The Eleventh International Conference on Advanced Computational Intelligence,
June 7-9, 2019, Guilin, China

Indoor Position Algorithm Based on the Fusion of Wifi and Image

Zhongshuai Wang[1,2], Pheng Sokliep[1,2],Chengpei Xu[3], Jiayu Huang[2], Linfa Lu[1,2], Zhuo Shi[1,2]
[1]Guangxi Key Laboratory of Intelligent Processing of Computer Images and Graphics
[2]National and Local Joint Engineering Research Center of Satellite Navigation and Location Service
Guilin University of Electronic Technology, Guilin 541004, China.
[3]School of Data and Computer Science, National Engineering Research Center of Digital Life, Sun Yat-sen University, China
Corresponding author: 22605918@qq.com

***Abstract*—In this paper, an image and WiFi multi-mode fusion localization method is proposed. WiFi fingerprint localization is one of the most commonly used methods at present, but WiFi fingerprint localization has the phenomena of multipath and non-line-of-sight, signal fluctuation, scattering and so on, which leads to the unstable positioning effect. In addition, the image uses feature matching to determine the user's location, which is less affected by the environment, more stable and lower cost. Taking advantage of the advantages of the two algorithms, a new fusion location estimation algorithm is proposed. In the offline phase, WiFi fingerprint and image are collected, fingerprint database is established, and the collected images are split and matched using AlexNet to determine the overall range. In the online stage, fingerprint location matching is carried out in the area through the fusion WiFi.After experiments, the results show that our method can effectively reduce the fluctuation of WiFi fingerprint positioning and improve the positioning accuracy; it can be widely used to provide indoor positioning services for people.**

***Keywords*-Indoor position; WiFi fingerprint positioning; multi-mode fusion; features match;**

I. Introduction

Indoor positioning plays an important role in indoor business activities, monitoring, logistics and transportation. With the emergence of Location-Based Service (LBS), indoor positioning is also more used for personal purposes (such as indoor navigation, indoor location-based service, supermarket shopping guide, etc.). However, compared with outdoor positioning based on global positioning system (GPS), which is mature, there are several major problems in indoor positioning: 1) environmental closeness and structural duplication. The impurity makes the accuracy of outdoor positioning unable to meet the practical application. 2) Indoor positioning with additional hardware is expensive and difficult to popularize. 3) Radio signal has serious interference, large attenuation and low positioning accuracy. Mobile phone indoor positioning has been a hot spot in personal indoor positioning because of its low threshold, low cost and easy promotion. However, due to the limitation of mobile phone hardware and computing platform, mobile phone indoor positioning has many limitations.

In the context of the widespread use of WiFi routers and smartphones, WiFi fingerprint-based positioning methods have received widespread attention. The RADAR system [1]introduced by Microsoft Research is the first example of WiFi indoor positioning. The RADAR method compares the real-time collected Wi-Fi Received Signal Strength (RSSI) vector with the data in the database to find the most similar position as the positioning result; Horus[2]is a classic Naive Bayesian probability model. WiFi fingerprint location system. The probability distribution of the Wi-Fi signal strength is used to find the degree of matching between the RSSI vectors in a probabilistic manner. Singh et al. [3]proposes an indoor positioning method that combines WiFi and GPS signals. However, since WiFi signals are affected by multipath interference, it is difficult to achieve satisfactory results by relying solely on WiFi fingerprint positioning. Using images for indoor positioning is another hot research method. Jiao et al [4]uses the analysis of image information and WiFi signals to visualize the WiFi signal and achieve position estimation by merging the visual image and the RGB image.

In view of the shortcomings of the above single indoor positioning method, this paper attempts to use the multi-modal fusion for precise positioning, and combines the image with the WiFi information for positioning, aiming at overcoming the defects of single-mode signal positioning and achieving better positioning effect. Both signals have the following advantages: no additional infrastructure is required, the hardware cost is low; the signal is relatively stable, the robustness is good; the acquisition is convenient and easy to promote. Based on the characteristics of the two signals, the image has better discrimination in a wide range, and the WiFi fingerprint algorithm can distinguish well in the local area. In the positioning process, the image information is first used for regional differentiation, and then the WiFi fingerprint matching is used to obtain more accurate positioning results. In such a multi-modal method, positioning from coarse to fine can have better adaptability and positioning effect.

