WI-VI FINGERPRINT: WIFI AND VISION INTEGRATED FINGERPRINT FOR SMARTPHONE-BASED INDOOR SELF-LOCALIZATION

Zhaozheng Hu, Gang Huang, Yuezhi Hu, Zhe Yang

ITS Research Center, Wuhan University of Technology, Wuhan 430063, China

ABSTRACT

Due to limited access to GPS signals, accurate and reliable localization still remains an open problem in indoor environments. This paper proposed a novel WiFi and Vision integrated fingerprint (Wi-Vi Fingerprint) for accurate and robust indoor localization. The method consists of two steps of fingerprint mapping and fingerprint localization. In the mapping step, the Wi-Vi fingerprints for all the sampling positions are computed by using EXIT signs as landmarks. In the localization step, a multi-scale localization strategy is proposed, which includes coarse localization with WiFi matching, image-level localization with visual matching, and finally metric localization for localization refinement. The proposed method has been tested in an indoor office building of 11,200 m^2 with different types of smartphones. Experimental results demonstrate that the proposed method can achieve 96% site recognition rates for image-level localization. The final localization errors after metric localization are less than half meter in average.

Index Terms—Indoor self-localization, visual features, Wi-Vi fingerprint, multi-scale localization

1. INTRODUCTION

Self-localization is a key technology for various industry applications, such as robots, location based services (LBS), navigation and path planning, etc. Nowadays, there are many mature solutions to outdoor solutions, such as the Global Positioning System (GPS), Beidou Systems, etc. However, indoor localization is much more challenging due to the limited access to GPS or Beidou signals.

Indoor localization has been intensively investigated due to the great demand in practice for the past decades. Researches on indoor localization can be mainly classified into three categories: 1) wireless network based; 2) sensor-based; and 3) vision-based methods [1-3]. WiFi fingerprint is the most popular method in the first category [4-5]. The principle is to generate the fingerprints from the Received Signal Strength Indication (RSSI) values from Access Points (APs). The collected fingerprint is then matched within the fingerprint map for accurate localization [6-8]. Niu et al [6] proposed a WiFi fingerprint and crowd-sourcing based indoor localization system called WicLoc. They collected WiFi, gyroscope and accelerometer data, and designed a fingerprint model to form fingerprints of each interested location. This system can achieve room-level localization with 5 *m* errors in average. Pedestrian Dead Reckoning (PDR) is a typical sensor-based localization method. PDR measures and statistics pedestrian's steps, step length and direction by using internally installed sensors in a smartphone, and calculate the walking track of pedestrian [10]. However, PDR only give out the position relative to the initial position. For a practical smartphone-based PDR, the localization errors dramatically increase with time. Image-based or vision-based localization have been paid more and more attention in indoor localization [11-14]. Wael et al [11] proposed using vision-based indoor localization method based on local feature computation and matching. They used Harris corners as local features and the ZNCC as descriptor. Guan et al [13] proposed a SURF and landmark based indoor localization method. They exploited SURF matching and homography computation for positioning. However, these methods are computational expensive and cannot achieve real-time localization especially on mobile devices. In addition, they are not robust enough because of the varying environments and flexible poses of the cameras when taking the pictures.

In this paper, we proposed a novel method for indoor localization. The core of this paper is to develop a novel Wi-Fi fingerprint for fast and accurate indoor localization. Contributions of the paper are as follows: 1) We proposed a Wi-Vi fingerprint by integrating WiFi fingerprint and visual features. Based on Wi-Vi fingerprints, we propose a multi-scale method for indoor localization; 2) The proposed method does not need additional equipment or modifications of the environments; 3) The proposed method uses EXIT signs for Wi-Vi fingerprint generation. Exit signs are ideal indoor landmarks and easily available within sight distance. EXIT signs also provide suitable sampling positions for WIFI fingerprint generation. Finally, the developed Wi-Vi fingerprints are easily integrated with other mainstream methods to dramatically enhance localization results.

2. WI-VI FINGERPRINT AND WI-VI MAP GENERATION

The proposed Wi-Vi fingerprint is the integration of WiFi features, sign visual features, and scene visual features (see

Fig. 1). The physical meaning of Wi-Vi fingerprint is that the WiFi signals and visual features from the images of the scenes and EXIT sign area are unique for one position. In other word, Wi-Vi fingerprint and the corresponding position are uniquely corresponded. Therefore, it is possible for self-localization from an input Wi-Vi fingerprint.

