm for each case. For example, when the context information shows that device is in "texting" or "talking" mode, the observation from camera can be integrated with GPS sensor to improve and validate the pedestrian dead-reckoning algorithm. The main issue in context-aware PNSs is detecting relevant context information using embedded mobile sensors in an implicit way. The contribution of this paper is to develop a visually-aided personal navigation solution using the smartphone embedded sensors which takes into account various user context.

2. VISION-AIDED PEDESTRIAN NAVIGATION

Recently, by the increase in the resolution of digital cameras and computing power of mobile devices, visual sensors have gained a great attention in the positioning research community. Therefore, they have been used for motion detection, obstacle avoidance, and relative and absolute localization. Vision-based navigation has been used for decades in navigation of robots (Corke et al. 2007); however, using it in pedestrian navigation has become a research topic only in the last few years (Ruotsalainen et al., 2011; Hide et al., 2011; Steinhoff et al., 2007). The focus of the vision-aided navigation research has been mainly in systems using a priori formed databases. When a match between images in the database and the ones taken by a pedestrian is found, the absolute position can be obtained. This procedure needs a priori preparations and highly depends on the availability of image database for that area. On the other hand, another group of algorithms with a wide range of applications deploy real-time motion estimation of a single camera moving freely through an environment. This estimation can be helpful in detecting displacement and orientation of the device and estimating the user's turns (Hide et al., 2011). This information can be incorporated in the position and heading estimation of pedestrian navigation. However, there are various problems when processing the video frames from a hand-held device's camera. First of all, the measurements are relative, therefore to estimate the absolute quantities, initialization of the parameters are required. Moreover, the scale of the observation cannot be obtained using only vision, and another sensor or a known dimension reference has to be used in order to retrieve the scale of the observation. Also the orientation of the mobile device affects the heading and velocity information. In this paper, we describe a low-cost context-aware personal navigation system that is capable of localizing a pedestrian using fusion of GPS and camera to robustly estimate frame-to-frame motion in real time (also known as visual odometry).

2.1 Computer Vision Algorithm

Motion estimation from video is a well-studied problem in computer vision. Approaches for motion estimation are based on either dense optical flow or sparse feature tracks (Steinhoff et al., 2007). In this paper a computer vision algorithm is developed to find the motion vector using the matched features between successive frames. The detected motion vectors are employed to estimate the forward motion velocity and the azimuth rotation angle between the two frames. To detect the motion vectors, interest points are detected from the frames using Speeded Up Robust Features (SURF) algorithm (Bay et al., 2008). The detected interest points of two successive frames are matched based on the Euclidean distance between the descriptors of these points. The vectors starting from an interest point in frame and ending at the corresponding matched point in the next frames are considered as candidate motion vectors.

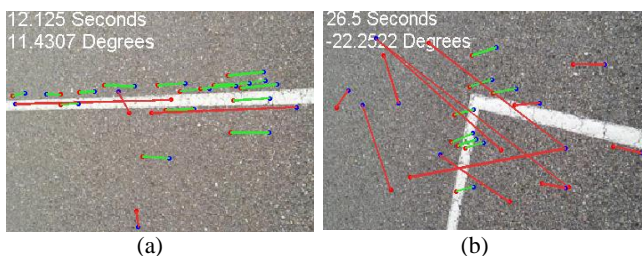


Figure 1. The matched features, candidate motion vectors (red), and acceptable motion vectors using RANSAC in two different cases: a) forward motion and b) change of the the heading.

As shown in figure 1, some matches could be incorrect due to the existence of repeated similar points in the frames. Therefore, the candidate motion vectors should be filtered out to remove the inconsistent vectors based on discrepancy in length or orientation of the vector (figure 1). The RANdom SAMpling Consensus algorithm (Fischler et al., 1981) is used to find the vector angle and vector length with the maximum number of compatible vectors. The accepted motion vectors are then averaged to get the average motion vector. The accuracy of their average motion vector is highly dependent on the number of the compatible vectors and variance of the angles and lengths of these vectors. Figure 2 shows the number of acceptable motion vectors from the first 20 motion vectors detected as the best matches in the successive frames. Under the assumption of having context information of the hand-held device alignment (texting mode and landscape/portrait forward alignment), the vertical component of the average motion vector is a measure of the forward motion speed between the two frames. The horizontal component of the average motion vector is a measure of the azimuth change between the two frames. To calibrate the scale approximation between the motion vector and both the forward velocity and the azimuth change, a reference track is navigated using the motion vector only. The transformation parameters between the motion vector and forward speed and azimuth change are computed so that the navigation solution matches the reference solution. Using the computed transformation parameters, the forward motion velocity and the azimuth change can be approximated between any two successive frames with the help of the average motion vector. The estimation of the velocity from camera can also be improved by the user mode such as walking, stairs, running context information. However, the relative measurements from the computer vision algorithm tend to accumulate error over time, resulting in long-term drifts. To limit this drift, it is necessary to augment such local pose systems with global estimations such as GPS.

