# ACCURATE POSITIONING SYSTEM BASED ON STREET VIEW RECOGNITION

*Chia-Hsiang Lee, Yu-Chi Su, and Liang-Gee Chen, Fellow, IEEE*

DSP/IC Design Lab., Graduate Institute of Electronics Engineering,
National Taiwan University, Taipei, Taiwan
Email: {ade,steffi,lgchen}@video.ee.ntu.edu.tw

## ABSTRACT

In this paper, an accurate and robust positioning system based on street view recognition is introduced. Vision-based technique is employed for dynamically recognizing shop or building signs on the GPS map. Two mechanisms including view-angle invariant distance estimation and path refinement are proposed for robust and accurate position estimation. Through the combination of visual recognition technique and GPS scale data, the real user location can be accurately inferred. Experimental results demonstrate that the proposed system is reliable and feasible. Compared with 20m error of position estimation provided by the GPS, our system only has 0.97m error estimation.

***Index Terms***— GPS, street view recognition, positioning, navigation

## 1. INTRODUCTION

Self-localization is important in many applications, such as automatic vehicle or pedestrian navigation, robotic path planning, and visually-impaired electronic-aids. The GPS system, which provides localization and huge map scale, is widely utilized in these applications. However, GPS has a fatal defect, positioning inaccuracy, which may be caused by satellite masking, multipath or cloudy weather. The positioning error may increase up to 20 meters when the user moves in urban environments. For many applications, such as navigation services for pedestrian, robot or the visually-impaired, 20 meters is unacceptable. An example is shown in Fig. 1. If a blind person is guided by a GPS system to get to his destination, the inaccurate position would lead him toward the opposite direction. To overcome this problem, an accurate positioning system is important and must be a basic requirement of GPS-based systems for advanced applications.

Many works have been proposed to improve GPS accuracy. These works can be classified into three categories. (1) Several works utilized mathematical models, such as Kalman filter [1], least square model [2] and frequency domain model [3], to eliminate the noise of positioning estimation and predict the most possible location. In general, these models do not have good performance in urban environments because the satellite signal is significantly masked or reflected by crowded buildings. (2) There are also some works combining different sensors to acquire multi-type data. For example, Maya Dawood [4] proposed a vehicle localization system with fusion of an odometer, reckoning sensors, 3D models and visual recognition methods. The approaches of the category are accurate but multi-type data fusion from too many sensors makes the system more complex. (3) Some works [5] [6] adopted visual recognition methods by tracking corresponding points across neighboring video frames for camera pose estimation. Shortages of these works are

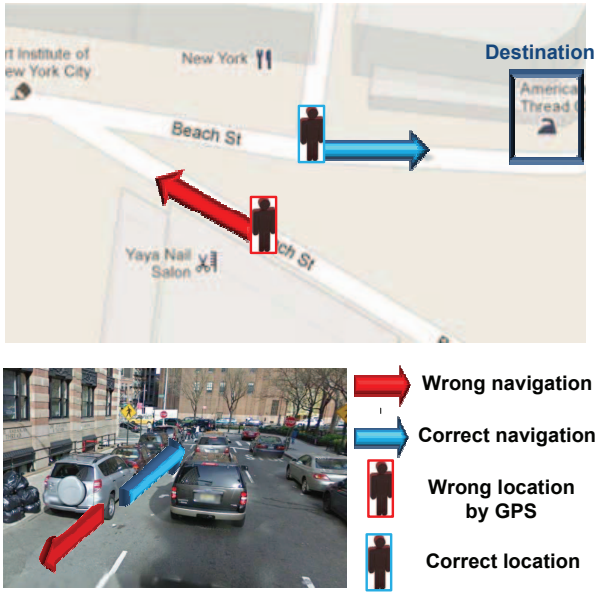


**Fig. 1**. The scenario of incorrect positioning estimation

the assumption of the known initial location and error estimations propagated during a longer period.

In this paper, we propose an accurate and robust positioning system based on street view recognition. Our system is equipped with a GPS receiver and a digital compass to catch global map information and estimate the current user orientation, respectively. Vision-based recognition technique is employed to capture dynamic street views, such as shop or building signs, which are tagged within the GPS map. With the targeted signs around the user on the street, a view-angle invariant distance estimation mechanism is developed to infer accurate user locations. The proposed system can be applied to many advanced applications, such as robot self-localization, visually-impaired aids, augmented reality and general navigation on smart phones.