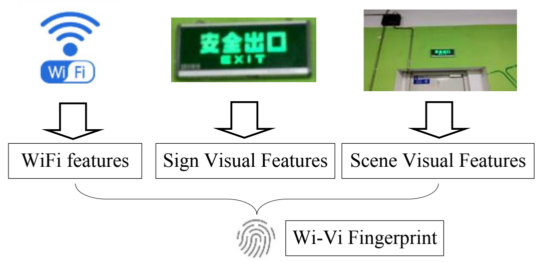

Fig. 1: Wi-Vi Fingerprint: WiFi and Vision Integrated Fingerprint computation by using EXIT sign as landmark.

We proposed using EXIT signs as the reference to sample the areas. The EXIT signs play important roles in the implementation. First, EXIT signs are enforced by fire law to distribute densely in the indoor situations. Hence, users can easily find EXIT signs within sight distances. Second, EXIT signs can provide reference localization information. Third, we set a scanning window in the image to box out sign area. We require that sign area approximate the scanning window as much as possible. As a result, users are slightly restricted to take images within expected distances.

The Wi-Vi fingerprint based method consists of two steps of mapping and localization. In the mapping step, every EXIT sign is a sampling site. We collected the WiFi and image data three times in every sampling site for the purposed of improving localization accuracy.

2.1. WiFi features computation

Thanks to the rapid progress in mobile devices, we can easily collect WiFi signals and the data of the APs with a smartphone. The data of an AP contains a lot of information, such as MAC address, Network SSID, RSSI strength values, etc. Among these collected data, we keep only the MAC address and the RSSI strength values for each AP. The MAC address is used for AP matching, while the RSSI values are for signal strength matching.

2.2. Visual Features Computation

Visual features consist of visual features from signs and visual features from the scene images. Visual features from signs are for the purpose of sign verification. Visual features from the scene images are for the purpose of sampling position recognition. For implemented on smartphones, visual feature extraction and matching method should be computationally effective enough. In this paper, we convert the image into a patch image. The center of the patch image is used as feature point position. And we compute the ORB descriptor as the visual features. "ORB" is the combination of oFAST (FAST with orientation) and rBRIEF (rotated BRIEF) [15].The moment of a patch image $I(x, y)$ is defined as follows

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \tag{1}$$

where m_{pq} is defined as the $(p+q)$-th-order moment of the image. The centroid of the image is computed as follows

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right) = \left(\frac{\sum_{x,y} xI(x,y)}{\sum_{x,y} I(x,y)}, \frac{\sum_{x,y} yI(x,y)}{\sum_{x,y} xyI(x,y)}\right) \tag{2}$$

The orientation of the patch image is thus defined from the moments as $\theta = \operatorname{atan2}(m_{01}, m_{10})$, where atan2 is the quadrant-aware version of arc tan. The BRIEF descriptor is a bit string description of an image patch. It is constructed from a set of binary intensity tests. In this paper, we chose a different strategy to generate the feature descriptor by comparing the difference of two pixel values to a threshold as follows

$$\tau(P; x, y) := \begin{cases} 1: P(x) - P(y) < \gamma \\ 0: P(x) - P(y) \geq \gamma \end{cases} \tag{3}$$

where $P(x)$ is the intensity of the image P at the point x, γ is a threshold value. The BRIEF descriptor is thus defined as a vector of n binary tests

$$f_n(P; x, y) := \sum_{i=1}^{n} 2^{i-1} \tau(P; x, y) \tag{4}$$

In this paper, we use a Gaussian distribution around the patch center and set the vector length n=256 [14]. Hence, an ORB feature descriptor is represented with a 256-bit string. In order to make the BRIEF descriptor invariant to rotation, we steer BRIEF according to the orientation of the key point. For any feature set of n binary tests at location, we define a $2 \times n$ matrix $S = \begin{pmatrix} x_1, \dots, x_n \\ y_1, \dots, y_n \end{pmatrix}$.From the orientation θ, we can compute the corresponding rotation matrix R_θ as

$$R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \tag{5}$$

We construct a "steered" version S_θ of S as $S_\theta = R_\theta S$. Then we can compute the ORB descriptor as follows

$$g_n(P, \theta) := f_n(P)(x_i, y_i) \in S_\theta \tag{6}$$

2.3. 3D coordinates for EXIT signs and coordinate system configuration

3D coordinates are also required in the step of map generation, which play the key role to provide position references. In our implementation, the 3D coordinates of the. In order to assure precise 3D coordinates, we manually measured 3D coordinates for all EXIT signs within the area.

$$[X_i \;\; Y_i \;\; Z_i]^T = R_0[x_i \;\; y_i \;\; 0]^T + t_0 \tag{7}$$

where $[X_i \;\; Y_i \;\; Z_i]$ and $[x_i \;\; y_i \;\; 0]$ are the coordinates of the same points on an EXIT sign. From Eq. (7), we can map the position into a global coordinate system on the map.