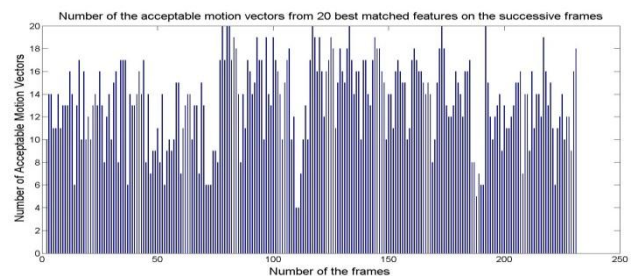


Figure 2. The numbe of the acceptable motion vectors from 20 best matched features on consecutive frames.

3. CONTEXT INFORMATION IN PNS

In order to achieve a context-aware "vision-aided pedestrian navigation" system, two important questions must be answered: what type of context is important for such a system and how can we extract it using the sensors on a hand-held device? The following section discusses these issues and investigates different methods for context extracting from a mobile device's sensors.

Context may refer to any piece of information that can be used to characterize the situation of an entity (person, place, or object) that is relevant to the interaction between a user and an application (Dey, 2001). While location information is by far the most frequently used attribute of context, attempts to use

other context information such as “user activity” have increased over the last few years (Baldauf et al., 2007). The primary contexts relevant to the navigation services in a mobile device can be divided into three categories: Environment, user, and device. As listed in table 1, *Environment* contexts include time and location of the user which are two fundamental dimensions of ubiquitous computation and have been discussed in various studies (Choujaa & Dulay, 2008). In contrast, detecting user’s activity is still an open topic in context-aware systems. *User activity* context refers to a sequence of motion patterns usually executed by a single person and at least lasting for a short duration of time, on the order of tens of seconds. In navigation services, another important issue is that “where the device is located with respect to the user”. In PNS, usually a mobile device can be carried out by the user in an arbitrary placement and orientation (e.g. in the pocket, in hand, on belt, in backpack, on vehicle’s seat, etc.).

Table 1. Contextual information coping with the proposed navigation systems

Context type		Context Values/samples
Environment	Time	Time of the day/night, Weekend/weekday, ...
	Location	Outdoor, Indoor, close by point of interests, ...
User	Activity	Static, Walking, Running, Stair up-down, Elevator, Driving, Cycling, ...
Device	Orientation	Horizontal/Vertical, Face-up/down, Landscape/Portrait
	placement	Dangling with hand swing, Texting (with one/two hand), In a pocket (pants, jacket), In hand bag, In backpack, On belt, Talking (close to ear/speaker)
	Sensor	Availability and Accuracy
	Network	Availability
	Battery	Power level

The contexts that are useful for vision-aided system include: device orientation (e.g. face-up/down, vertical or portrait modes), device location (Texting with 1 hand or 2 hand mode) and activity of the user (e.g. walking mode). By texting we refer to the position of the user while texting and therefore it includes all similar positions such as surfing, playing, reading and etc. Texting mode requires the user to hold the device in front of himself using one or both hands. Since information gathered by a single sensor can be very limited and may not be fully reliable and accurate, in this research a new approach has been proposed based on the multi-sensor fusion to improve the accuracy and robustness of context aware system (Saeedi et al., 2011).