## 2. THE PROPOSED POSITIONING SYSTEM

### 2.1. Approach Overview

The main idea of our approach is to combine street view recognition with shops or building information tagged within the GPS map. The system flow diagram is shown in Fig. 2. There are one camera, the digital compass and the GPS receiver equipped with the pro-

posed system. Our approach can be divided into two parts: street view recognition and position estimation. The aim of street view recognition is to recognize shop or building signs, which are also tagged on GPS map. In this stage, we utilize feature-based recognition method. Feature extraction and feature matching for the current frame of the input video are performed first. The following stage is position estimation. It includes three main functionalities: view-angle invariant distance estimation, location reconstruction, and path refinement. View-angle invariant distance estimation calculates the distance between the user and the recognized shop signs. Location reconstruction infers the location of the user based on geometrical relationship. Finally, path refinement is executed to correct the moving track of the user by considering both previous motions of the user and the estimated location at each timestamp.

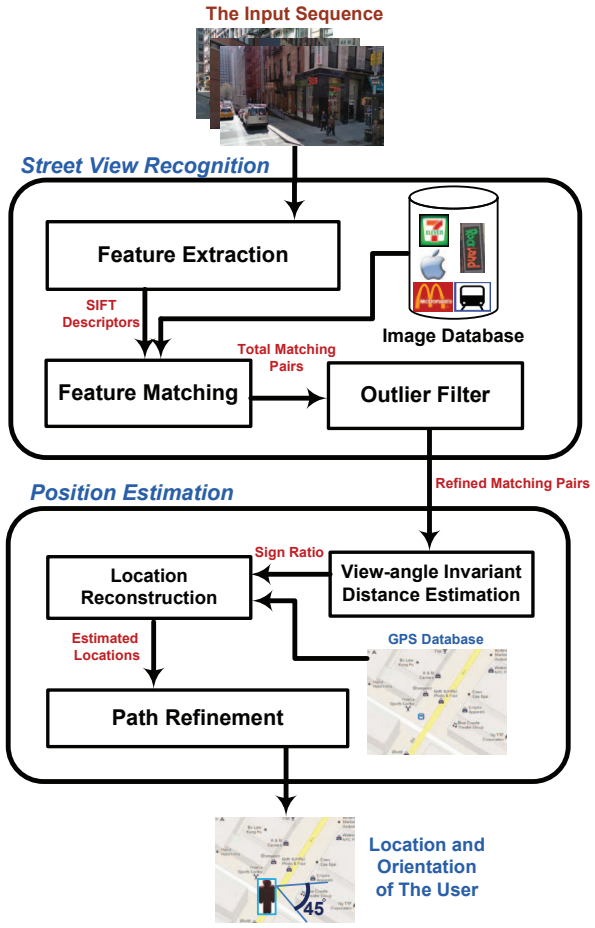


**Fig. 2**. System flow diagram

### 2.2. Street View Recognition

In order to identify shops or buildings on the street while the user is moving, the technique of visual recognition is adopted. Firstly, SIFT [7] features, a local descriptor with good scale, rotation, and luminance invariance, are extracted from the input video. Next, for each feature in the input video, feature matching is performed to find the nearest neighbors among reference images in the database. To speed up the processing time of the matching stage, kd-tree [8] is adopted as the index of the image database. After this step, hundreds of matching pairs for each frame in the input video are obtained. In addition, to remove false matching among matching pairs, RANSAC [9] is employed to filter outlier pairs. The RANSAC algorithm iteratively selects samples at random among matching pairs and estimates their homography matrix as the fitting model. Finally, the remaining matching pairs which fit the model within a user given tolerance are the final answers and sent to the next stage.

### 2.3. Position Estimation

#### *2.3.1. View-angle Invariant Distance Estimation*

To estimate the user position, the first step is to calculate the distance between the recognized shop sigh and the user. Intuitively, the scale ratio of the size of the shop sign in the input video frame to that in the database image, called $SR$, can be utilized to infer the distance. Fig. 3 shows the observation of different shapes of the shop signs under various view-angles from the user. From this figure, we can see that the pixel displacement of y-component keeps in a constant value even when the view-angle of the user changes. Thus, we define $SR$ for each matching pair as shown in the following equation.

$$r_{ij} = |\{y(f_i) - y(f_j)\}/\{y(F_i) - y(F_j)\}|, \quad (1)$$

where $r_{ij}$ is the $SR$ between the i-th and j-th features. $y(f_i)$ is the y-component of the i-th recognized feature $f_i$ and $F_i$ is the the corresponding feature of $f_i$ in the image database

