Bimodal Localization in Cellular Networks Utilizing Particle Filters

Arash Tabibiazar
Electrical and Computer Eng. Dept.
University of Waterloo
Waterloo, Canada
atabibiazar@uwaterloo.ca

Otman Basir
Electrical and Computer Eng. Dept.
University of Waterloo
Waterloo, Canada
obasir@uwaterloo.ca

Abstract – *Location based services in wireless networks is a quite demanding application especially in urban areas. Cellular network provides measurements regarding the signal attenuations from serving and neighbouring base stations for managing radio resources. Localization based on this inconsistent received signal strength is a challenging problem. This paper describes a novel bimodal localization idea for mobile users in cellular networks. A series of vision-based algorithms are applied to extract user position from monocular vision and then augment it with extracted location in cellular network. A probabilistic framework based on particle filters developed to fuse the bimodal data as well as localize the mobile user precisely from inconsistent measurements. This approach can be easily implemented to utilize available online visual databases to increase accuracy of conventional localization methods for wireless networks even in indoor environments that other navigation signals are not available.*

Keywords: bimodal data fusion, monocular vision, particle filters.

1 Introduction

Localization is a key component in wireless networks. The most popular localization sensor for outdoor is the GPS receiver that allows highly accurate localization. Such accuracy is affected by a number of factors, including satellite positions, noise in the radio signal, atmospheric conditions, and natural barriers in signal path. Unfortunately, in dense urban areas or indoors, buildings can mask received signal and in this case the localization accuracy drops considerably. So it is necessary to use other data types such as visual signals to solve localization problem specially in such places where the odometry is difficult and vision based localization system could make a good complementary sensor to other sensors. In our approach the received signal strength (RSS) information - available in cellular networks - is augmented with extracted position from visual signal to find user position in network precisely. A probabilistic approach based on particle filters is developed for localization and data fusion. One desired application of the proposed method can be delivering the location based services (LBS) to mobile users in cellular network with cellphones equipped to camera. A cellphone with camera in such networks acts as a real-time position sensor. From a single shot (monocular vision) taken by user sent to mobile switching center (MSC) through MMS, relative position is extracted, then the rough estimation of global position - that has already computed through the network - is augmented with calculated position from monocular vision to compute geographical coordination (longitude and latitude) of the user with higher accuracy. In cellular networks, received signal strength by cellphone from 3 to 7 communication towers are frequently reported to mobile switching center for both calling and non-calling users. These information can be used for multilateration to have a rough estimation of mobile user location. Upon receiving a camera shot from mobile user, a task run to find another monocular vision based position of user. We assume that have access to available local maps and videos like ones provided by Google with geographical information tag (longitude and latitude). Recently, localization within a probabilistic framework has achieved good performance in robotics and sensor networks. Probabilistic inference is typically made by observing multiple sets of noisy evidences. Particle filtering uses a probabilistic transition model between a state and a probabilistic observation model to estimate the next state using observed evidences. It represents the posterior distribution by a set of weighted samples (particles), that approximate probability distribution. Most of the existing localization algorithms did not consider the practical situation of multiple measurements and cannot be directly applied. The novelty in this paper is in utilizing particle filters for both localization and fusion of bimodal data that is organized as follows. Section 2 is literature review on related works. In section 3 the proposed method is explained and then evaluated in section 4. Finally, conclusions are given in section 5.

2 Related works and remarks

There is extensive literature on localization methods in wireless networks. In this section we review the monocular vision methods and state-space concepts in wireless networks.