3.1 Context Recognition Module

Most of the current approaches for context recognition are data-driven (Yang, et al., 2010; Pei, et al., 2010). In this research we aim at integrating the data-driven paradigm with the knowledge-oriented paradigm to solve context detection problems considering expert’s rules and other information sources. Activity recognition module follows a hierarchical approach (Avci, 2010) for fusing accelerometer and gyroscope sensors in feature level. As it is shown in figure 3, the raw data captured by sensors is pre-processed for calibration and noise reduction. Then, signal processing and statistical algorithm are used to derive an appropriate set of features from the measurements. The potential number of features that can be used is numerous;

however, the used features need to be carefully selected to perform real-time and robust context recognition. After feature extraction, pattern recognition techniques can be used to classify the feature space. There is a wide variety of classification techniques and often selecting the best one depends on the application (Saeedi et al., 2011).

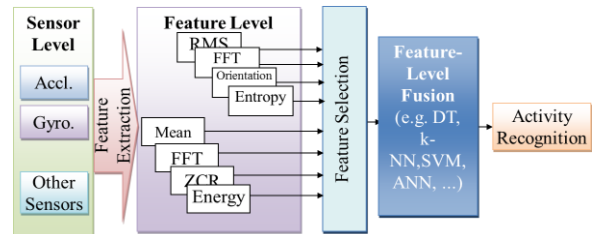


Figure 3. Feature recognition module (Saeedi et al., 2011)

Figure 4 presents an example of accelerometer sensors’ output in different placement scenarios after sensor calibration and low-pass filtering. Some modes are easy to identify, such as the dangling mode in which accelerometer has significantly large magnitude due to the arm swing. However, other modes are quite similar to each other and require pattern recognition algorithms for classification.

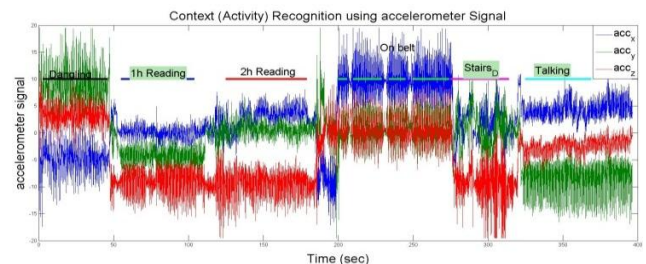


Figure 4. Tri-axial accelerometers output in different placement mode

In this research the following features (table 2) has been used in time and frequency domains for context detection based on inertial data.

Table 2. The useful time and frequency domain features for context detection

Feature Space		Description
Time-Domain Features	Mean	$\bar{y} = \frac{\sum_{j=1}^N (y_j)}{N-1}$
	Variance	$\text{Var} = \left(\frac{1}{N-1} \sum_{j=1}^N (y_j - \bar{y})^2 \right)^{1/2}$
	Inter-axis Correlation	it has good results in detecting device orientation and placement
	Zero or Mean Crossing Ratio	useful for detecting device placement
Frequency-Domain Features	Frequency Range Power	computes the power of the discrete Fourier Transform for a given frequency band
	Spectral Energy	distinguish inactive activities from dynamic activities
	Spectral Entropy	a measure of the distribution of the frequency components in the frequency band

In order to increase robustness of activity recognition and reduce computations, a k-NN based feature selection method is applied and a set of twelve features has been selected with the

same accuracy for classification approach. These selected features have been used as inputs for the classification and recognition methods. We evaluated and compared several classifiers provided by WEKA, namely Bayesian Network, the Support Vector Machine (LibSVM), k-Nearest Neighbor (kNN), and Artificial Neural Network (ANN) (Saeedi et al., 2011). The SVM has the best performance in this case and have been shown in figure 6.

Uncertainty is an integral part of the extracted context information and it is mostly caused by the imperfectness and incompleteness of sensed data and classification models. For example, it is difficult to detect if the phone is in the pocket or on the belt based on low-level sensing information such as accelerometer signals. Therefore, in our work we used Fuzzy Inference Engine (FIS) to transform the data into higher-level descriptions of context information. The hybrid method is capable of handling the uncertainty of the activities recognized using signal processing, removing the conflicts and preserving consistency of detected contexts, and filling the gaps. The list of linguistic variables and their corresponding membership functions is mentioned in table 3).