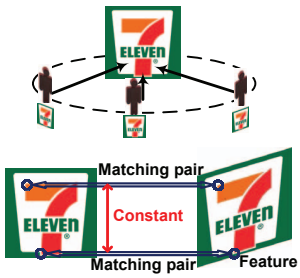


**Fig. 3**. Sign shapes under different view-angles

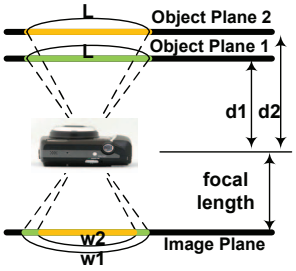


**Fig. 4**. Geometric relationship between the object distance and the projected width in the image plane

However, shapes of shop signs under various view-angles are usually not perfectly transformed. The effect would cause the error of $SR$ estimation. Therefore, a spatial consistency filter is introduced to eliminate outliers with extreme $SR$ values. This filter calculates the average value of $SR$ among all matching pairs. Next, the matching pairs with $SR$ larger than the average value under a given tolerance would be removed. This process repeatedly proceeds until the percentage of inliers is larger than a given threshold. The screening condition could be represented as follows.

$$m_{ij}^k = \begin{cases} 1 & \text{if } \left|\mathrm{r_{ij}} - R^{k-1}\right| < p \times R^{k-1}, \\ 0 & \text{others,} \end{cases} \quad (2)$$

where $p$ is the tolerable error rate of $R^{k-1}$, which is the refined ratio in (k-1)-th iteration. $m_{ij}^k$ is a binary value which decides whether the i-th and j-th matching pair $r_{ij}$ is inlier or not in k-th iteration.

The $r_{ij}$ would be regarded as an inlier if it is closed to $R^{k-1}$. Otherwise, it would not be adopted in the next iteration and the cor-

responding $m_{ij}^k$ would be labeled as zero. After inliers are acquired, $R^k$, the value of $R$, could be updated as follows.

$$R^k = (\sum_{i=2}^{N}\sum_{j=1}^{i-1} m_{ij}^k r_{ij}\ )/(\sum_{i=2}^{N}\sum_{j=1}^{i-1} m_{ij}^k\ ), \tag{3}$$

where $N$ is the total number of matching pairs and $R^k$ would be iteratively refined until the following condition is reached.

$$\begin{aligned}\sum_{i=2}^{N}\sum_{j=1}^{i-1} m_{ij}^k &>= n(n-1)/2 - an(an-1)/2 \\ &= n(1-a)(an+n-1)/2,\end{aligned} \tag{4}$$

where $a$ is the tolerable error rate of outliers.

Eq. 2 eliminates extreme $SR$ values. The $R^k$ would be iteratively refined by preforming operations shown in Eq. 2 and 3 until the condition illustrated in Eq. 4 is reached. Next, $SR$ values of the recognized shop signs are calculated and the corresponding distance from the user could also be estimated by Eq. 5. The geometrical relationship is shown in Fig. 4,

$$d1 = \frac{w2}{w1} \times d2 = R \times d2, \tag{5}$$

where $d1$ is the distance between the user and shop signs and $d2$ is the constant value stored in the image database. $w$ is the pixel width of shop signs appearing in the video frame.

Refer to the $d2$ and $R$, $d1$ could be obtained and it would be sent to the next stage to reconstruct the user location.

### 2.3.2. *location Reconstruction*

In order to estimate current user location on the GPS map, the estimated distance between shop signs and the user should be combined with the information of the user orientation. Fig. 5 illustrates definitions of the user orientation, pattern angle $\theta_{pi}$ and view angle $\theta_h$. $\theta_h$ can be acquired by the digital compass. $\theta_{pi}$, representing the pattern angle of the i-th recognized shop signs, can be calculated by the pixel position of the signs. Combining the user orientation, estimated distance from shop signs, and GPS information, the user location in the global map could be estimated by Eq. 6.

$$\begin{aligned} x &= \sum_{i=1}^{Np} [\ x_{pi} - d_i \cos(\theta_h + \theta_{pi})]/Np \\ y &= \sum_{i=1}^{Np} [\ y_{pi} - d_i \sin(\theta_h + \theta_{pi})]/Np \end{aligned} \tag{6}$$

where $Np$ is the number of recognized sings. The $x_{pi}$ and $y_{pi}$ are the coordination of the i-th recognized signs which are stored in GPS database. If $Np > 1$, we calculate the mean of these reconstructed locations.