The rest of this paper is organized as followed. Section 2 reviews related work and the algorithm framework in Section 3. Section 4 discusses the methodology. Application and analysis are discussed in Section 5, and finally Conclusion

978-1-5386-7732-2/19/$31.00 ©2019 IEEE

and Future works will be featured in Section 6.

II. Related Work

The indoor positioning system uses sensor information to realize position estimation. Our work is to realize multi-modal position fusion position estimation by using WiFi fingerprint location and image position classification.

A. WiFi fingerprint positioning

WiFi has two working modes, one is a peer-to-peer working mode, wireless interconnection through a wireless network card; the other is to install a wireless connection device, builds an infrastructure network model, and forms a unique WiFi hotspot, which is used by a large number of users. Such as large shopping malls, airports, etc.Different wireless receiving points (APs) broadcast radio signals to the surrounding at a specific frequency. The data frames contained in the signals carry some specific information to indicate their network status, especially the received signal strength (RSSI) and MAC address, WiFi fingerprint. Positioning obtains the target position by obtaining the corresponding position information by real-time matching of the received signal strength, and the WiFi positioning is divided into two stages, an offline training stage and an online positioning stage. The offline stage constructs location coordinates and WiFi signal strength, and locates the fingerprint database. The online stage compares the measured WiFi fingerprint information with the stored data in the database. Commonly used WiFi measurement methods are TOA, TDOA, and AOA triangulation methods. This paper proposes a weighted fusion fingerprint location method.

B. Image-based positioning

Image-based localization uses pattern matching to directly or indirectly match the image of the current environment with the pre-prepared image. The most direct way of matching is to extract features from real-time images, and find out the most similar database images by feature similarity, using the location of the image identification as the location. For example, some researchers use some obvious landmarks as markers to achieve localization by image recognition of landmarks. The use of landmark positioning requires the ease of discovery and uniqueness of landmarks, which is more suitable for outdoor open environment. In the indoor environment, because of its space limitations, the feasibility of using landmarks alone for positioning is low.

In addition, many studies have focused on the use of image fingerprint for matching and positioning. The feature extraction methods focus on scale-Invariant Feature Transform (SIFT), Speed-Up Robust Feature (SURF) and other features of the captured image, and establish image fingerprint database for matching. The position corresponding to the image with the highest matching degree will be used as the users adjacent location. There are also image semantics [5] information used as features, using semantic features for matching. Because image location only needs image as input, it can also be used for outdoor location. Image localization requires a more distinct distinction between locations, and it is closely related to the quality of the image.

The main methods of using computer vision are 2D image retrieval and 3D modeling. The 2D approach typically requires image feature matching to determine where the user is taking the photo. A typical representative is MoVIPS [6], which uses the SURF feature to estimate the relative distance of the real geographic location using the pixel distance, and finally the specific location where the user took the photo. The 3D model-based positioning method usually uses the SfM (Structure from Motion) algorithm to construct a three-dimensional structure of an indoor scene using a two-dimensional image sequence. For example, the iMoon method proposed by Jiang et al. [7] compiles a navigation grid from the generated 3D model by building a 3D model of the indoor environment through crowdsourced 2D images. Find the mapping between the 2D feature points and the 3D point cloud on the photo to get the positioning result. The positioning method based on computer vision is relatively intuitive, with good stability and convenient data collection. However, the method based on 3D model localization is complicated, and the amount of data construction is large. The method based on 2D retrieval is relatively inaccurate, and it is necessary to further combine the other methods to optimize the positioning effect.

C. Multi-mode Fusion Indoor positioning

By utilizing the advantages of both, the multi-mode data can be effectively utilized in a multi-modal manner to obtain a better positioning effect. This paper is mainly based on the indoor positioning method based on the fusion of WiFi information and image data information. The convenience of WiFi and the stability of the image are used to improve the positioning performance. Firstly, the image data is trained in the offline stage to divide the location area, the WiFi fingerprint information is collected, and the image information is divided; the WiFi data is matched in real time in the online stage, and then get the positioning result.

III. Algorithm Framework

An algorithm for image localization using multi-modal fusion using image and WiFi fingerprint information is proposed to overcome the shortcomings of single-mode signal localization. Based on the characteristics of image and WiFi fingerprint location, in the process of positioning, the image information is first used for area division. Based on this, the WiFi fingerprint information is used for further fine-grained matching. Positioning is done in a coarse to fine manner.