Table 3. Definition of fuzzy input variables

Walking pattern correlation	<i>Proper</i>	$>.7$
	<i>Medium</i>	$>.4 \text{ \& } <.7$
	<i>Poor</i>	$<.4$
Connectivity between activities	<i>high</i>	$<.7$
	<i>Medium</i>	$>.4 \text{ \& } <.7$
	<i>low</i>	$>.4$
GPS DOP	<i>Good</i>	$1-4$
	<i>Moderate</i>	$5-10$
	<i>fair</i>	$10-20$
	<i>Poor</i>	>20
GPS velocity	<i>Driving</i>	$>10 \text{ (m)}$
	<i>Pedestrian</i>	$<10 \text{ (m)}$

In the next step fuzzy rules between the input and the output membership functions has been defined. These rules are determined using an experienced human. Based on the defined membership functions and the rules, fuzzy reasoning for the conjugate point determination is carried out in a Mamdani type (Zadeh, 1965) fuzzy reasoning structure. In the following four sample rules for detecting context information are presented:

If walking correlation of dangling is proper and connectivity of dangling is high then context is dangling
If GPS velocity is driving and GPS-DOP is good or moderate then environment is outdoor

In designing rule repository, the designer can define specific constraints to incorporate common-sense knowledge. This will reduce the amount of required training data and makes the rule mining computationally efficient. An example of such a constraint is that a person cannot drive while in an indoor environment. Therefore our rule repository is composed of a number of predicates generated by the user and designer along with the mined association rules. These rules are stored in a knowledge-base (KB) that facilitates the modification, updating or removing the rules. In the rule based engine, different types of rules have different levels of confidence and reliability.

4. NAVIGATION SENSOR INTEGRATION

The core of the vision-aided pedestrian navigation system consists of GPS location and velocity information for retrieving

absolute positioning while the position aid (velocity and heading change rate) information is provided from frame to frame camera images. These measurements are integrated using a KF filter (Aggarwal et al., 2010) that is presented briefly in the following section. The design of the integrated pedestrian navigation algorithm is shown in figure 5. The contexts that are useful for vision-aided system include: device orientation (e.g. face-up/down, vertical or portrait modes), device location (texting mode) and activity of the user (e.g. walking mode). Also, the context information about sensor's availability and accuracy can be used to select the device dynamic and observation model in the KF.

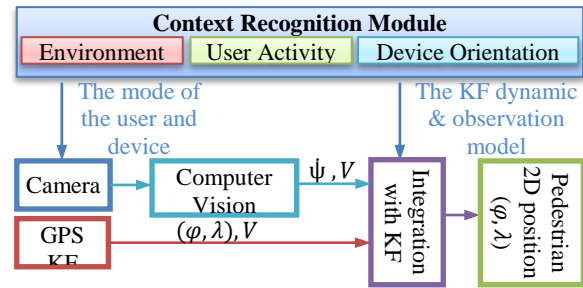


Figure 5. The multi-sensor pedestrian navigation diagram using context-aware vision-aided observation

In this paper the dynamic system is based on whether the user is in *texting* mode while walking in an outdoor environment. In order to model the characteristics of the two-dimensional motion of a walking user we have used Dead Reckoning (DR) algorithms. DR is the determination of a new position from the knowledge of a previously known position, using the current distance and heading information. In a 2D-navigation, the current coordinates (E_t, N_t) with respect to a previously known position (E_{t-1}, N_{t-1}) can be computed as follows:

$$E_t = E_{t-1} + S_{[t-1,t]} \sin \psi_{t-1} \quad (1)$$

$$N_t = N_{t-1} + S_{[t-1,t]} \cos \psi_{t-1} \quad (2)$$

where $S_{[t-1,t]}$ denote the distance travelled by the user since time $t - 1$ and ψ_{t-1} is the user's heading since time $t - 1$.

4.1 Kalman Filter

The absolute position observations from GPS and heading measurements obtained from camera have been integrated using a KF. This filter uses the dynamic model to make a prediction of the state in the next time step. Then, it uses an observation model to compare the predicted and observed states. The dynamic equation of a KF is (Aggarwal et al., 2010):

$$x_k = \Phi_{k-1} x_{k-1} + w_{k-1} \quad (3)$$

where, x_k is the state vector, Φ_{k-1} represents the transition matrix that relates the state of a previous time to the current time, and w_k is the process noise which is assumed to be drawn from a zero mean multivariate normal distribution with covariance Q_k ($w_k \sim N(0, Q_k)$). In this case, the dynamic equations for vision aided GPS is:

$$E_{k+1} = E_k + V_k \cdot \sin \psi_k \cdot \Delta t_k + w_1 \quad (4)$$

$$N_{k+1} = N_k + V_k \cdot \cos \psi_k \cdot \Delta t_k + w_2 \quad (5)$$

$$\psi_{k+1} = \psi_k + w_3 \quad (6)$$

$$\psi_{k+1} = \psi_k \cdot \Delta t_k + \psi_k + w_4 \quad (7)$$

$$V_{k+1} = V_k + w_5 \quad (8)$$

where E, N represent the absolute position in the East and North coordinate, both in meters, V (m/s) is the speed, ψ (radian) the heading defined with the origin North and clockwise positive, and $\dot{\psi}$ (radian/s) the heading change rate. The variable Δt presents the time between two epochs. The state vector in our system is:

$$\mathbf{x}_k = [E_k \quad N_k \quad \dot{\psi}_k \quad \psi_k \quad V_k]^T \quad (9)$$

To avoid linearization, the state transition matrix is defined here simplified as:

$$\Phi_k = \begin{bmatrix} 1 & 0 & 0 & 0 & \sin \psi_{k-1} \Delta t_k \\ 0 & 1 & 0 & 0 & \cos \psi_{k-1} \Delta t_k \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \Delta t_k & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (10)$$

Φ_k is approximated as a constant matrix at every time epoch k . Observation Model general form is presented in equation (11) and is defined according to the information provided by the GPS and visual sensor.

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \quad (11)$$

where \mathbf{z}_k , is the observation vector, \mathbf{H}_k is the observation model which relates the state space into the observed space and \mathbf{v}_k is the observation noise which is assumed to be zero mean Gaussian white noise with covariance \mathbf{R}_k ($\mathbf{v}_k \sim N(0, \mathbf{R}_k)$). The number of measurements fed to the filter is varied on an epoch-to-epoch basis based on the availability of the sensors and its data rate. The non-availability situation of the visual aiding is based on the matching accuracy and was discussed in the computer vision section of this paper. The accuracy of the GPS sensor is also available on the android smartphones. The full-scale measurement vector (\mathbf{z}_k) is as follows:

$$\mathbf{z}_k = [E_{GPS} \quad N_{GPS} \quad V_{GPS} \quad \dot{\psi}_{cam} \quad V_{cam}] \quad (12)$$

The KF works in two phases: the prediction and the update. In the first phase, the filter propagates the states and state's accuracies using the dynamic matrix Φ_{k-1} and $\hat{\mathbf{x}}_{k-1}^+$ (estimated in the previous epoch), based on this equation: $\hat{\mathbf{x}}_k^- = \Phi_k \hat{\mathbf{x}}_{k-1}^+$. Then the covariance matrix \mathbf{P}_k can be estimated using \mathbf{P}_{k-1}^+ . The usual equation to calculate \mathbf{P}_k^- is $\mathbf{P}_k^- = \Phi_k \mathbf{P}_{k-1}^+ \Phi_k^T + \mathbf{Q}_{k-1}$. In the update phase the state is corrected by a robust blending of prediction solution with the update measurements based on the following equation:

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \bar{\mathbf{K}}_k (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^-) \quad (13)$$

where $\bar{\mathbf{K}}$ is the Kalman gain obtained by: $\bar{\mathbf{K}} = \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k)^{-1}$. The update of the covariance takes place with the equation: $\mathbf{P}_k^+ = \mathbf{P}_k^- - \bar{\mathbf{K}}_k \mathbf{H}_k \mathbf{P}_k^-$.

5. EXPERIMENTS AND RESULTS

The potential of the proposed method are evaluated through comprehensive experimental tests conducted on a wide variety of datasets using a Samsung Galaxy Note smartphone. Multiple sensors are integrated on the circuit board including MEMS tri-axial accelerometers (STMicroelectronics k3dh), three orthogonal gyros (K3G), a back camera (Samsung S5K5BAF-2MP that can record video frames in HD format, and a GPS receiver module. To gather data from the phone, an application called TPI android logger (developed by MMSS research group at the University of Calgary) is used. These applications can be used in real time and collect data with a timestamp.

