

# A Hybrid of Smartphone Camera and Basestation Wide-area Indoor Positioning Method

Jichao Jiao<sup>1</sup>, Zhongliang Deng<sup>1</sup>, Lianming Xu<sup>2</sup> and Fei Li<sup>1</sup>

<sup>1</sup>Laboratory of Intelligent Communication, Navigation and Micro/Nano-Systems  
Beijing University of Posts and Telecommunications  
Beijing, 100876 – China

[e-mail: jiaojichao@bupt.edu.cn]

<sup>2</sup>Beijing RTMAP Technology Co. Ltd  
Beijing, 100876 - China

[e-mail: xulianming@rtmap.com]

\*Corresponding author: Jichao Jiao

*Received September 30, 2015; revised November 19, 2015; accepted December 9, 2015;  
published February 29, 2016*

---

## Abstract

Indoor positioning is considered an enabler for a variety of applications, the demand for an indoor positioning service has also been accelerated. That is because that people spend most of their time indoor environment. Meanwhile, the smartphone integrated powerful camera is an efficient platform for navigation and positioning. However, for high accuracy indoor positioning by using a smartphone, there are two constraints that includes: (1) limited computational and memory resources of smartphone; (2) users' moving in large buildings. To address those issues, this paper uses the TC-OFDM for calculating the coarse positioning information includes horizontal and altitude information for assisting smartphone camera-based positioning. Moreover, a unified representation model of image features under variety of scenarios whose name is FAST-SURF is established for computing the fine location. Finally, an optimization marginalized particle filter is proposed for fusing the positioning information from TC-OFDM and images. The experimental result shows that the wide location detection accuracy is 0.823 m ( $1\sigma$ ) at horizontal and 0.5 m at vertical. Comparing to the WiFi-based and ibeacon-based positioning methods, our method is powerful while being easy to be deployed and optimized.

---

**Keywords:** Indoor positioning, smartphone camera, fusion of telecommunication and navigation, image feature descriptor, marginalized particle filter

## 1. Introduction

With the rapid growth in the use of handheld devices such as smartphones and tablets, location-based services (LBSs) have become increasingly popular. The demand for an indoor positioning service has also accelerated. That is because people spend most of their time indoor environment [1]. Over the last decade, researchers have studied many indoor positioning techniques [2]. Indoor positioning systems based on wireless local area networks are growing rapidly in importance and gaining commercial interest. Moreover, with the development of the integrated circuit technology, multi-sensors, for example, camera, Earth's magnetic field, WiFi, Bluetooth, inertial module, have been integrated in smartphones. Therefore, smartphones are powerful platforms for location-awareness. At the same time, smartphone positioning is an enabling technology that is used to create a new business in the navigation and mobile location-based services industries.

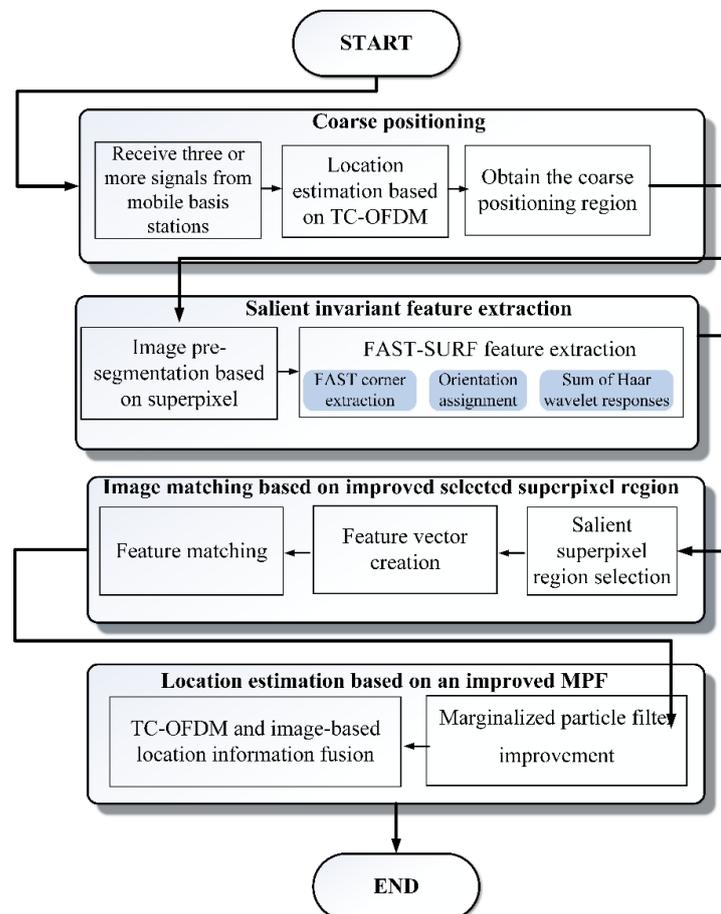
Recently, lots of researchers have been focused on the indoor positioning studies. According to the localization performance, indoor positioning is classified into local positioning and wide positioning [3], [4] and [5]. Local positioning methods are implemented based on wireless local area network. WiFi-based indoor positioning [6] have been more widely used in buildings than other local methods like RFID [7], Ultra Wideband (UWB) [8], Zigbee [9], Bluetooth [10] and Pseudolite [11]. However, the local methods gain a poor accuracy that is more than 5 meters except UWB-based and Pseudolite-based methods that are very expensive to be implemented. Furthermore, almost of the local methods can only work inside buildings, which make users depend on the GNSS in outdoor positioning. Therefore, the popularization of indoor LBS application is seriously limited.

Wide-area positioning methods are based on cellular network base stations that are widely distributed in the world, which have wide indoor signal coverage. 2G/3G/4G mobile communication systems can provide indoor location. However, the positioning accuracy is too poor to meet most of the requirements due to the interference of Non Line of Sight (NLOS) [12], multipath and the poor time synchronization among Base Stations (BSs) [13]. The Time & Code Division-Orthogonal Frequency Division Multiplexing (TC-OFDM) system, which is a typical navigation and communication integrated system, can achieve 1~3 meters in horizon and 0.5 meter in vertical. However, in urban canyon environment or inside the large buildings, the TC-OFDM signal cannot cover those areas [14]. Moreover, the indoor awareness information cannot be obtained by using WLAN-based positioning system [15]. Therefore, Wang and her colleges presented a wireless sensor network based indoor positioning systems for context-aware applications [16]. In [17], a maximum likelihood-based fusion algorithm that integrates a typical Wi-Fi indoor positioning system with a PDR system was proposed for indoor pedestrian navigation. However, these methods that using existing wireless networks have low deployment costs, but the positioning error can be up to several meters because of NLOS, multipath and signal attenuation. Therefore, smartphone camera-based indoor positioning is a promising approach for accurate indoor positioning without the need for expensive infrastructure like access points or beacons.

More recently, indoor positioning based on images has been popular [18][19][20]. Meanwhile, all of those research works mainly focus on improving image matching accuracy. Some of these algorithms are, however, quite demanding in terms of their computational complexity and therefore not suitable to run on mobile devices, which need smartphones with high hardware configuration. Although smartphones are inexpensive, they have even more limited performance than the aforementioned Tablet PCs. Smartphones are embedded

systems with severe limitations in both the computational facilities and memory bandwidth. Therefore, natural feature extraction and matching on phones has largely been considered prohibitive. To address these issues, we proposed a new image feature detector named FAST-SURF.