2.1 Monocular vision localization

Simultaneous Localization and Mapping (SLAM) is one of the fundamental problems in many research areas as robotics. Although, there is some commonalities with the researches in vision-based approaches held in the SLAM community but they can be categorized into bottom-up and top-down approaches. They are natural inheritances of structure from motion (SFM) approaches of the vision community and Kalman filter (KF) approaches of the control community. Real-time SLAM using only monocular vision has been achieved by Davison [10], but is feasible only for a small environments with less than 100 landmarks and it is not suitable if the robot needs to travel for hundreds of meters. It is also possible to compute egomotion by using only visual data as done by Nister et al. [11]. In this case, maintaining a large map is not required but the localization accuracy decreases as the distance traveled increases. Moreover, in two successive navigation experiments, the robot may not use the same landmarks and the resulting trajectory may be different. Another approach for achieving robot navigation - according to a human guided experience - represents trajectory as a set of key images that the robot has goes from one key frame to another one. As a possibility, a relative position of the robot with reference to the key frame can be computed from features and match using a wide baseline techniques. After that, a displacement vector is computed to reach the next key frame. Some approaches, first build a map of the environment through off-line learning step. After that, the robot is able to localize itself with this map. This is not always the case in SLAM approaches because map building and localization must be done simultaneously in real-time. The map can be built by using different techniques and different sensors. Cobzas et al. [9] use a rotating camera along with a laser range finder to build a set of panoramic images enhanced with 3D information. With this map, a single 2D image is enough to localize the camera. Kidono et al. [13] build a 3D reconstruction from a stereo camera and an odometer under the assumption that the ground is planar. Then they use the map generated in the 3D reconstruction process to localize the robot in real-time. Monocular vision is utilized in our approach to find the distance of camera from a landmark that have already addressed in several works as [4]. One of the drawback of current approaches in visual SLAM is constructing the search area. We have a tremendous data sources available online like Google. These online sources can be used for SLAM. In our approach, available Google videos or provided videos by the administration offices in sport complex, shopping mall, etc. can be imported for pose estimation from monocular vision[6, 7].

2.2 Filtering for the state-space model

State space models originated from dynamic control theory. A state space model usually consists of two sets of equations, the system equation: $x_{t+1} = f_t(x_t, u_t)$ and the observation equation: $y_t = h_t(x_t, v_t)$, where the mappings $f_t : R^m \times R^p \to R^m$ and $h_t : R^m \times R^q \to R^k$ are assumed to be known. The system (dynamic) equation models the dynamics of state variables and the observation (output) equation models the observed state variables. For a linear Gaussian state-space model, the well-known Kalman filtering approach provides optimal estimates for state variables based on the information from the two sources, the dynamic equations and the observations. However, for nonlinear and non-Gaussian state-space models, it is quite challenging to estimate (as well as filter, smooth and forecast) the state variables and model parameters. In general, the state x_k is frequently evaluated from the conditional probability density function $p(x_k|y_{1:k})$ and a set of measurements $y_{1:k} \equiv \{y_1, ..., y_k\}$ up to sample k by the Chapman-Kolmogorov equation:

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1} \quad (1)$$

For every measurement y_k, the posterior state probability density function can be updated via the Bayes rule:

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_y|y_{1:k-1})}{p(y_k|y_{1:k-1})} \quad (2)$$

where $p(y_k|y_{1:k-1})$ is a normalizing term. This equation can be written in recursive mode as:

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|x_{1:k-1})}{p(y_k|y_{1:k-1})}p(x_{1:k-1}|y_{1:k-1}) \quad (3)$$

The analytical solution to the above equations is intractable. So, many computational methods such as Sequential Monte Carlo (SMC) approaches (A.K.A Particle Filters) and Gibbs sampler techniques have been developed for nonlinear and non-Gaussian state space models in the past decades [14, 12].