3. WI-VI FINGERPRINT BASED LOCALIZATION METHOD

In the localization step, we proposed a multi-scale localization strategy, which includes coarse localization with WiFi matching, image-level localization with visual matching, and metric localization for localization refinement.

3.1. Coarse localization with WiFi fingerprints

The coarse localization area is determined from WiFi matching (or the conventional WiFi fingerprint matching). In this paper, we proposed a novel matching method to compare two WiFi fingerprints as follows

$$M(x, y) = \prod_{i=1}^{N}(1 + \lambda)\exp(-((x_i - y_i)/x_i)^2) \tag{8}$$

where $\lambda \geq 0$, x_i and y_i are the RSSI values for the i-th matched APs from the collected WiFi signals x and y, N is the total number of matched APs. We compute a number of candidate positions that have the highest matching scores by using Eq. (8). As a result, we can select these candidate positions for an input WiFi fingerprint such that we can narrow down the searching spaces with WiFi fingerprint matching to improve the localization accuracy and computational efficiency.

3.2. Image-level based localization

Based on the coarse localization results from WiFi matching, we applied image-level based localization by visual feature matching. As discussed in Section 2.2, we proposed using a single ORB feature descriptor to represent the scene. In the localization step, we also use ORB descriptors to compute scene visual features. As an ORB descriptor is a 256-bit string, we use Hamming distance as the metric to compute the similarity between two ORB descriptors as follows

$$D_R(H^1, H^2) = \sum_{i=1}^{256} XOR(H_i^1, H_i^2) \tag{9}$$

With Eq. (9), we can compute the Hamming distances of an input image with all the candidates computed from coarse localization results. Therefore, we can compute a total of N numbers of candidates that has the minimal Hamming distances with the input. As there are three sampling positions for each site in the mapping step, some derived candidates correspond to the same site. Hence, we proposed using the K-Nearest Neighbors (KNN) to derive the final site from the N candidates. This step is thus called image-level localization which allows us to determine the correct site (or EXIT sign) for more accurate localization.

3.3. Metric localization

Metric localization is finally applied to refine the localization results. As the map includes 3D coordinates of all EXIT signs, we can derive the absolute position from image-level localization results. Before metric localization, we first establish feature point correspondences across the localization image and the matched image from the map. The correspondence results are further corrected by using RANSAC method as $[u_m \;\; v_m \;\; 1]^T \cong H_{ml}[u_l \;\; v_l \;\; 1]^T$. The relation between the matched image from the map and the physical plane is $[u_m \; v_m \;\; 1]^T \cong H_m[X \;\; Y \; 1]^T$. As a result, we can establish the correspondence between the physical plane and the localization image as follows

$$[u_l \;\; v_l \;\; 1]^T \cong {H_{ml}}^{-1} H_m [X \;\; Y \; 1]^T = H_l[X \;\; Y \;\; 1]^T \tag{10}$$

From projective geometry, the homography is $H_l \cong K[r_1 \;\; r_2 \;\; 1]^T$.Therefore, the pose of the camera with respect to the EXIT sign is computed as follow

$$\begin{cases} R = \left[\frac{K^{-1}H_l^{(1)}}{\left\|K^{-1}H_l^{(1)}\right\|} \; \frac{K^{-1}H_l^{(2)}}{\left\|K^{-1}H_l^{(2)}\right\|} \; \frac{K^{-1}H_l^{(1)}}{\left\|K^{-1}H_l^{(1)}\right\|} \otimes \frac{K^{-1}H_l^{(2)}}{\left\|K^{-1}H_l^{(2)}\right\|}\right] \\ t = \frac{K^{-1}H_l^{(3)}}{\left\|K^{-1}H_l^{(1)}\right\|} = \frac{K^{-1}H_l^{(3)}}{\left\|K^{-1}H_l^{(2)}\right\|} \end{cases} \tag{11}$$

The computed pose can be finally mapped into the global coordinate systems.

$$P = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{smallmatrix}\right)(-R_0 R^{-1} t + t_0) \tag{12}$$

where $(R_0 \;\; t_0)$ is the transform matrix, $\left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{smallmatrix}\right)$ is the orthogonal projection matrix that projects the center of the phone camera onto the ground plane.