This paper proposed a hybrid algorithm combining vision-based positioning approach with BSs-assisted approach that was TC-OFDM for wide-area indoor, and this algorithm was named TC-Image. First of all, the coarse location is calculated by using TC-OFDM system. Then, the salient invariant features are extracted from the images taken by the smartphone camera, and feature vectors are built. Third, the fine positioning information can be obtained by using an improved matching method. Finally, an improved marginalized particle filter (MPF) is used to fusion the positioning results from TC-OFDM and images. The proposed algorithm is implemented on Android operation system, and the experiment demonstrates that the performance improvement can be achieved. **Fig. 1** shows the procedure of our algorithm.



**Fig. 1.** Framework of the system

## 2. Related Work

Valgren and his colleague proposed a camera-based outdoor positioning by using SURF feature for speeding up the image matching [21]. Li and Wang [22] introduced A-SIFT

feature for image matching achieved by RANSAC, which increased the matching accuracy. Tian and his co-workers [23] proposed a similar method to [22] for indoor positioning. However, those two complex computational methods are not suitable for smartphone based indoor positioning. This is because of limited computational resources of mobile devices.

### 3. The Proposed Algorithm

How to take advantage of heterogeneous wireless networks (WiFi, Zigbee, RFID, Bluetooth, Cellular Network, etc.) to realize the seamless outdoor and indoor positioning in wide area based on smartphones has become the hot issues of LBS. Since both kinds of systems have their own advantages and disadvantages, a novel particle filter-based fusion algorithm is proposed in this paper. In this method, it is to integrate the TC-OFDM indoor positioning system with images that taken by using cameras of the smartphones.

#### 3.1 Coarse positioning based on TC-OFDM

TC-OFDM is a wide-area indoor and outdoor seamless positioning system based on mobile base stations. TC-OFDM system is a typical navigation and communication integrated system. The cellular network carries the TC-OFDM signals who multiplex communication and navigation signal in same frequency band, which is shown in Fig. 2.

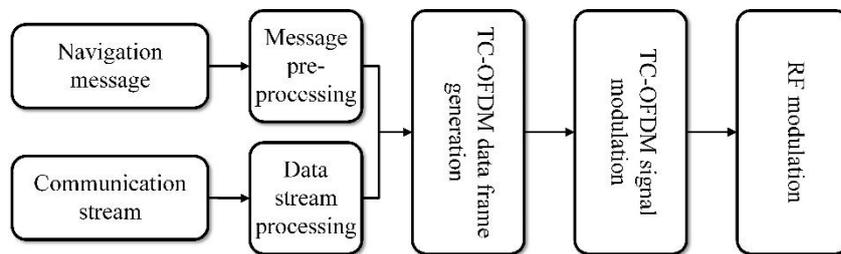


Fig. 2. Flowchart of TC-OFDM signal generation

The wide-area indoor signal coverage is achieved by mobile BSs with high-precise time synchronization. Terminals demodulate the TC-OFDM signals and obtain the navigation message for positioning. Besides, it can also assist GNSS to enhance the performance and robustness for outdoor positioning accuracy and Time To First Fix (TTFF). TC-OFDM system can provide seamless indoor and outdoor positioning in wide area with one-meter accuracy. Positioning information can be sent to the location server system on the network for location management to provide location information to the third party LBS. However, in small number of large scale constructions, signal coverage of BSs may not be well, and indoor supplement system of TC-OFDM will be used for signal coverage.

In order to receive and process the TC-OFDM positioning signal, an accessory named WINP (Wide Indoor Navigation and Positioning) have been designed and it can be connected with smartphone by using Bluetooth. Then, the initial location information is computed based on TDOA (Time Difference of Arrival) when the WINP receives signals from three more base stations, which is shown by Fig. 8.

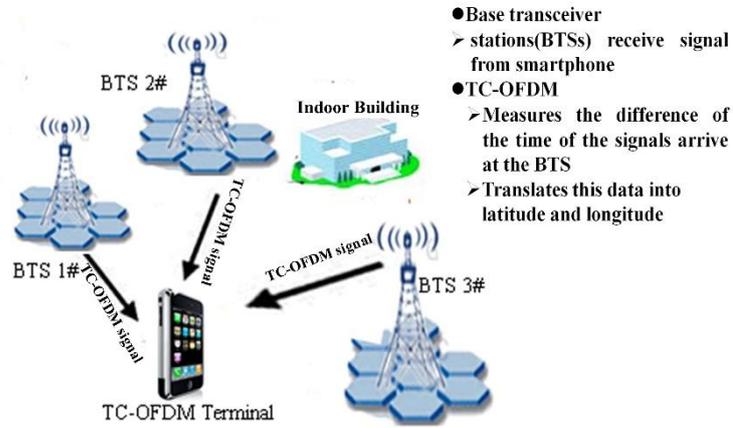


Fig. 3. The TDOA used in TC-OFDM system

In our experiment, the accuracy of time synchronization between BTS achieved to 5 ns, which a key factor for obtaining a precision positioning result.

**3.2 Salient invariant feature extraction**

Based on the initial location, the regional reference images are fast located in the database, which increases the positioning speed. Then, in order to make our algorithm robust to the indoor illumination and complex indoor background, we extract salient features including Speeded Up Robust Features (SURF) [24], and features from accelerated segment test (FAST) corners [25] from the images taken by the smartphone camera and reference database. Specially, the features are combined into one feature vector for building a robust representation model.

1) FAST-SURF invariant feature detection

SURF is a novel scale- and rotation-invariant interest point detector and descriptor [17]. It approximates or even outperforms previously proposed schemes with respect to repeatability, distinctiveness, and robustness, yet can be computed and compared much faster. This is achieved by relying on integral images for image convolutions, and a Hessian matrix-based measure is used for the corner detector. Hessian detector is shown as follows:

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \tag{1}$$

where  $\sigma$  is used to obtain the scale-space representation that is a set of images represented at different levels of resolutions. Different levels of resolution are in general created by convolution with the Gaussian kernel:  $L(x, \sigma) = G(\sigma) * I$ .  $G(\sigma)$  is the Gaussian kernel.

$L_{xx}(x, \sigma)$  is the convolution of the Gaussian second order derivative  $\frac{\partial^2}{\partial x^2} g(\sigma)$  with the image  $I$  in point  $x$ , and similarity for  $L_{xy}(x, \sigma)$  and  $L_{yy}(x, \sigma)$ .