3 Proposed method

Our localization method in this paper is a RSS-based localization system, augmented with position information extracted from monocular vision. For estimation and fusion, particle filter is utilized that can be easily implemented in Mobile Switching Centers (MSC) to

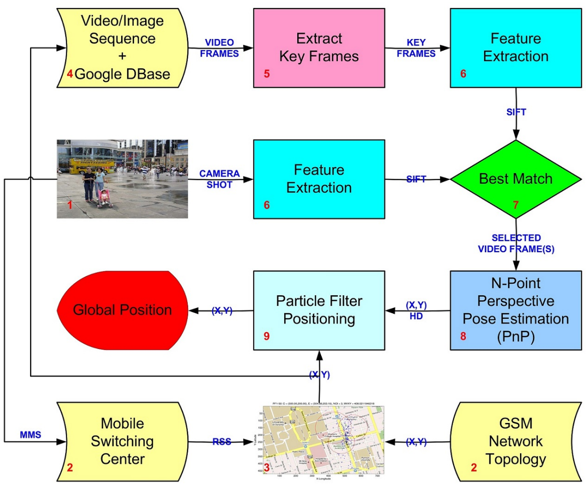

Figure 1: Bimodal localization in wireless networks. Extracted user position from RSS information is augmented by estimated distance between a mobile user and landmark from monocular vision.

find the position of mobile user in the areas with inconsistent navigation signals like indoors. Most of conventional approaches use the mathematical methods like triangulation or multilateration in the first step to extract the coordinates of mobiles for feeding a tracking algorithm like Kalman filter. Figure 1 shows an overall view of system with its sub-modules. For global localization from monocular vision in cellular network, we need to have access to local visual databases and maps with geographical tags like Google maps as well as Network Measurement Report (NMR) files respectively. To extract the real distance between Mobile Stations (MS) and Base Transceiver System (BTS), base station physical geometry is required too. Some of the communication channel models require more specific information about the antenna type, antenna direction, beam width and etc. that usually provided by service providers.

3.1 Monocular pose estimation

The proposed algorithm starts from a single shot taken by mobile user in network. This shot is then used to search with in the available visual databases to match with video frame or image. The first step for video processing is key frame selection. If there is not enough camera motion between two frames, the computation of the epipolar geometry will be an ill conditioned problem. So first we select frames with highest camera motion between frames while still they can be matched. The first frame of the sequence is always selected as a key frame (I_1). The second key frame I_2 must have at least M common interest points with I_1. Then key frame I_{n+1} must have at least M common interest points with I_n and at least N common interest points with I_{n-1}. This ensures that there are enough point matches between key frames to compute camera

Figure 2: Key frames selection in reference video to avoid ill conditioned problem and have maximum camera motion between frames.

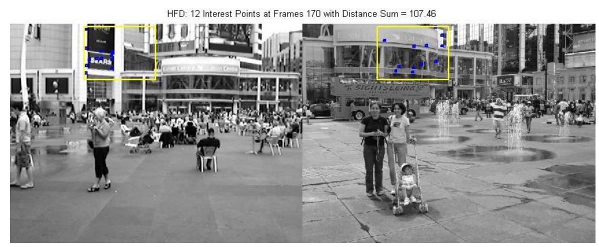

Figure 3: Feature extraction for frame matching. Both images have been divided to M subframes with minimum N features. Hamming Distance can be used to find the best matched frame of local video with the taken shot by the user.

motion (Figure 2). After selecting the key frames, image matching algorithm is run. Image matching is a fundamental part of many problems such as object or scene recognition. The extracted features for matching must be invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint and well localized in both the spatial and frequency domains. In our approach a Scale Invariant Feature Transform (SIFT) method is used to extract features for image matching between taken shot and reference video/image sequences [16]. Figure 3 shows one iteration of feature extraction and matching process. In this result, frames have been divided to 16 sub-regions for feature extraction [3]. Given the perspective projection of three points constituting the vertices of a known triangle in 3D space, it is possible to determine the position of each of the vertices. This problem is an important problem in computer vision (Figure 4) and there are six different solutions for this problem [15]. Each of the six solutions begins from three equations generated by the law of cosines (Equation 4).