The designed indoor positioning algorithm is mainly divided into two stages: the offline training stage and the online positioning stage. The positioning process is as shown:

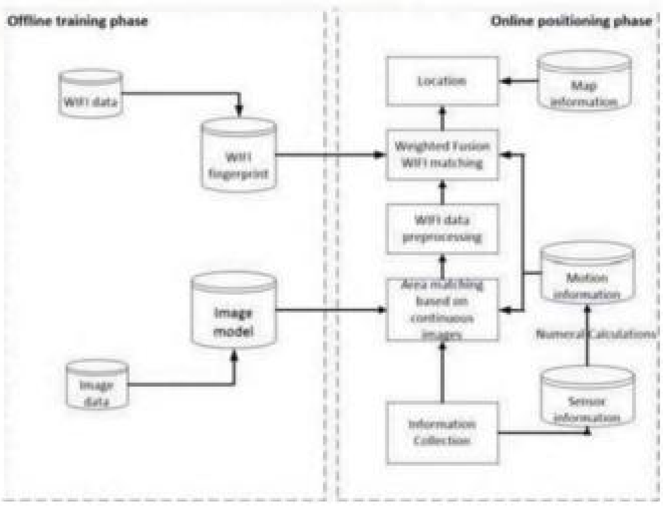

Figure 1. The system work flow of our positioning process

In the offline training stage, it is necessary to combine the test scenario plan with the preset data acquisition related information to collect and analyze the WiFi signal strength and image data. This phase mainly includes two parts: (1) WiFi signal strength processing: using the K-means algorithm to perform fingerprint clustering training on signal strength to construct a corresponding WiFi fingerprint map; (2) Image training: Using AlexNet to construct an image classification model.

In the online positioning stage, after the relevant data acquisition is completed, the region matching based on the continuous image: for each segment, the trained neural network model is used for image classification, and successive images are analyzed to reduce the chance and determine the overall region range. Weight-based fused WiFi signal strength matching: The location information is filtered using the region range determined by the image and the signal strength obtained online. To determine the closest location.

A. *Construction of indoor coordinates and grid point division*

This paper chooses to build indoor maps in Guilin Intelligent Industrial Park, builds indoor maps using indoor actual scenes, establish a two-dimensional coordinate system, and maps each position of the indoor plane to coordinates.

The indoor map is meshed evenly, and Wi-Fi fingerprint data and multiple image data are collected at multiple locations and recorded in two-dimensional coordinate positions. In the case of Wi-Fi fingerprint collection, due to the obvious fluctuation of some low-intensity Wi-Fi APs, this paper adopts the strategy of time-division multiple acquisition. For image data, by acquiring continuous video, subsequent frame extraction is performed to acquire a related position image.

The grid points are divided by a fixed length to collect data for the walkable range of the indoor environment. In this paper, the sampling points are divided by 0.5 m steps. Figure 2(c) shows the point distribution information for indoor sampling in the indoor reachable range, and the origin of which is set at the left lower corner of the scene.

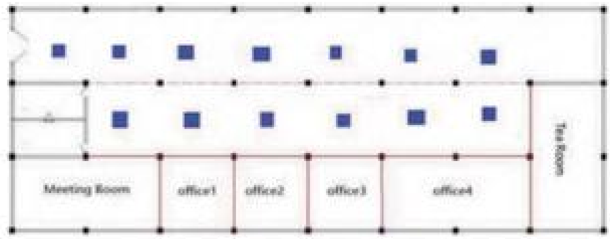

Figure 2. Indoor coordinates

IV. Methodology

First, the area to be located is divided, and then the corresponding collection point is determined according to the fixed AP point. According to the indoor environment and accuracy requirements, the distance between each fixed point can be set between 1 meter and 5 meters. The higher the precision, the smaller the fingerprint distance of the dot. The information such as the BSSID of the AP, the physical location of the sample reference point, the received signal strength value, and the timestamp of the RSSI are acquired. The original data is filtered to eliminate the useless AP information, which can improve the computational efficiency of the algorithm, and then average the original data to reduce the interference of other signals, so that the fingerprint information is more stable. The pre-processed data is trained by fingerprint clustering using K-means to further classify, reduce online matching fingerprint search time, and improve corresponding efficiency, so that it can adapt to relatively large scenes. After some columns data processing, the fingerprint database is finally generated and stored in the server.