In order to improve the performance of the proposed algorithm in positioning time, FAST corner detector is used to search the salient corner because of its good performance in computation time, scale-invariant and accuracy. The formula of FAST corner detector is shown as follows:

$$S_n = \begin{cases} d, & I_n \leq I_{cp} - t (\text{darker}) \\ s, & I_{cp} - t \leq I_n \leq I_{cp} + t (\text{similar}) \\ l, & I_n \geq I_{cp} + t (\text{lighter}) \end{cases} \quad (2)$$

where  $I_{cp}$  is the gray value of the candidate FAST corner,  $I_n$  is the gray value of the contiguous pixels in the circle whose center is the candidate FAST corner.  $t$  is a threshold for measuring the contrast of  $I_{cp}$  and  $I_n$ .  $d$ ,  $s$  and  $l$  are used to show the contrast status of a pixel in the image. This detector exhibits high performance, but the threshold  $t$  is a fixed value, and  $t \neq 1$ . Therefore, in order to make the corner detector more robust, this paper proposed an approach for calculating the threshold as following:

$$t = a\mu + \eta \quad (3)$$

where  $\mu$  is the image mean,  $\eta$  is the image variance,

$$a = \frac{\sum_{i=0}^{M \times N} [r(x_i, y_i)] \cdot \sum_{i=0}^{M \times N} f(x_i, y_i)}{\sum_{i=0}^{M \times N} f(x_i, y_i) \sum_{i=0}^{M \times N} [r(x_i, y_i)]} \cdot f(x_i, y_i) \text{ is an image taken}$$

by the smartphone camera, and  $r(x_i, y_i)$  is a reference image from the database.

In order to generate the FAST-SURF descriptor, the first step consists of constructing a square region centered on the FAST point, and oriented along the orientation selected in the previous section, which is the same as SURF [23]. Considering finer subdivisions appeared to be less robust and would increase matching times too much, in our proposed method, the dimensions of FAST-SURF is 64, which characterize the local appearance of an object.

### 3.3 Salient invariant feature matching based on improved selected superpixel region

#### 1) Selected region detection

After extracting FAST-SURF features, a speed-up feature matching approach based on the pre-segmentation is proposed. An improved method based on superpixels that is refined simple superpixel is introduced for the high-resolution image pre-segmentation, and then the matching feature space are selected from the superpixel regions, which reduces the running time in searching feature space.

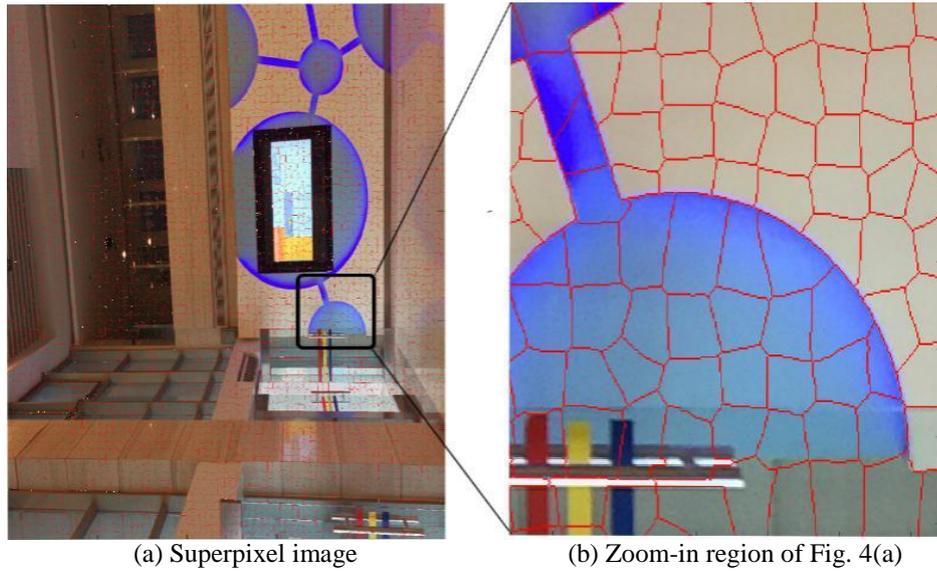
A superpixel is a patch whose boundary matches the edge of an object. The aim of superpixels is to reduce the speed-up problem by replacing pixels with regularly-spaced, similarly-sized image patches whose boundaries lie on edges between objects in the image, so a pixel-exact object segmentation can be accomplished by classifying superpixel patches rather than individual pixels. **Fig. 4** shows the pre-segmentation result.

According to our research, the salient invariant features are mainly in the complex superpixels. Therefore, entropy that measures the image complexity is introduced to choose the superpixel regions that include salient features. The following formula is used to detect the selected superpixel region:

$$\Phi(S_i) = \alpha \times \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} + \beta \times \left( \sum_{i=1}^N p(x_i) \ln p(x_i) \right) \quad (4)$$

where  $\Phi$  is the candidate salient region score,  $S_i$  is the  $i$ th superpixel,  $x_i$  is the pixel gray value in the  $i$ th superpixel,  $\mu$  is the mean of superpixel region,  $\alpha$  and  $\beta$  are the weights, and  $\alpha + \beta = 1$ .

Then, we detect the salient regions by using a threshold  $thr$ . If  $\Phi_i > thr$ , this superpixel is the selected region for feature matching.



**Fig. 4.** The pre-segmentation image by using superpixels.

### 2) FAST-SURF feature matching

After detecting the salient superpixel, FAST-SURF feature will be matched. The matching is performed by computing the Mahalanobis distance between the FAST-SURF features by using the following function:

$$M_d = \sqrt{(\mathbf{FS}_{cam} - \mathbf{FS}_{sen})^T \Lambda (\mathbf{FS}_{cam} - \mathbf{FS}_{sen})} \quad (5)$$

where  $FS_{cam}$  and  $FS_{sen}$  are the FAST-SURF features extracted from smartphone images and database, respectively.  $\Lambda$  is the FAST-SURF feature covariance matrix.

### 3) Vision-based Position and Orientation Determination

Given a set of correspondences between known 3D reference points and their 2D positions in images, camera position and orientation can be determined [25]. Then any object appeared in the view can also be estimated in 6 degrees of freedom (DOF).

The fundamental function model for photogrammetric 6 DOF pose estimation (space resection) is called collinearity equations, which represent the geometry between projection centers, the world coordinates of an object and its image coordinates. For the object-centered projection, the image position for the 3D point  $(X, Y, Z)$  in the world coordinate frame is given by the projective transformation,

$$\begin{pmatrix} x \\ y \\ f \end{pmatrix} = \frac{f}{z} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \\ f \end{pmatrix} \quad (6)$$

where  $(x, y)$  is the 2D point in the image,  $(x_0, y_0)$  is a 2D point in the camera interior parameters,  $f$  is the camera's focal length. In order to build the relationship between  $(x, y)$  and its corresponding 3D point  $(X_s, Y_s, Z_s)$  in the camera coordinate frame, the transformation equation between the world coordinate and the camera coordinate is shown as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} X_s \\ Y_s \\ Z_s \end{bmatrix} \quad (7)$$

where  $(a_i, b_i, c_i)$  are the elements of the rotation matrix between the image and the object coordinate system. Then Eq. 1 and Eq. 2 can be written as

$$\begin{aligned} x - x_0 &= -f \frac{a_1(X - X_s) + b_1(Y - Y_s) + c_1(Z - Z_s)}{a_3(X - X_s) + b_3(Y - Y_s) + c_3(Z - Z_s)} \\ y - y_0 &= -f \frac{a_2(X - X_s) + b_2(Y - Y_s) + c_2(Z - Z_s)}{a_3(X - X_s) + b_3(Y - Y_s) + c_3(Z - Z_s)} \end{aligned} \quad (8)$$

### 3.4 Location estimation based on an improved MPF

As mentioned above, TC-OFDM system can provide continuous tracking, and thus, they can be used to overcome the fluctuation of RSS-based Wi-Fi positioning. Moreover, one of the advantages of TC-OFDM system is that it can provide the altitude information for the users. However, in the complex environment where the SNR of TC-OFDM signal is lower than  $-135dBm$ , which makes the positioning accuracy worse. Besides, TC-OFDM system cannot support indoor context information. Conversely, the vision-based system can make up for assisting in increasing positioning accuracy and supporting the around scene information for customers. Therefore, in this section, an MPF-based fusion algorithm is proposed to fuse the TC-OFDM and the vision-based positioning information effectively. We prefer the MPF-based fusion scheme to the particle filters, due to the high computational complexity of particle filters, which need to update and maintain the states of a large number of particles whenever a new location is provided.