$$S_1^2 = L_A^2 + L_B^2 - L_A \times L_B \times \cos\theta_{AB} \tag{4}$$

Solving Perspective-n-Point (PnP) problem is corresponding to find the roots of Equation 5. By solving this equation, unknown distances L_A, L_B and L_C can be defined from known values s_1, s_2 and s_3 and perspective angles θ_{AB}, θ_{BC} and θ_{AC}.

$$G_0 + G_1x + G_2x^2 + G_3x^3 + G_4x^4 = 0 \tag{5}$$

The accuracy of monocular pose estimation depends on several parameters including changes in environment

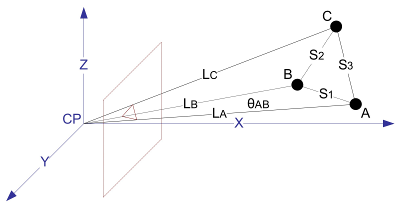

Figure 4: Perspective-3-Point pose estimation. This illustrates the geometry of the three point space resection problem. The problem is to determine the lengths $L_A, L_B \& L_C$ from 3D point positions $A, B \& C$ [15].

like occlusion by parked cars, moving people, trees that can happen from shot to shot [15].

3.2 RSS-based localization

Signal attenuation is the most simple and inexpensive to use in estimating the node-to-node distance; however, they suffer from noisy and inaccurate measurements or delays due to fading channels. Most of the current range-based localization methods, first estimate the distance based on the empirical channel models, and then infer position. Trilateration is the most basic and intuitive way for positioning. This method computes a nodes position via the intersection of three circles. In real-world applications, the distance estimation inaccuracies as well as the inaccurate position information of reference nodes - as depicted in Figure 5 - results in an infinite set of possible positions. Furthermore, when a larger number of reference points are available, we can use multilateration to compute the node position and an over determined system of equations must be solved. Usually, over determined systems do not have a unique solution ($Ax = b$), but can be easily solved using standard methods like the least squares method. In our approach, Equation 6 has been used to extract radius of the circles which mobile user moves inside them with highest probability [1, 2].

$$(x - \frac{k^2 x_2 - x_1}{k^2 - 1})^2 + (y - \frac{k^2 y_2 - y_1}{k^2 - 1})^2 = (\frac{kD}{k^2 - 1})^2$$

$$\log_{10}(k) = \frac{A_1 - A_2}{10n} + N(0, \frac{2\sigma^2(1-\rho)}{100n^2}) \tag{6}$$

In this equation, A_i is signal attenuation (dB), n is path loss exponent, $k = \frac{d_1}{d_2}$ is distance ratio, D specifies distance between two base stations (m), $\sigma^2 = 2.2{-}8.3$ is standard deviation (dB) and $\rho = 0.3 - 0.8$ is correlation coefficient of shadow components of two paths[5, 8].

3.3 Bimodal data fusion

The dynamics of the system can be described by the system equation:

$$x_k = f(x_{k-1}, m_k, u_k) \tag{7}$$

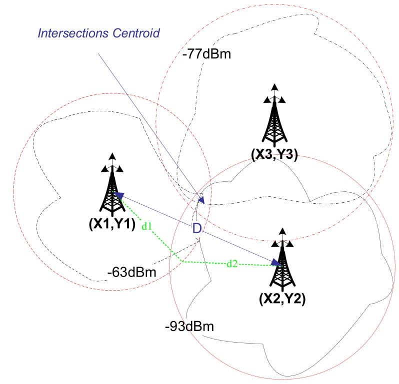

Figure 5: Trilateration in wireless network using inconsistent RSS information. It is a basic building block of localization, however, has not yet overcome the challenges of poor ranging measurements, dynamic and noisy environments and fluctuations in wireless communications.