For the context recognition, extensive pedestrian field tests have been performed. First, training datasets for accelerometer and gyro signals were collected for 10 minutes: three users were asked to perform walking around a tennis court repeatedly with different activities and device orientations such as on belt, in pocket, carting in the backpack, in-hand dangling, texting and talking modes. After the activity recognition step, the classified results were compared with the known placement configurations as shown in figure 6 to evaluate the accuracy of the context recognition.

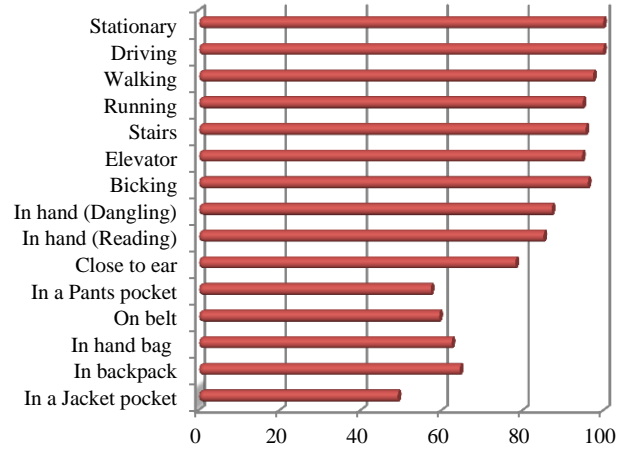


Figure 6: Recognition rates for different activities using Feature-level fusion algorithm (SVM)

Figure 6 shows the recognition rate for each activity using SVM. By investigating each activity's recognition rate, it can be inferred that the user activities such as: texting, driving, walking, running, taking stairs and elevator modes have an accuracy of 95%. In contrast, the classification models cannot distinguish between the device placements such as in pocket and on belt. This is expected because the way the users put their navigators in pocket and bags are quite ambiguous. In the case of vision-aided pedestrian navigation, we only need the *texting*-mode and this mode can be detected from accelerometer sensor with the accuracy of almost 82%. In this mode, the orientation of the device (i.e. landscape or portrait mode) can be detected with an accuracy of almost 93%.

Finally, a dataset with two combined user context was collected for testing the total context-aware and navigation solution. The user walked along the side-line of a tennis court in a close loop. During the loop, the user changed the placement twice before and after making turns which represents a very challenging situation for vision navigation. Using the classification algorithm, the system recognized the mode change and adapts the most suitable vision-based heading estimation automatically. Then, to accomplish vision-aided solution, the frame rate of four images per second was used. The resolution of the images was down-sampled to 320x240 pixels. The frame rate of 4 Hz was chosen because the experiments show that it provides sufficient information to capture meaningful motion vectors in different scenarios. A comparison of integrated navigation solutions is shown in Figure 7. The tennis court is located between two buildings and therefore, the smartphone's GPS navigation solution has been degraded. As it can be seen from the figure, without using the context-aware vision-aided navigation, the GPS solution in comparison with the vision sensor is not accurate enough and unable to discern turns.

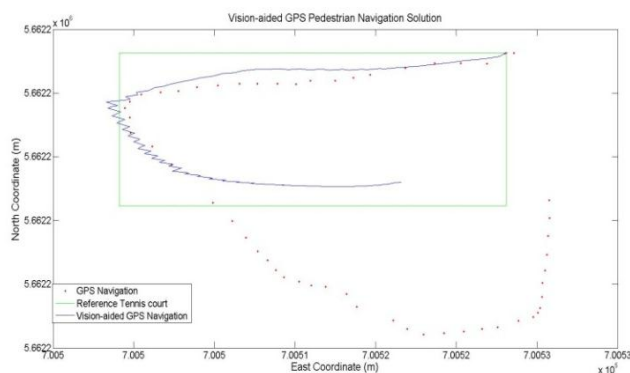


Figure 7: Field test using phone in two modes while user walking around a tennis court: the reference solution (green), GPS position (red), Vision aided GPS navigation (blue)

6. CONCLUSION

This paper concentrates on detecting the most important context information in personal navigation for users carrying smartphones. The field test shows that texting mode (which is the proper mode for vision sensor) can be detected from accelerometer sensor with the accuracy of 82%. In this mode, the orientation of the device (i.e. landscape or portrait mode) can be detected with an accuracy of 93%. Once context detection is performed, proper computer vision algorithm can be applied accordingly to find the motion vectors from successive frames to extract user's motion.