Therefore, eight common motion states used during indoor navigation are detected by fusing information gained from built-in sensors of the smartphone. The aim of fusing TC-OFDM and image positioning information can be as trying to estimate the following probability density function (PDF),

$$I(\eta(x_t)) = E_{p(x_t|Y_t)}[\eta(x_t)] = \int \eta(x_t) p(x_t | Y_t) dx_t \quad (9)$$

where  $\eta(x_t)$  is the status function,  $p(X_t | Y_t)$  is the distribution which is shown as follows:

$$p(x_t | Y_t) = \frac{p(y_t | x_t) p(x_t | Y_{t-1})}{p(y_t | Y_{t-1})} \quad (10)$$

$$p(y_t | Y_{t-1}) = \int p(y_t | x_t) p(x_t | Y_{t-1}) \quad (11)$$

where the status variance at a certain time can be described as follows:

$$p(x_t^l, X_t^n | Y_t) = p(x_t^l | X_t^n, Y_t) p(X_t^n | Y_t) \quad (12)$$

where  $p(x_t^l | X_t^n, Y_t)$  could be estimated by Linear Kalman Filter and  $p(X_t^n | Y_t)$  is estimated by a general particle filter.

### 3.5 Map matching

In order to increase the accuracy, knowledge of the building layout is as important as the positioning technology itself. As our system is only designed to provide location and navigation in public areas of a building, a simple map matching technique has been implemented, which forces the calculated position to stay in those areas. If the position drifts outside, it is simply projected back to the closest position in a public area.



Fig. 5. The wide-area outdoor and indoor seamless system for our experiment in BUPT

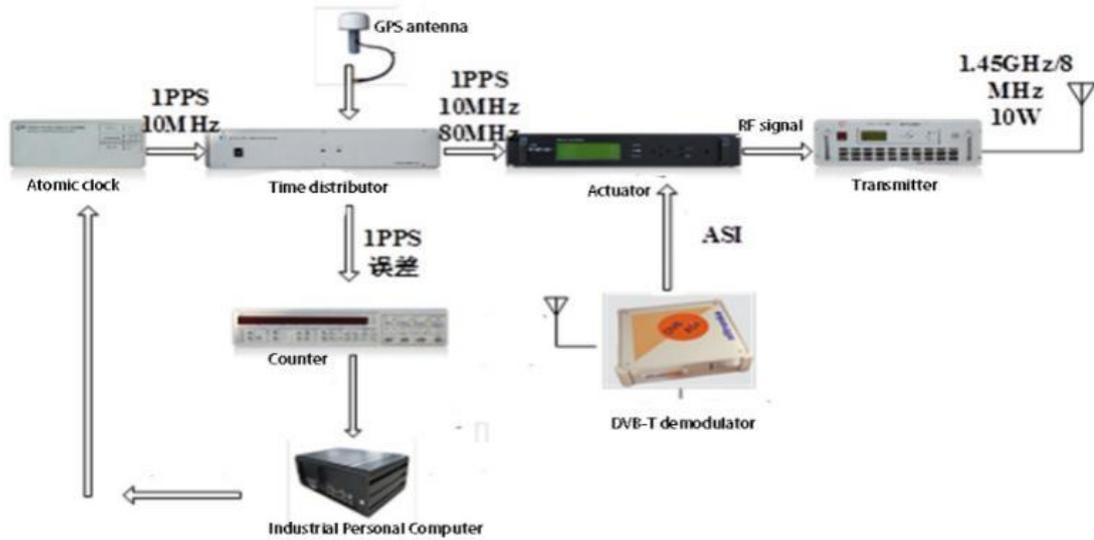


Fig. 6. The equipments of the TC-OFDM positioning base station

## 4. Experiment Results

### 4.1 Study materials and system configuration

We conducted experiments at the New Research Building locating in Beijing University of Posts and Telecommunications, which is shown by Fig. 5. Four base-stations are equipped for covering  $100000m^2$  area including two buildings. It is noticed that the experimental area is  $2000m^2$ . The hardware structure of the positioning base station is shown by Fig. 6. Furthermore, we obtained 600 omnidirectional panoramic reference images and 1237 supplemental images for image matching, and the image resolution is  $3264 \times 2448 pixels$ . It is noticed that all the images are taken by a smartphone camera, and two of them are shown by Fig. 7. Moreover, a static measurement system based on TC-OFDM and Beidou Real Time Kinematic (RTK) is introduced. By using this system, the scalable locations with positioning accuracy (0.6 ~ 1 meter) are obtained. The BUPT dataset covers four buildings and results in total of 1986 positions.



(a) Reference image



(b) Matching image



**Fig. 7** The scenarios of the hall and hallway in BUPT

In our experiment, a smartphone running Android OS 4.4 is used to test the positioning methods used in this paper. The technical details of this phone are shown in the **Table 1**.



**Fig. 8.** The positioning terminal, left one is WINP module

**Table 1.** The key technical parameters

Parameter	Value
CPU	Qualcomm Snapdragon
CPU Processor	4 Core×2.5GHz
GPU	Adreno 330×578MHz
OS	Android 4.4
Memory	RAM: 3GB, ROM: 2GB
Camera	13MP

**Table 2.** The Summary of User-specified parameters

Name	Symbol	Value
Cut off areas	-	2m
Number of fingerprint labels	-	389
The velocity of the object motion	$v$	0.75 m/sec
Sampling period	$\Delta t$	2.0 sec
The initial RSSI value at one-meter	$P_{do}$	3.0 dBm
The breakpoint distance	$R_o$	2 m
Image Resolution	$P$	2048×2048 pixels
Reference point density	-	5 m

In order to evaluate the proposed algorithm, WiFi-based and ibeacon-based positioning system are implemented. In this article, we assumed the measurement noise or wall lose is the normal random variable with zero mean and variance  $N(0; \sigma^2)$ . There are two patterns generated by the DR method, which were used to simulate the path of kinematics model in actuality, circular-path. The algorithm starts with the initial condition are shown in **Table 2**.