where $x_k \in R^{n_x}$ is the system base state (position), and $u_k \in R^{n_u}$ is the state noise, with $k \in N$ that N is the set of natural numbers. The modal state m_k characterizes the system different modalities that can take values over a finite set $m_k \in M$. The measurement equation has the form of:

$$y_k = h(x_k, v_k) \tag{8}$$

where $y_k \in R^{n_z}$ is the observation, and $v_k \in R^{n_v}$ is the measurement noise. Functions $f(.)$ and $h(.)$ are nonlinear in general. It is assumed that the observations are taken at discrete time points from either radio or visual signals. For bimodal localization process, we have utilized hybrid particles to fully characterize the state x_k which is frequently evaluated from the conditional probability density function and a set of bimodal measurements $y_{1:k} = \{y_1, y_2, ..., y_k\}$ up to time instant k. A general particle filter method can be applied in three stages for time instances $k = 0, 1, 2, \ldots$ as follows:

- for $i = 0, 1, 2, \ldots, N$, sample $x_k^i \sim p(x_k | x_{0:k-1}^i, y_{0:k})$ that $x_{0:k}^i \equiv (x_{0:k-1}^i, x_k^i)$
- for $i = 0, 1, 2, \ldots, N$, compute weights $\hat{w}_k^i = \hat{w}_{k-1}^i \times p(y_k | x_k^i) p(x_k^i | x_{k-1}^i) / p(x_k | x_{0:k-1}^i, y_{0:k})$
- for $i = 0, 1, 2, \ldots, N$, normalize weights $\tilde{w}_k^i = \hat{w}_k^i / \sum_{j=1}^{N} \hat{w}_k^j$

The estimate $\hat{I_N}(f_n)$ of the posterior expectation $I_N(f_n)$ is obtained using the equation:

$$\hat{I_N}(f_n) = \sum_{i=1}^{N} f_n(x_{0:n}^i) \frac{\hat{w}_k^i}{\sum_{j=1}^{N} \hat{w}_k^j} \tag{9}$$

where $\hat{w}_k^i$ is the non-normalized weight for sample i at time k. Figure 6 shows the working flow of implemented particle filter for data fusion. In this procedure, two main stages can be distinguished: transition

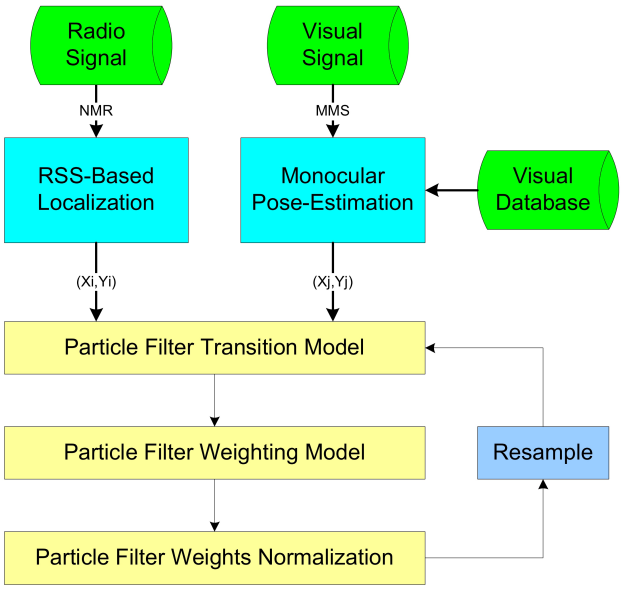

Figure 6: Data fusion workflow by particle filters. Particles are generated based on posteriori distribution probabilities and then updated based on likelihood function.

and resampling. During the transition, each particle is modified according to the state model in presence of noise. Then in the update stage, each particle's weight is re-evaluated based on the recent measurement. A resampling procedure is dealing with the elimination of particles with small weights and replicates them with higher weights. Figure 7 shows the localization results after a few iterations with generated particles. In summary, the proposed scheme can be simply described in the following steps. These steps are applied for every available measurement through received signal strength level or extracted distance from camera shots. This approach has an implied data fusion between different modalities of measured signal.