### 2.3.3. *Path Refinement*

Although the user location could be estimated in the previous stage, there is still small error of $R$. Therefore, Kalman filter is adopted to refine the path. It can combine all estimations and previous status to predict the most possible position. In this paper, the previous motion trajectory, GPS raw data, and the estimated distance from shop signs are combined to predict accurate location of the user. The following equations are the main mathematical operations in Kalman filter.

$$\hat{x}_t = A\hat{x}_{t-1} + Bu_{t-1}, \tag{7}$$

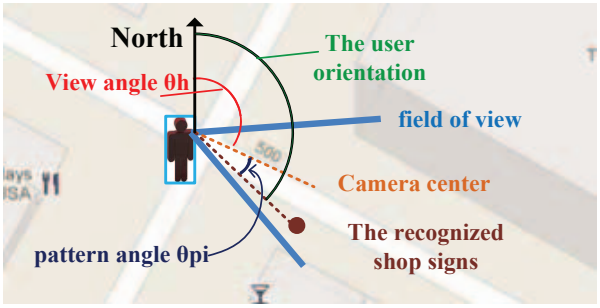


**Fig. 5**. Definition of the user orientation, pattern angle, and view angle

$$x_t = \hat{x}_t + K_t(z_t - H\hat{x}_t), \tag{8}$$

where the $\hat{x}_t$ is the predicted user location according to the previous motion on $t$-th. The $u_{t-1}$ is the previous motion trajectory of the user on ($t$-1)-th. The $x_{t-1}$ is the previous decision of user location. The $z_t$ is the GPS raw data and the reconstructed location presented in section 2.3.2. The $K_t$ is the confident ratio for the predicted location.

## 3. EXPERIMENTAL RESULTS

In the proposed work, real sequences captured from a CMOS front-mounted camera with 1920x1080 resolution video input are utilized. Firstly, a user walks along several streets and his walking paths are recorded as correct answers. While the positioning information outputs from GPS, the proposed system is tested simultaneously with the GPS information. Two experiments are conducted to compare positioning accuracy between GPS and the proposed system.

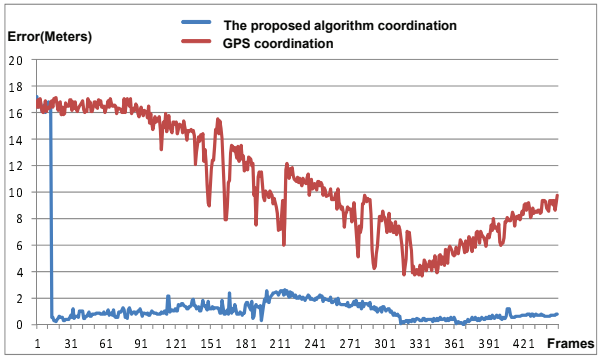


**Fig. 6**. Accuracy comparison between GPS and the proposed method.

The first experiment compares the error of location estimation calculated by GPS and the proposed system. The experimental result is shown in Fig. 6. From this figure, we can see that estimated error of the proposed system is the same as GPS in first 16 frames. It is reasonable because there are few recognized shop signs utilized for location reconstruction in the beginning. As the number of shop signs appears in subsequent frames increases, the estimated errors of locations by the proposed system are below 2 meters. Compared with location information provided by GPS with 10 meters error on average, the proposed system significantly improves positioning accuracy.