#### 4.2 Coarse positioning based on TC-OFDM

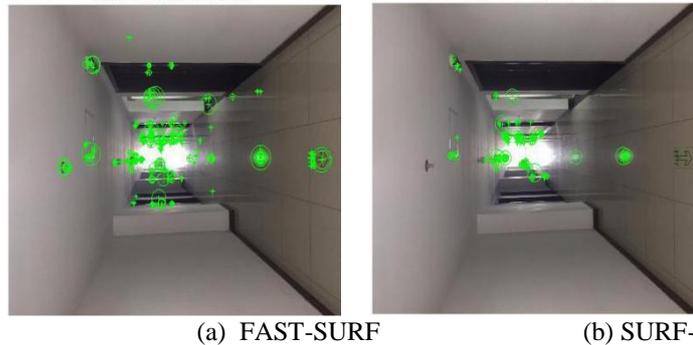
TC-OFDM achieved the positioning information based on the TDOA and the positioning result is shown by **Fig. 9**. The mean of the positioning results of the TC-OFDM system is 2.516 m.

**Fig. 9.** The corset positioning result based on TC-OFDM

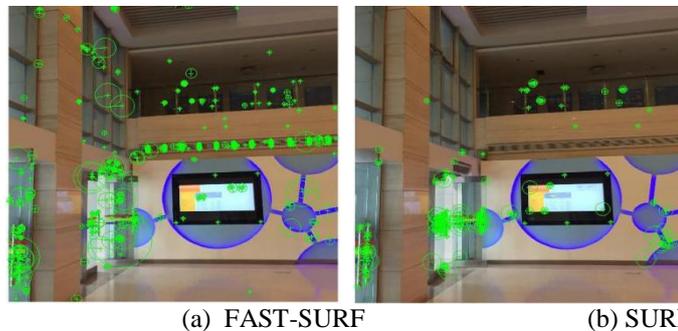
#### 4.3 FAST-SURF feature matching

##### 1) Feature extraction

**Fig.10** is used to show the result of FAST-SURF feature extraction. In this experiment, FAST corners are used to replace the Harris corners in the SURF. In order to evaluate the proposed method, the feature extraction result based on SURF-128 is introduced. The feature extraction results gotten by FAST-SURF and SURF-128 are shown by **Fig. 10** and **Fig. 11**. Moreover, tested on 100 images, the average runtime of FAST-SURF is 73 ms, which is 20 ms faster than SURF-128.



**Fig. 10** The comparison result of feature extraction based on FAST-SURF and SURF-128 for the passage image



**Fig. 11.** The comparison result of feature extraction based on FAST-SURF and SURF-128 for the hall image

According to the **Fig. 10** and **Fig. 11**, we can find that the proposed approach is more robust than SURF-128, which more feature descriptors are extracted by FAST-SURF from the low contrast region.

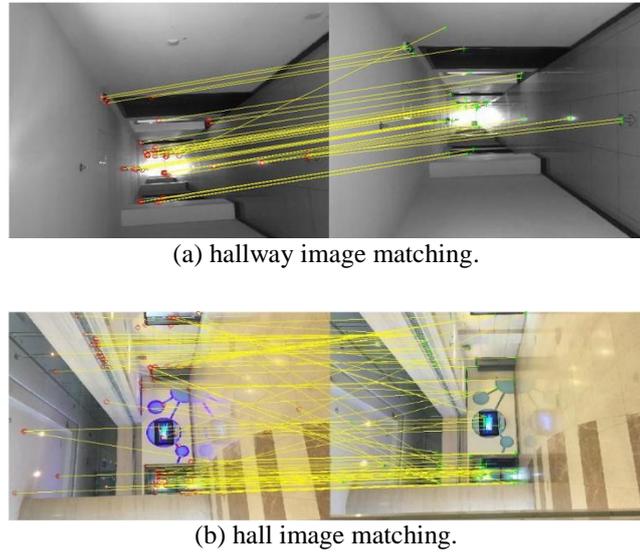
#### 2) Feature matching

According to **Fig. 12**, we can find that SURF-128 gives a large number of matches (perhaps because the number of detected interest points is higher), but more matches are also wrong. The matching results of FAST-SURF and SURF-128 are shown in **Table 3**. With a percentage of only 78% correct matches, it is clearly the worst of the algorithms. Again, FAST-SURF comes out on shared first position with 85% correct matches. However, FAST-SURF finds fewer matches than the other algorithms.

The reason behind is that such feature-based methods depend on the choice of correspondence on local information and fail to consider global context. When an image has repeated patterns, ambiguities will occur when the local information for the similar parts is identical. Moreover, FAST corner detector is used to replace the Harris detector, which make FAST-SURF robust to the illumination.

**Table 3.** Total Number of Matches

Algorithm	Total matches	Total correct matches	Percentage correct	Running time
FAST-SURF	768	653	85%	33.28
SURF-128	479	383	78%	53.74

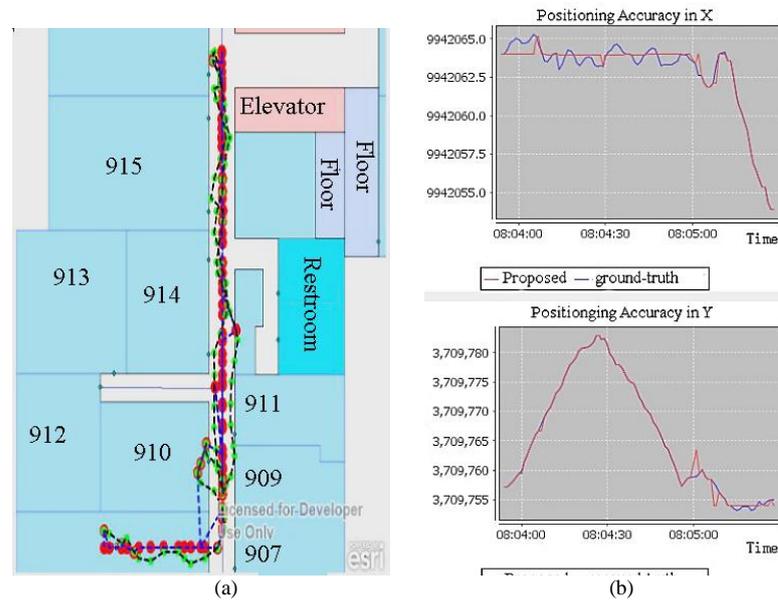


**Fig. 12.** The feature matching results based on FAST-SURF.

#### 4.4 Fine positioning

We have presented a sensor fusion approach to combine TDOA with image-based measurements from a base-station system for 3D location estimation. The approach is experimentally shown to result in accurate position and height estimates when compared to data from an independent optical reference system. To be able to use the TC-OFDM measurements in the sensor fusion approach, the TC-OFDM setup has to be calibrated, i.e. the smartphone with the WINP positions have to be computed. We have solved the WINP calibration problem using a novel approach, taking into account the possibility of delayed TDOA measurements due to NLOS and/or multipath. Furthermore, images taken by smartphone are used to compute the precise positions because of solving feature matching problem under the complex environment. Throughout this work, we have used a marginalized particle filter to model the linear/nonlinear location state. This model is shown to lead to accurate position estimation even from challenging data containing a fairly large amount of outliers in a new multi-lateration approach. In order to fusion the positioning information, Rao-Blackwellization approach is used to compute the model.