4 Evaluation

To implement the derived particle filter and localization algorithm, we have developed a simulation framework to design, simulate and evaluate bimodal monocular vision positioning and RSS-based localization methods under this platform simultaneously. The implemented system has several adjustable parameters or characteristics that affects the overall performance as well as computation cost. These parameters include camera parameters, video frames resolution, number of features in SIFT, number of partitions in video frames, window size of each feature for feature matching, number of circles for multilateration, number of particles and error mean threshold. In the following results the original image (shot and video frame) is divided to 16 partitions and for every partition at least 12 features are

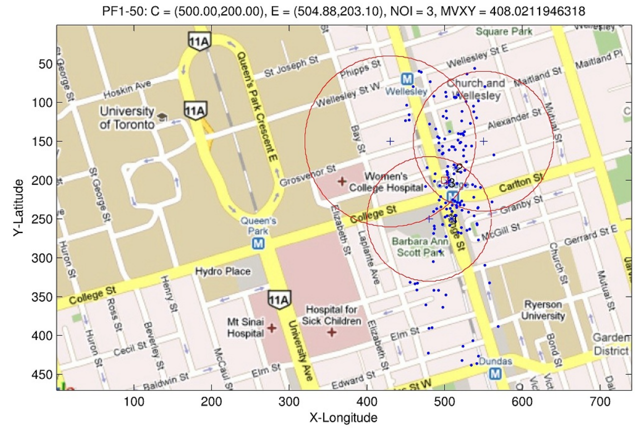

Figure 7: Particle filtering evolution with 50 hybrid particles after 3 iterations around the primary user position, located by trilateration in cellular network.

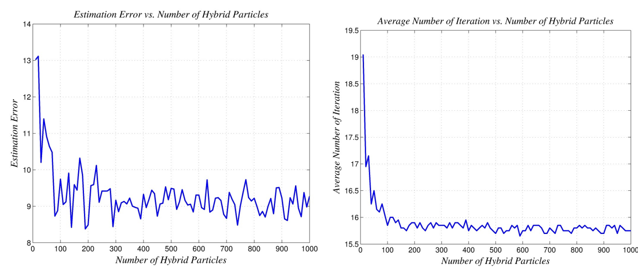

Figure 8: Estimation Error and Number of iterations vs. Number of Hybrid Particles for a typical threshold

extracted. By computing the Hamming Distance (HD) of features in partitions, we find the best match of the taken shot with video frames. These HD values can be used directly for computing the posteriori probabilities. System may fail for some reasons including high occlusion, especially in the regions with highly moving object like people and cars. In such scenes, lot of features (interest points) are assigned to moving objects that are misleading for frame matching algorithm. Figure 8 (left) shows the estimation error versus the number of particle in system. Number of particles play a key role in system convergence time as well as the accuracy of localization. Respectively, Figure 8 (right) shows the average number of iterations versus the number of particles. Figure 9 shows the effect of error mean threshold on estimated error and convergence time.

5 Conclusion

To deliver location based services in wireless networks, we shall first solve the positioning problem from inconsistent measurements. We have proposed an approach to use a single camera shot for positioning of mobile users in cellular networks. Integration of these provided location information with in particle filter shows the possibility of a real-time bimodal positioning system.

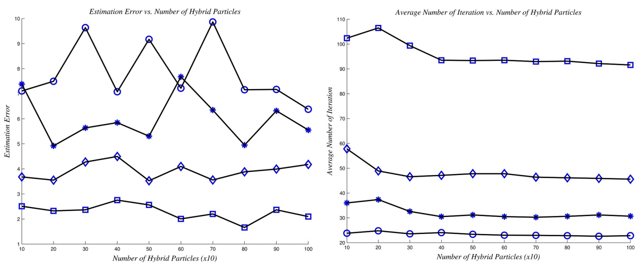

Figure 9: Number of iteration vs. number of particle for different error mean thresholds: 400 (square), 300 (diamond), 200 (asterisk) and 100 (circle).