Moreover, a vision-aided pedestrian navigation algorithm is proposed to improve GPS solution. To model the characteristics of the two-dimensional motion of a walking user, Dead Reckoning algorithm is used as a dynamic model in Kalman Filter. The measurements fed to the filter are the GPS positions, velocity and vision-based velocity and the changes in heading angles when available. Pedestrian field tests were performed to verify the algorithms. The results are promising for combined modes and showed great potential for accurate, reliable and seamless navigation and positioning.

7. REFERENCES

- Aggarwal P., Syed Z., Noureldin A., El-Sheimy N., 2010. *MEMS-Based Integrated Navigation*, ARTECH HOUSE, Boston, London.
- Avci, A. a.-P.-P. (2010). Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey. In: *23th Int. Conf. on Architecture of Computing Sys, ARCS 2010*, 22-23 Feb, Germany. (pp. 167-176).
- Baldauf, M., Dustdar, S., & Rosenberg, F., 2007. A survey on context-aware systems. *Int. Journal of Ad Hoc and Ubiquitous Computing archive*, Vol. 2(4).
- Bay H., Ess A., Tuytelaars T., Gool L.V., 2008. SURF: Speeded Up Robust Features, *Computer Vision and Image Understanding (CVIU)*, Vol. 110, No. 3, pp. 346—359.
- Choujaa, D., & Dulay, N., 2008. TRAcME: Temporal Activity Recognition using Mobile Phone Data. *IEEE/IFIP Int. Conf. on Embedded and Ubiquitous Computing*.
- Corke P., Lobo J., and Dias J., 2007. An introduction to inertial and visual sensing, *The International Journal of Robotics Research*, Vol 26, No 6., pp. 519-535.
- Dey, A., 2001. Understanding and Using Context. *Personal and Ubiquitous Computing*, 5(1) Feb 2001, pp.4 7.
- Hide, C., Botterill, T., Anredotti, M., 2011. Low Cost Vision-Aided IMU for Pedestrian Navigation, *Journal of Global Positioning Systems*, Vol.10, No.1 :3-10.
- Fischler M. A. and Bolles R. C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Comm. of the ACM* 24 (6): 381–395.
- Lee, U., & Gerla, M., 2010. A survey of urban vehicular sensing platforms. *Computer Net.*, V.54 , pp. 527-544.
- Mokbel, M., & Levandoski., J. J., 2009. Towards context and preference-aware location-based services. *Proc. of the ACM Int. Workshop on Data Engineering for Wireless and Mobile Access*, (pp. pp. 25–32).
- Pei, L., Chen, R., Liu, J., Chen, W., Kuusniemi, H., Tenhunen, T., et al. (2010). Motion Recognition Assisted Indoor Wireless Navigation on a Mobile Phone. *ION GNSS 2010*. Portland, USA.
- Ruotsalainen L., Kuusniemi H., Chen R., 2011. Visual-aided Two-dimensional Pedestrian Indoor Navigation with a Smartphone, *Journal of Global Positioning Systems*, Vol.10, No.1 :11-18.
- Saeedi, S., El-Sheimy N., Zhao X., Sayed Z., 2011. Context Aware Mobile Personal Navigation Using Multi-level Sensor Fusion, *ION GNSS 2011*, Portland, USA.
- Steinhoff U., Omerčević D., Perko R., Schiele B., and Leonardis A., 2007. How computer vision can help in outdoor positioning, *Proceedings of the European conference on Ambient intelligence*, Springer-Verlag Berlin Heidelberg, LNCS 4794, pp. 124-141.
- Yang, J., Lu, H., Liu, Z., & Boda, P. P., 2010. *Physical Activity Recognition with Mobile Phones: Challenges, Methods, and Applications*. Springer-Verlag London Limited , In Multimpp. 185-213.
- Zadeh, L.A., 1965. "Fuzzy sets". *Information and Control*, 8 (3): 338–353. doi:10.1016/S0019-9958(65)90241-X. ISSN 0019-9958.
- Zhang W., and Kosecka J., 2006, Image based localization in urban environments, *The 3rd Int. Symp. on 3D Data Processing, Visualization and Transmission*, IEEE Computer Society, 14-16 June, 2006, Chapel Hill, NC, USA, pp. 33-40.