In this paper, two-level fusion strategy is introduced. First of all, the positioning result based on TC-OFDM is used to compute the image feature extraction space, which speed up the image-based positioning. In the second fusion level, the positioning results from TC-OFDM and images are fused by using RMPF (Rao-blackwellization Marginalized Particle Filter). The positioning results are shown as follows:



**Fig. 13.** The positioning result based on TC-Image. (a) the ground-true pass line (blue) and estimated pass line (red); (b) the comparison positioning result of the ground-truth (blue) and estimated based on TC-Image (red).

According to **Fig. 13(a)**, we can find that the estimated positioning line practically coincides to the ground-truth, which is also be improved by **Fig. 13(b)**. From the **Fig. 13(b)**, it is noticed that the positioning result in the horizontal direction is worse than that in the vertical direction. It is because the width of the building corridor is 1.5 m, which constrains the positioning error by using map matching. Moreover, the mean positioning of TC-Image is 0.823 m ( $1\sigma$ ), which shows that the proposed method has achieved to the indoor sub-meter positioning.

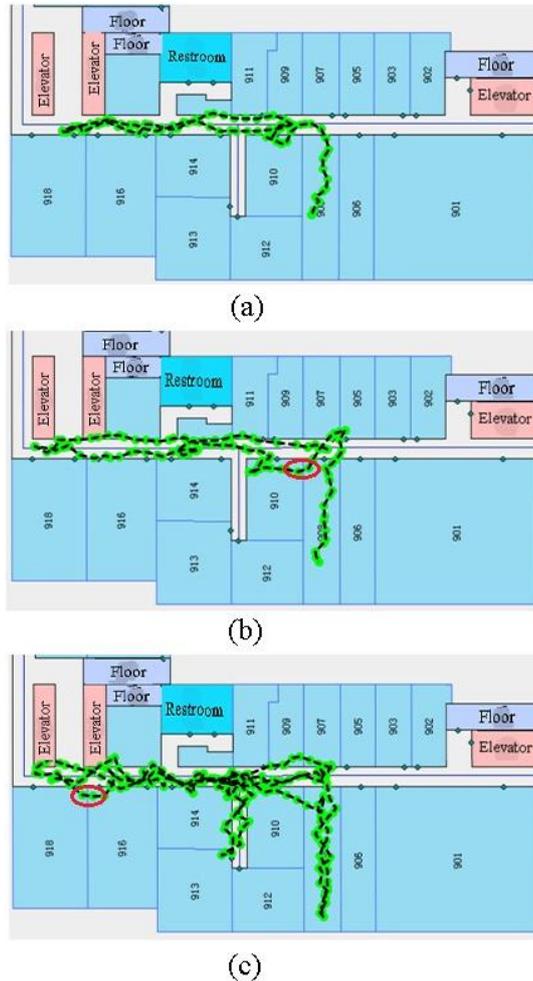
#### 4.5 Evaluation

In this article, in order to check the estimation performance of the proposed method. Two approaches are introduced to compare to the TC-Image on the accuracy.

##### 1) Track Estimation Comparison

First of all, we tested the tree methods at the ninth floor of the New Research Building in our university. Three researchers took three positioning terminals to walk along the same track at the same time. The positioning results stored in the terminals are shown in **Fig. 14**.

**Fig. 14** summarizes the performance of the TC-Image comparing to other indoor positioning methods, which shows that TC-Image based method obtained the best performance among the three approaches. As shown in **Fig. 14(a)**, the user's locations in the room 908 and 910 were precisely estimated. However, as shown in **Figs. 14(b)** and **14(c)**, some error positioning points are calculated based on WiFi and ibeacon system, which shown by yellow and red circle. From those two figures, we can find that the users' tracks passed the wall between two neighboring room, which were caused by signal fluctuations because of NLOS and without map matching.



**Fig. 14.** The positioning curve based on (a) TC-Image, (b) WiFi-based and (c) iBeacon.

In order to characterize positioning accuracy, we first manually ground truth the position and pose of each query image taken. This is done by using the 3D model representation of the mall and distance measurements recorded during the query dataset collection. For each query image, we are able to specify a ground truth yaw and position in the same coordinate frame as the 3D model and the output of the pose recovery step. Then, the estimated locations were compared to the ground truth location, which are shown by [Fig. 15](#).

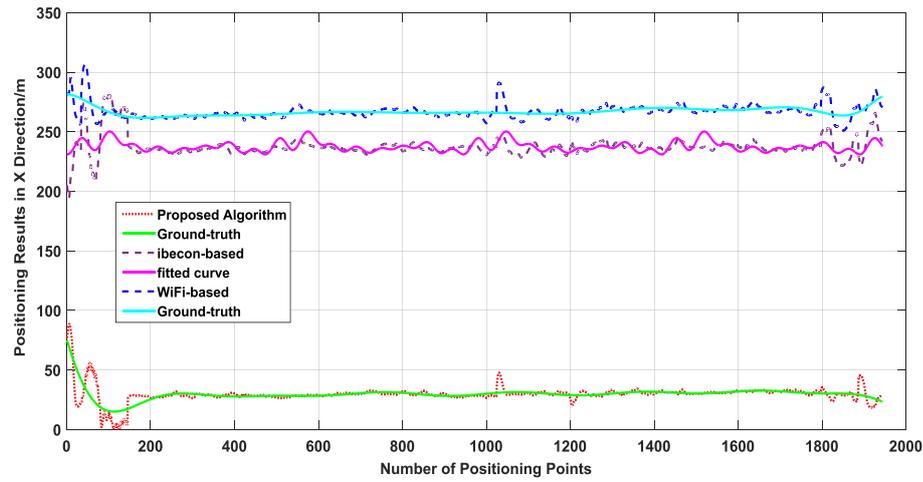
#### 2) Comparison of Positioning Results between Estimation and Ground Truth

According to the [Fig. 15\(a\)](#), we can find that the estimated locations in the horizontal direction are almost unanimous to the ground truth. However, at the initial location, the wave motions of the three methods are bigger because of the terminal initialization.

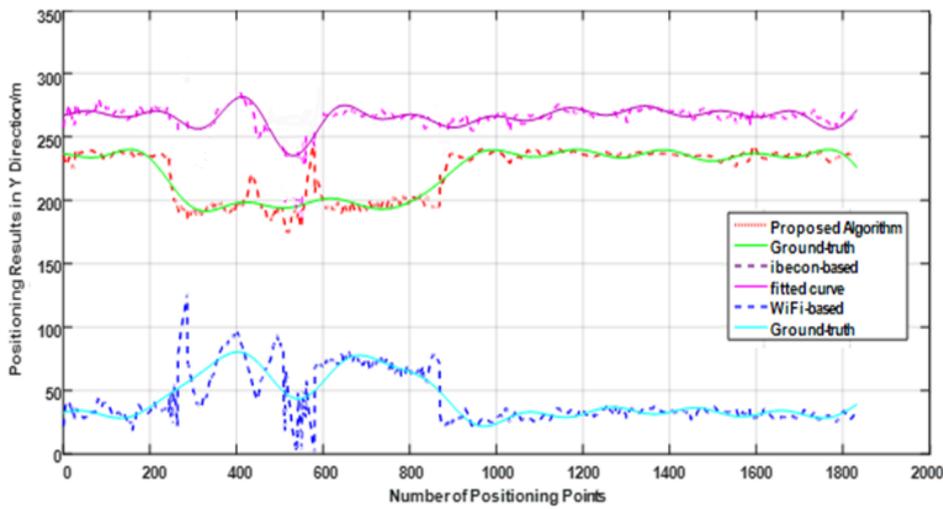
From the [Fig. 15\(b\)](#), we plot the estimated and ground truth locations of the plane direction onto the ninth 2D floorplan of the New Research Building. As seen from this figure, there is close agreement between the two based on our proposed method, which better than WiFi-based and iBeacon-based method.