In the offline fingerprint data collection stage, data is collected according to the reference point, and R = $[r_1, r_2, \ldots, r_m]$ be a two-dimensional matrix for reference points.At each reference point, nth RSSIs are received. In order to reduce the measurement uncertainty, the signal strength is obtained from R_0 to R_t in t time, and the received signal strength is averaged, and computed as:

$$\bar{R}_t = \frac{1}{t}\left(\sum_{n=0}^{n\le t} R_t\right). \tag{1}$$

In the online stage, the WiFi signal strength is obtained in real time, and the Euclidean distance [8] is used to calculate the real-time signal strength and the signal strength distance in the fingerprint database. The Euclidean distance is:

$$D_j = \sqrt{\sum_{n=1}^{i}(r_n - \bar{R}_t)}. \quad (2)$$

Taking kth reference points with the smallest Euclidean distance and then estimates the user position with the corresponding position:

$$(x, y) = \frac{\sum_{j=1}^{k}[(x_i, y_i)]}{k}. \quad (3)$$

A. Image classification and positioning based on AlexNet

When collecting image data, the collector only needs to collect continuous video according to the direction of travel, without taking a separate point to take a photo. This way can be synchronized with the acquisition of the magnetic field sequence, and the acquisition speed is faster. After the video acquisition is completed, frame extraction is performed on the collected video. The images of the corresponding frames are extracted according to a certain number of frames, and are mapped to specific path coordinates according to time intervals. Since the similar geographical location images have local similarity, each picture can be classified according to its actual physical location, that is, the images of a certain length interval are classified into the same class.

After completing the classification, use AlexNet [9] to perform supervised learning classification modeling on the image for regional scoping in the online positioning stage. Compared with some of the later models, such as GoogLeNet [10], AlexNet has a relatively shallow network structure, but it can also extract more features. For the small amount of data in this article, using the AlexNet model will have better accuracy and efficiency.

As can be seen from the basic structure diagram, AlexNet consists of eight layers of structure, including five convolutional layers and three fully connected layers. Each convolutional layer contains an excitation function (RELU) and a local response normalization processing layer (LRN), which is processed by downsampling (also called pooling layer, which is not included in conv3 and conv4) to reduce the amount of computation. And data dimensions. Among them, the LRN (Local Response Normalization) layer is a special calculation layer in AlexNet, which performs local area normalization on the input data, and the function is to smooth the output result of the current layer, thereby improving the precision. The calculation formula is as shown:

$$b_{x,y}^{i} = a_{x,y}^{i} / (k + a \sum_{min(N-1,i+n/2)}^{j=max(0,i-n/2)} (a_{x,y}^{j})^2)^{\beta} \quad (4)$$

In practical applications, the first input layer is 227 × 227 × 3, instead of 224 × 224 × 3 in the original network, the main reason is that the input of 227 × 227 is more conducive to the divisible calculation of the convolutional layer. In the last layer of the original model, 1000 neurons are used, and 4096 neurons in the seventh layer are fully connected. Through the Gaussian filter, 1000 values of type float are obtained, that is, 1000 prediction probabilities are obtained. In actual use, the appropriate number of neurons is selected according to the number of classifications. After the image data is classified, the Caffe Model can be directly called in the online positioning stage, and the newly collected image data is classified, so that the area is divided according to the information, and it's better to be used to the WiFi positioning matching.

V. Application and analysis

This experiment mainly tests the positioning algorithm in Guilin Intelligent Industrial Park. It uses OnePlus 5T mobile phone as the test equipment. In the offline stage, it collects WiFi and image data, and performs WiFi fingerprint and multi-angle image acquisition at fixed points. We selected 200 sets of test data and compared the cumulative error distribution functions of the test data, as shown in the figure. The abscissa is the positioning error and the ordinate is the corresponding cumulative probability. The red solid line is the multi-modal indoor positioning algorithm used in this paper, that is, the positioning error curve combined with the magnetic field and the image; the green thick dotted line is the positioning error curve of the WiFi positioning mode, and the positioning is performed using the WiFi fingerprint information in the test data; The thin color dashed line is the image mode positioning error curve, and the image data is combined with feature extraction for positioning.