4. EXPERIMENTAL RESULTS

The proposed method had been tested in a two-floor office building, located on Yujiatou Campus, Wuhan University of Technology (as shown in Fig. 2). The total area is about 11,200 m^2. The WiFi signals are variable due to a large number of floating people. Moreover, the numbers of APs are also varying in different time. Besides, the illuminations vary dramatically in different times of a day. Hence, these conditions make it challenging for existing WiFi or vision-based methods. All EXIT signs were inventoried in advance to collect the Wi-Vi fingerprints and the corresponding 3D coordinates for map generation.

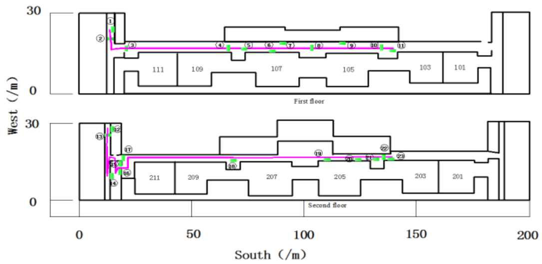

Fig. 2: Map of the test site with all EXIT signs represented as green circles and test points represented as purple line.

The proposed localization systems were implemented on Android platforms with the developed APP installed on different mobile devices for indoor localization. The user is expected to put the phone camera to shoot the exit sign as shown in Fig. 3. And sign area boxed out by a white selection window on the User Interface (UI).

Fig. 3: Illustration of the proposed method with smartphone

In the experiment, 23 EXIT signs in the scene were randomly assigned to three users by using different smartphones (SAMSUNG T710, MIUI 1, and MIUI 2A). For each sign, 10 trials were performed by the user in different time and poses. In the coarse localization, a total of 10 candidates were derived from WiFi Fingerprint matching. The image-level localization was performed thereafter for site localization. We define site recognition rate, which is the ratio of the correct trial number to total trials. Fig. 4 illustrates the image-level localization results. It can be observed that the proposed method can achieve 96% site recognition rate for image-level localization.

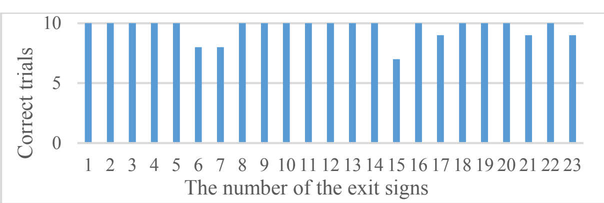

Fig. 4: Image-based localization (site recognition) results

The image-level localization results were finally refined by using metric localization. CDF (Cumulative Distribution Function) of final localization errors are shown in Fig. 5. From Fig. 5, it can be observed that the percentage of the localization error less than 0.2 *m* is 78%, and the percentage of the localization errors less than 1 *m* is 96%. The results demonstrate that the proposed method can generate accurate and reliable localization results.

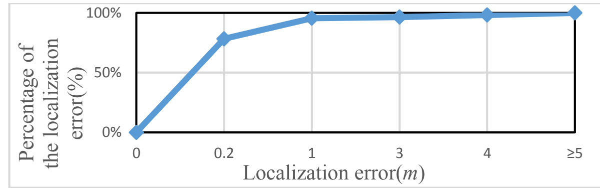

Fig. 5: CDF of localization errors after metric localization

We compared localization error results and average time consuming with the methods in [7], [11] and [13] by using the same dataset and setup (see Table 1). From Table 1, it can be observed that average localization error of the proposed method is 0.3 *m* which is smaller than that of WiFi [7] (3.5 *m*), Harris-ZNCC [11] (0.8 *m*) and SURF based method [13] (1.0 *m*). The percentage of the localization error less than 1 *m* of the proposed method is higher than that of [7], [11] and [13]. And the average time consuming of the proposed method is only 0.1 *s* which means the proposed method can achieve real time localization.