As seen in [Fig. 12](#), when the location error is less than 1 meter, the FAST-SURF features of corresponding store signs present in both query and database images are matched together by

the RANSAC homography [17]. Conversely, in less accurate cases of location estimation where the positioning result error exceeds 4 meters, the RANSAC homography finds “false matches” between unrelated elements of the query and database images. In the example shown by Fig. 14(b), different objects in the corridor of the two images are matched together. In general, we find that images with visually unique signs perform better during location estimation than those lacking such features. Therefore, the proposed approach of extracting features from salient regions achieved to increase the positioning accuracy.



(a) In the horizontal direction



(b) In the vertical direction

Fig. 15. The positioning result based on TC-Image.

### 3) Root Mean Square Error

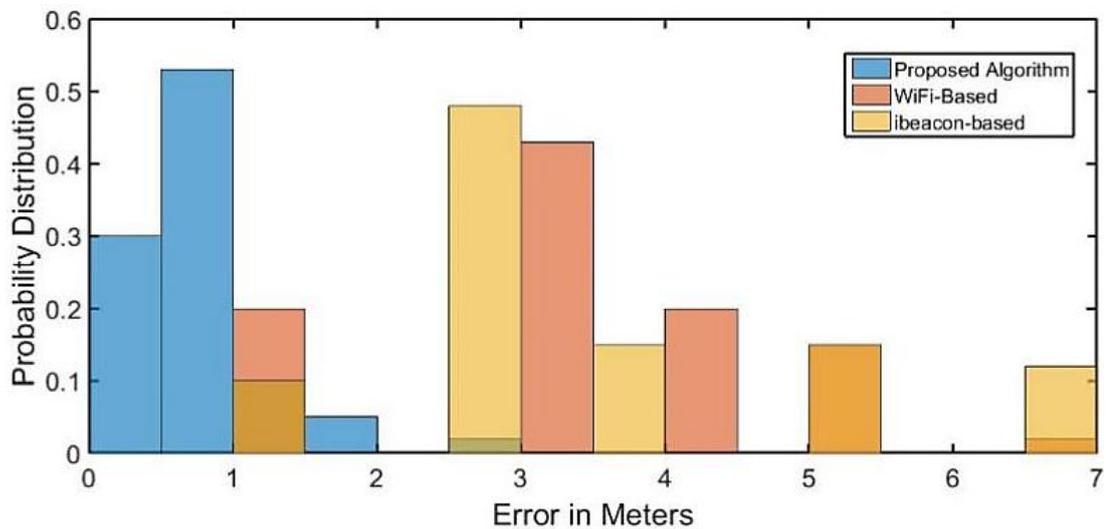
Root Mean Square Error (RMSE) is introduced to evaluate the proposed algorithm performance. The positioning accuracy can be computed as RMSE between the real position and their estimated position. It should be noticed that the WiFi-based approach obtained the location based on the fingerprint matching and the ibeacon-based approach obtained the location at the 2.4GHz in a circle area with 50 meters radius.

The estimation accuracy comparisons are listed in **Table 4**. As shown in **Table 4**, if the range measurement based on single sensor is unreliable, wireless signal with map matching cannot achieve a high accurate estimation. The RMSE of WiFi-based method is 2.317 m by using fingerprint matching, which is negative by sharp fluctuations of RSSI. Moreover, the RMSE of ibeacon-based method achieved to 3.143 m, which is negative by NLOS and signal fading. In contrast, our TC-Image based positioning algorithm is highly robust and can achieve a very accurate estimation with RMSE of 0.823 m. The comparison result improves that multi-sensors could obtain higher accuracy than single-senor. Furthermore, the image-based positioning method is robust to the NLOS and signal strength variance.

**Fig. 16** summarizes the performance of the probability distribution of the positioning error. As shown in Figs. 16, we are able to localize the position to within sub-meter level of accuracy for over 83% of the query images. Furthermore, 99% of the query images are successfully localized to within 2 m of the ground truth position. While, the positioning error of the WiFi-based and ibeacon-based methods exceed one-meters. 83% of the WiFi-based positioning results within 4 m, and 70% of the ibeacon-based positioning results within 4 m.

**Table 4.** Performance Comparison in Accuracy

Algorithm	RMSE(m)	Min Error(m)	Max Error(m)	Running Time(ms)
Proposed Algorithm	0.823	0.392	2.764	78.75
WiFi-based	2.317	1.286	5.819	67.51
ibeacon-based	3.143	1.074	6.673	89.06



**Fig. 16.** The probability distribution of the positioning error for three methods

## 5. Conclusion

This paper presented a smartphone indoor positioning method. The proposed solution is a hybrid solution, fusing multiple smartphone sensors with mobile communication signals and images. The smartphone sensors are used to measure the motion dynamics information of the mobile user.

This paper provides the experimental results of a system utilizing only the sensors available on a smartphone to provide an indoor positioning system that does not require any prior knowledge of floor plans, transmitter locations, radio signal strength databases, etc. Experimental results demonstrated the positioning accuracy is 0.823 m at horizontal and 0.5 m at vertical. The comparison to WiFi-based and iBeacon-based indoor positioning system shows that our approach is robust, while still being efficient. Because the operation of this method only uses the built-in hardware and computational resources of a smartphone, the positioning solution presented here is more cost-efficient and convenient for integration with related applications and services than alternative systems presented previously.

### Acknowledgments

This work is supported by Supported by National Natural Science Foundation of China (NO. 61401040) the National High Technology Research and Development Program ("863"Program) of China (NO. 2014AA123103).