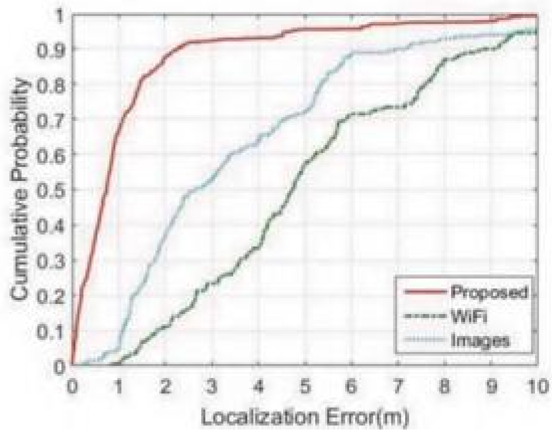

Figure 3. Guilin Intelligent Industrial Park test cumulative probability error function

This scene is mainly based on the corridor area. Some of the test results based on the image information can also obtain better positioning results, but the ratio of errors greater than 2 meters is significantly higher than that of the multi-modal method. The positioning error based on the WiFi fingerprint method is less than 2 meters. In general, compared with the other two methods, the multi-modal positioning method fully combined the advantages of the signal, and helps to narrowing the range and obtain a small positioning error.

VI. Conclusion and Future works

Indoor positioning has always been important for indoor business activities, monitoring, logistics, etc.Many scholars have studied Wi-Fi positioning, but due to the instability of Wi-Fi, Wi-Fi positioning focuses on indoor room and area positioning, and it is difficult to coordinate positioning in more common situations. There are also many studies that use WiFi and images for point location, but their image localization is mostly based on preset image tags, using the image projection in the camera and its prestored spatial geometry for point location, but it is necessary to place image tags that are clearly identifiable in a plurality of locations, and it is limited by the viewing angle of the viewing angle and is easy to cause visual interference to the user. The fusion strategy proposed in this paper is to divide the existing area by image, and then use WiFi to perform real-time positioning through weighted fusion positioning algorithm, which realizes accurate point location in the case of small number of WiFi APs. The positioning algorithm has considerable practical application value for industrial production, but it still needs improvement in several aspects: (1) For realtime performance, it is often used as one of the indicators for evaluating positioning performance, and the AlexNet image training is used to obtain the classification area, which increases the time overhead. (2) When the user makes a fast move, the image matching accuracy is low for image shake. However, in the case of low real-time requirements and smooth movement, the algorithm proposed in this paper has a higher practical significance.

Acknowledgements

This project was sponsored in part by the National Natural Science Foundation of China (Nos.61772149 and 61762028),the Guangxi Key Research and Development Program (No.AB17195057), the Guangxi Innovation Drive Project (No.AA18118039), and Guangxi Natural Science Foundation (No.2018GXNSFAA138084).

References

[1] P. Bahl, V. N. Padmanabhan, V. Bahl, and V. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," 2000.

[2] R. Van Renesse, K. P. Birman, B. B. Glade, K. Guo, M. Hayden, T. Hickey, D. Malki, A. Vaysburd, and W. Vogels, "Horus: A flexible group communications system," Cornell University, Tech. Rep., 1995.

[3] R. Singh, M. Guainazzo, and C. S. Regazzoni, "Location determination using wlan in conjunction with gps network (global positioning system)," in *2004 IEEE 59th Vehicular Technology Conference. VTC 2004-Spring (IEEE Cat. No. 04CH37514)*, vol. 5. IEEE, 2004, pp. 2695–2699.

[4] J. Jiao, F. Li, W. Tang, Z. Deng, and J. Cao, "A hybrid fusion of wireless signals and rgb image for indoor positioning," *International Journal of Distributed Sensor Networks*, vol. 14, no. 2, p. 1550147718757664, 2018.

[5] C. Colombo, A. Del Bimbo, and P. Pala, "Semantics in visual information retrieval," *Ieee Multimedia*, vol. 6, no. 3, pp. 38–53, 1999.

[6] M. Werner, M. Kessel, and C. Marouane, "Indoor positioning using smartphone camera," in *2011 International Conference on Indoor Positioning and Indoor Navigation*. IEEE, 2011, pp. 1–6.

[7] J. Dong, Y. Xiao, M. Noreikis, Z. Ou, and A. Ylä-Jääski, "imoon: Using smartphones for image-based indoor navigation," in *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2015, pp. 85–97.

[8] N. Ravi, P. Shankar, A. Frankel, A. Elgammal, and L. Iftode, "Indoor localization using camera phones," in *Seventh IEEE Workshop on Mobile Computing Systems & Applications (WMCSA'06 Supplement)*. IEEE, 2005, pp. 1–7.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.