Method	Average time consuming(*s*)	Average localization error(*m*)	Error ≤1 *m*
Proposed method	0.1	0.3	96%
Method in [7]	0.1	3.5	30%
Method in [11]	1.1	0.8	77%
Method in [13]	87.5	1.0	94%

Table 1: Localization error results and time consuming comparison

5. CONCLUSION

This paper has proposed a method for smartphone-based indoor self-localization by introducing a novel Wi-Vi fingerprint. The method consists of two steps of mapping and localization. In the mapping step, the Wi-Vi fingerprints for all the sampling positions are computed. In the localization step, we proposed a multi-scale localization strategy with WiFi matching, visual feature matching, and finally metric localization. The results demonstrate that the proposed method can achieve real time localization on smartphones, and the localization errors after metric localization are less than half meter in average. The proposed method outperforms existing state-of-the-art methods in the comparative experiment.

6. ACKNOWLEDGEMENTS

The work presented in this paper was funded by the National Natural Science Foundation of China (No.51679181), the Major Project of Technological Innovation in Hubei Province (No.2016AAA007), and the Science-technology Funds for Overseas Chinese Talents of Hubei Province (No.2016-12).

7. REFERENCES

[1] Q. Liu, J. Qiu and Y. Chen, "Research and development of indoor positioning," *China Communications*, vol.13, pp.67-79, 2017.

[2] D. Gabrie, C. Kevin and C. Joan, "A Survey of Active and Passive Indoor Localization Systems," *Computer Communications*, pp.1939-1954, 2012.

[3] Z. Deng, Y. Yu, X. Yuan, N. Wan and L. Yang, "Situation and Development Tendency of Indoor Positioning," *China Communications*, pp.42-55, 2013.

[4] C. Wu, Z. Yang, Y. Liu and W. Xi, "WILL: Wireless Indoor Localization without Site Survey," *IEEE Transactions on Parallel and Distributed Systems*, vol.24, no.4, pp.839-848, 2012.

[5] C. Yang and H. Shao, "WiFi-based indoor positioning," *IEEE Communications Magazine*, vol.53, no.3, pp.150-157. 2015.

[6] J. Niu, B. Wang, L. Cheng, and J. J. Rodrigues. "WicLoc: An indoor localization system based on WiFi fingerprints and crowdsourcing," *Proceedings of2015 IEEE International Conference on Communications*, pp.3008-3013, 2015.

[7] W. Li, L. Jin, and X. Chen. "Indoor positioning system design and implementation based on Android platform," *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, pp. 422-424, 2013

[8] I. Bisio, M. Cerruti, F. Lavagetto, M. Marchese, M. Pastorino, A. Randazzo and A. Sciarrone, "A Trainingless WiFi Fingerprint Positioning Approach Over Mobile Devices," *IEEE Antennas and Wireless Propagation Letters*, vol. 13, pp.832-835. 2014.

[9] W. Li, D. Wei, H. Yuan. "A novel method of WiFi fingerprint positioning using spatial multi-points matching," Proceedings *of the 2016 International Conference on Indoor Positioning and Indoor Navigation* (*IPIN*). 2016.

[10] Q. Tian, Z. Salcic, K. I. Wang and Y. Pan, "A Multi-Mode Dead Reckoning System for Pedestrian Tracking Using Smartphones," *IEEE Sensors Journal*, vol. 16, no.7, pp. 2079-2093, 2016.

[11] W. Elloumi, A. Latoui, R. Canals. "Indoor Pedestrian Localization With a Smartphone: A Comparison of Inertial and Vision-Based Methods," *IEEE Sensors Journal*, vol.16, no.13, pp. 5376-5388, 2016.

[12] H. Xue, L. Ma and X. Tan, "A fast visual map building method using video stream for visual-based indoor localization," *Proceedings of Wireless Communications and Mobile Computing Conference* (*IWCMC*), pp.650-654, 2016.

[13] K. Guan, L. Ma, X. Tan,"Vision-based indoor localization approach based on SURF and landmark," *Proceedings of the 2016 Wireless Communications and Mobile Computing Conference* (*IWCMC*), pp.655-659, 2016.

[14]W. Elloumi,K. Guissous, A. Chetouani and S. Treuillet, "Improving a vision indoor localization system by a saliency-guided detection," *Proceedings of Visual Communications and Image Processing*, pp. 149-152, 2014.

[15] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," *Proceedings of IEEE International Conference on Computer Vision*, pp. 2564–2571, 2011.