### References

- [1] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779-782, 2008. [Article \(CrossRef Link\)](#).
- [2] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, pp. 1067-1080, 2007. [Article \(CrossRef Link\)](#).
- [3] Y. Gu, A. Lo, and I. Niemegeers, "A survey of indoor positioning systems for wireless personal networks," *Communications Surveys & Tutorials, IEEE*, vol. 11, pp. 13-32, 2009. [Article \(CrossRef Link\)](#).
- [4] R. Mautz and S. Tilch, "Survey of optical indoor positioning systems," in *Proc. of Indoor Positioning and Indoor Navigation (IPIN), 2011 International Conference on*, pp. 1-7, 2011. [Article \(CrossRef Link\)](#).
- [5] F. Evennou and F. Marx, "Advanced integration of WiFi and inertial navigation systems for indoor mobile positioning," *Eurasip journal on applied signal processing*, vol. 2006, pp. 164-164, 2006. [Article \(CrossRef Link\)](#).
- [6] A. Bekkali, H. Sanson, and M. Matsumoto, "RFID indoor positioning based on probabilistic RFID map and Kalman filtering," in *Proc. of Wireless and Mobile Computing, Networking and Communications, 2007. WiMOB 2007. Third IEEE International Conference on*, pp. 21-21, 2007. [Article \(CrossRef Link\)](#).
- [7] M. J. Kuhn, M. R. Mahfouz, N. Rowe, E. Elkhoully, J. Turnmire, and A. E. Fathy, "Ultra wideband 3-D tracking of multiple tags for indoor positioning in medical applications requiring millimeter accuracy," in *Proc. of Biomedical Wireless Technologies, Networks, and Sensing Systems (BioWireLeSS), 2012 IEEE Topical Conference on*, pp. 57-60, 2012. [Article \(CrossRef Link\)](#).
- [8] S.-H. Fang, C.-H. Wang, T.-Y. Huang, C.-H. Yang, and Y.-S. Chen, "An enhanced ZigBee indoor positioning system with an ensemble approach," *Communications Letters, IEEE*, vol. 16, pp. 564-567, 2012. [Article \(CrossRef Link\)](#).
- [9] M. Muñoz-Organero, P. J. Muñoz-Merino, and C. Delgado Kloos, "Using bluetooth to implement a pervasive indoor positioning system with minimal requirements at the application level," *Mobile Information Systems*, vol. 8, pp. 73-82, 2012. [Article \(CrossRef Link\)](#).

- [10] J. Barnes, C. Rizos, J. Wang, D. Small, G. Voigt, and N. Gambale, "Locata: A new positioning technology for high precision indoor and outdoor positioning," in *Proc. of 2003 International Symposium on GPS\ GNSS*, pp. 9-18, 2003. [Article \(CrossRef Link\)](#).
- [11] S. Mazuelas, F. A. Lago, J. Blas, A. Bahillo, P. Fernandez, R. M. Lorenzo, et al., "Prior NLOS measurement correction for positioning in cellular wireless networks," *Vehicular Technology, IEEE Transactions on*, vol. 58, pp. 2585-2591, 2009. [Article \(CrossRef Link\)](#).
- [12] D. Zhongliang, Y. Yanpei, Y. Xie, W. Neng, and Y. Lei, "Situation and development tendency of indoor positioning," *Communications, China*, vol. 10, pp. 42-55, 2013. [Article \(CrossRef Link\)](#).
- [13] Y. Cui and S. S. Ge, "Autonomous vehicle positioning with GPS in urban canyon environments," *Robotics and Automation, IEEE Transactions on*, vol. 19, pp. 15-25, 2003. [Article \(CrossRef Link\)](#).
- [14] J. Wang, R. V. Prasad, X. An, and I. G. Niemegeers, "A study on wireless sensor network based indoor positioning systems for context-aware applications," *Wireless Communications and Mobile Computing*, vol. 12, pp. 53-70, 2012. [Article \(CrossRef Link\)](#).
- [15] J. Z. Liang, N. Corso, E. Turner, and A. Zakhor, "Image based localization in indoor environments," in *Proc. of Computing for Geospatial Research and Application (COM. Geo), 2013 Fourth International Conference on*, pp. 70-75, 2013. [Article \(CrossRef Link\)](#).
- [16] L.-H. Chen, E. H.-K. Wu, M.-H. Jin, and G.-H. Chen, "Intelligent fusion of Wi-Fi and inertial sensor-based positioning systems for indoor pedestrian navigation," *Sensors Journal, IEEE*, vol. 14, pp. 4034-4042, 2014. [Article \(CrossRef Link\)](#).
- [17] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Proc. of Computer vision—ECCV 2006*, ed: Springer, pp. 404-417, 2006. [Article \(CrossRef Link\)](#).
- [18] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. of Computer Vision—ECCV 2006*, ed: Springer, pp. 430-443, 2006. [Article \(CrossRef Link\)](#).
- [19] Y. Benezeth, B. Emile, H. Laurent, and C. Rosenberger, "Vision-based system for human detection and tracking in indoor environment," *International Journal of Social Robotics*, vol. 2, pp. 41-52, 2010. [Article \(CrossRef Link\)](#).
- [20] C. Chen, W. Chai, Y. Zhang, and H. Roth, "A RGB and D vision aided multi-sensor system for indoor mobile robot and pedestrian seamless navigation," in *Proc. of Position, Location and Navigation Symposium-PLANS 2014, 2014 IEEE/ION*, pp. 1020-1025, 2014. [Article \(CrossRef Link\)](#).
- [21] Valgren, Christoffer, and Achim J. Lilienthal, "SIFT, SURF and Seasons: Long-term Outdoor Localization Using Local Features," *EMCR.*, 2007. [Article \(CrossRef Link\)](#).
- [22] Li, Xun, and Jinling Wang, "Image matching techniques for vision-based indoor navigation systems: performance analysis for 3D map based approach," *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*. IEEE, 2012.
- [23] Z. Tian, X. Tang, M. Zhou, and Z. Tan, "Fingerprint indoor positioning algorithm based on affinity propagation clustering," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, pp. 1-8, 2013. [Article \(CrossRef Link\)](#).
- [24] J. Wang, R. V. Prasad, X. An, and I. G. Niemegeers, "A study on wireless sensor network based indoor positioning systems for context-aware applications," *Wireless Communications and Mobile Computing*, vol. 12, pp. 53-70, 2012. [Article \(CrossRef Link\)](#).
- [25] L.-H. Chen, E. H.-K. Wu, M.-H. Jin, and G.-H. Chen, "Intelligent fusion of Wi-Fi and inertial sensor-based positioning systems for indoor pedestrian navigation," *Sensors Journal, IEEE*, vol. 14, pp. 4034-4042, 2014. [Article \(CrossRef Link\)](#).



**Jichao Jiao** received the PhD degrees in signal and information processing from Beijing Institute of Technology. He is an Assistant Professor in Beijing University of Posts and Telecommunications since 2013. His research activities include computer vision, machine learning, and indoor positioning.



**Zhongliang Deng** received PhD degree in mechanical manufacture from Tsinghua University, China. He is a professor and doctoral supervisor at School of Electronic Engineering in Beijing University of Posts and Telecommunications now. He holds the position of director of research with the Laboratory of Intelligent Communication, Navigation and Micro/Nano-Systems (ICNMNS). He is also the executive vice President of BUPT. His research interests include indoor and outdoor seamless positioning, GNSS, satellite communications, MEMS and multimedia.



**Liming Xu** received the PhD degree in the Department of electronic science and technology, Beijing University of Posts and Telecommunications, Beijing, China. He is the CTO of Beijing RTMAP Technology Co. Ltd, Beijing, China.



**Fei Li** is a graduate student in the Department of Electronic Engineering at Beijing University of Posts and Telecommunications. Her research interests include image processing, pattern recognition, computer vision and visual-aided indoor/outdoor navigation.