A series of vision-based algorithms applied for feature extraction, feature matching, perspective-n-point depth inversing and multilateration. Finally, the provided location information or priori probabilities extracted from downlink signal strength or monocular vision camera shot are fused by particle filters.
This approach improves the accuracy of RSS-based methods and augments them with visual data information. The main difficulty with the proposed method is to have a video of the environment up to date. Even if the experiments have shown that this method is robust to some changes, it may not be enough for highly changing environment. Sometimes occlusion for example in a city, parked cars, moving people, trees evolve according to the season, buildings are change from shot to shot.
For future study, other data fusion and inferencing methods may be considered to weigh particles. Another very interesting research topic is to implement a distributed positioning system to run part of system on cellphones.

References

[1] H. Ren and M. Q. Meng, "Power Adaptive Localization Algorithm for Wireless Sensor Networks Using Particle Filter", *IEEE Transactions on Vehicular Technology*, vol. 58, no. 5, pp. 2498–2508, June 2009.

[2] M. Khalaf-Allah, "Nonparametric Bayesian Filtering for Location Estimation and Global Localization of Mobile Terminals in Outdoor Wireless Environments", *EURASIP Journal on Advances in Signal Processing*, January 2008.

[3] X. D. Nguyen, B. J. You and S. R. Oh, "A Simple Framework for Indoor Monocular SLAM", *International Journal of Control Automation, and Systems*, vol. 6, no. 1, pp. 62–75, February 2008.

[4] E. Royer, M. Lhuillier, M. Dhome and J. M. Lavest, "Monocular Vision for Mobile Robot Localization and Autonomous Navigation", *International Journal of Computer Vision*, vol. 74, no. 3, pp. 237–260, 2007.

[5] D. B. Lin and R. T. Juang, "Mobile Location Estimation Based on Differences of Signal Attenuations for GSM Systems", *IEEE Transactions on Vehicular Technology*, vol. 54, no. 4, pp. 1447–1454, July 2005.

[6] M. Pupilli and A. Calway, "Real-time camera tracking using a particle filter", *Proceeding of British Machine Vision Conference*, 2005.

[7] T. Goedeme, T. Tuytelaars, L. V. Gool, G. Vanacker and M. Nuttin, "Feature based omnidirectional sparse visual following", *International Conference on Intelligent Robots and Systems*, pp. 1003–1008, 2005.

[8] V. Otsason, A. Varshavsky, A. LaMarca and E. de Lara, "Accurate GSM Indoor Localization", *Ubicomp*, pp. 141–158, 2005.

[9] D. Cobzas, H. Zhang and M. Jagersand, "Image-based localization with depth-enhanced image map", *International Conference on Robotics and Automation*, 2003.

[10] A. J. Davison, "Real-time simultaneous localisation and mapping with a single camera", *the 9th International Conference on Computer Vision*, 2003.

[11] D. Nister, "An efficient solution to the inverse-point relative pose problem", *Conference on Computer Vision and Pattern Recognition*, pp. 147–151, 2003.

[12] Z. Chen, "Bayesian filtering: From Kalman filters to particle filters, and beyond," *Technical Report, Adaptive System Lab, McMaster University*, 2003.

[13] K. Kidono, J. Miura and Y. Shirai, "Autonomous visual navigation of a mobile robot using a human-guided experience", *Robotics and Autonomous Systems*, vol. 40, no. 2–3, pp. 124–132, February 2002.

[14] A. Doucet, N. de Freitas, and N. J. Gordon, "Sequential Monte Carlo Methods in Practice", *New York, Springer*, 2001.

[15] R. Haralick, C. Lee, K. Ottenberg and M. Nolle, "Review and analysis of solutions of the three point perspective pose estimation problem", *International Journal of Computer Vision*, vol. 13, no. 3, pp. 331–356, 1994.

[16] C. Harris and M. Stephens, "A combined corner and edge detector", *Alvey Vision Conference*, pp. 147–151, 1988.