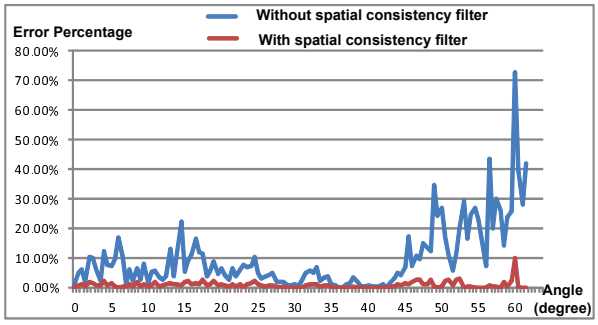


**Fig. 7**. Comparison of the error rate in $SR$ estimation

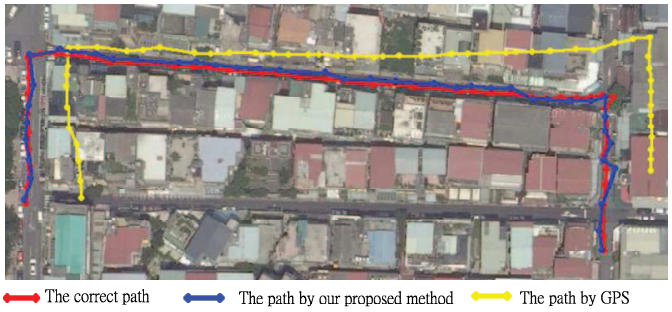


**Fig. 9**. Path estimation by different methods

The accuracy of $SR$ is important for distance estimation and location reconstruction. In the second experiment, we measure the improvement of $SR$ estimation by the stage of view-angle invariant distance estimation. Fig. 7 shows the comparison result. The y-axis represents the percentage of the error estimation to the real value for the $SR$. For example, if the estimated $SR$ is 4.9 and the real value is 5, the error percentage is 0.1/5 = 2%. The x-axis represents the view-angle of the detected signs. The view-angle of the facade of the sign is regarded as zero degree. From Fig. 7, we can see that without the proposed spatial consistency filter, the error percentage increases dramatically as the view-angle is larger than 40-degree. On the contrary, with the proposed spatial consistency filter, the error percentage is 0.8% on average even when the view-angle increases. This result proves that the view-angle invariant distance estimation method supports accurate $SR$ and correct distance estimation. Fig. 8 plots the estimated user location on the GPS map and shows the corresponding street view captured by the camera. Based on the spatial relationship between the user and recognized shop signs, the proposed system supports accurate positioning in practice. Fig. 9 depicts paths estimated by GPS and the proposed system. It is obvious that the estimated path provided by the proposed system is closer to the correct answer.


**Fig. 8**. The information on the GPS map and the corresponding recognized shop signs in the video frame

## 4. CONCLUSION

We propose an accurate and robust positioning system based on street view recognition. With combination of dynamic street view recognition, huge GPS map and shops information, the proposed system provides accurate user locations. The view-angle invariant distance estimation and path refinement mechanisms are proposed to achieve positioning with high precision. Compared with 20 meters error of localization results provided by GPS, the error of the proposed system is 0.97m on average. Experimental results demonstrate our system is reliable and feasible. In addition, the proposed system can be applied to many innovative navigation systems.

## 5. REFERENCES

[1] S. Yamaguchi and T. Tanaka, "Gps standard positioning using kalman filter," in *SICE-ICASE, 2006. International Joint Conference*, oct. 2006, pp. 1351 –1354.

[2] A.M. Bilgicl Y. Hel, R. Martinl, "Approximate iterative least squares algorithms for gps positioning," in *Signal Processing and Information Technology (ISSPIT), 2010 IEEE International Symposium on*, dec. 2010, pp. 231 –236.

[3] M.R. Mosavi, "Frequency domain modeling of gps positioning errors," in *Signal Processing, 2006 8th International Conference on*, nov. 2006, vol. 4.

[4] M. Dawood, C. Cappelle, M.E. El Najjar, M. Khalil, and D. Pomorski, "Vehicle geo-localization based on imm-ukf data fusion using a gps receiver, a video camera and a 3d city model," in *Intelligent Vehicles Symposium (IV), 2011 IEEE*, june 2011, pp. 510 –515.

[5] M. Shah A. Hakeem., R. Vezzani and R. Cucchiara, "Estimating geospatial trajectory of a moving camera," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, 0-0 2006, vol. 2, pp. 82 –87.

[6] F. Lin, B.M. Chen, and T.H. Lee, "Vision aided motion estimation for unmanned helicopters in gps denied environments," in *Cybernetics and Intelligent Systems (CIS), 2010 IEEE Conference on*, june 2010, pp. 64 –69.

[7] D.G. Lowe., "Distinctive image features from scale-invariant keypoints," in *International Journal of Computer Vision*, 2004, vol. 60, pp. 91–110.

[8] S. Wess, K. Althoff, and G. Derwand, "Using k-d trees to improve the retrieval step in case-based reasoning," in *Topics in Case-Based Reasoning*, Stefan Wess, Klaus-Dieter Althoff, and Michael Richter, Eds., vol. 837 of *Lecture Notes in Computer Science*, pp. 167–181. Springer Berlin / Heidelberg, 1994.

[9] M.A. Fischler and R.C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," in *Communications of the ACM*, June 1981, vol. 24, pp. 381–395.