Fusion of WiFi and Vision based on Smart Devices for Indoor Localization

Jing Guo
College of Information Science and Technology
Northwest University
Xi'an, China
guo@nwu.edu.cn

Shaobo Zhang
College of Information Science and Technology
Northwest University
Xi'an, China
201531437@stumail.nwu.edu.cn

Wanqing Zhao*
College of Information Science and Technology
Northwest University
Xi'an, China
zhaowq@nwu.edu.cn

Jinye Peng
College of Information Science and Technology
Northwest University
Xi'an, China
pjy@nwu.edu.cn

ABSTRACT

Indoor localization is an important problem with a wide range of applications such as indoor navigation, robot mapping, especially augmented reality(AR). One of most important tasks in AR technology is to estimate the target objects' position information in real environment. The existed AR systems mostly utilize specialized marker to locate, some AR systems track real 3D object in real environment but need to get the the position information of index points in environment in advance. The above methods are not efficiency and limit the application of AR system, so that solving indoor localization problem has significant meaning for the development of AR technology. The development of computer vision (CV) techniques and the ubiquity of intelligent devices with cameras provides the foundation for offering accurate localization services. However, pure CV-based solutions usually involve hundreds of photos and pre-calibration to construct an densely sampled 3D model, which is a labor-intensive overhead for practical deployment. And a large amount of computation cost is difficult to satisfy the requirement for efficiency in mobile device. In this paper, we present iStart, a lightweight, easy deployed, image-based indoor localization system, which can be run on smart phone and VR/AR devices like HTC Vive, Google Glasses and so on. With core techniques rooted in data hierarchy scheme of WiFi fingerprints and photos, iStart also acquires user localization with a single photo of surroundings with high accuracy and short delay. Extensive experiments in various environments show that 90 percentile location deviations are less than 1 m, and 60 percentile location deviations are less than 0.5 m.

*Corresponding author

VRCAI '18, December 2–3, 2018, Hachioji, Japan

ACM ISBN 978-1-4503-6087-6/18/12…$15.00
https://doi.org/10.1145/3284398.3284401

CCS CONCEPTS

• **Computing methodologies** → *Scene understanding*; *3D imaging*; *Reconstruction*;

KEYWORDS

indoor localization, image-based localization, WiFi fingerprint, smart devices, CV model

ACM Reference format:
Jing Guo, Shaobo Zhang, Wanqing Zhao, and Jinye Peng. 2018. Fusion of WiFi and Vision based on Smart Devices for Indoor Localization. In *Proceedings of International Conference on Virtual Reality Continuum and its Applications in Industry, Hachioji, Japan, December 2–3, 2018 (VRCAI '18),* 8 pages.
https://doi.org/10.1145/3284398.3284401

1 INTRODUCTION

Accurate indoor location has a great effect on commercial location-based services for smart devices. Meanwhile, indoor location through smart phone and VR/AR devices provides users a good experience. As a technology to connect the real world, indoor location is important to the AR field. If a device wants to provide a 3D effect AR, it is obvious that the motion tracking module is absolutely necessary, location (motion tracking) and building maps form the core function of indoor location. Therefore, indoor location can be said to be the most basic module in AR, belonging to the category of device-aware surrounding environment. Based on the above situation, we need a localization system that is easily de-ployable with high accuracy (sub-meter) and short delay. Indoor location which only uses common camera as equipment to collect data without extra hardware, has become an aroused general interest [Li et al. 2012; Rai et al. 2012].

Crowdsourced WiFi-based fingerprinting [Jiang et al. 2012; Yang et al. 2012]and image-based method [Mautz and Tilch 2011; Mulloni et al. 2009; Tilch and Mautz 2013] are two mainstream easy-to-use approaches,which usually achieve meter-level accuracy. In these systems, crowdsourced data is used to construct model and keep up-to-data on model. In crowdsourcing schemes, it takes a long time to collect users' data which need a sustainable incentive mechanism to ensure data quality. The image-based methods require

the high-quality CV model which is constructed by a huge image database. The SFM-based methods utilize photos from network and crowdsourced data to construct the CV model [Frahm et al. 2010; Snavely et al. 2010; Wu 2013] , which needs a large number of photos for bundle adjustment [Wu et al. 2011]. And it still has the problem of incompleteness of CV model. The use of extra hardware such as deep cameras, binocular cameras and infrared range finders can help on the build process. As a promising alternative is to combine the two methods described above [Dong et al. 2015], which achieves the WiFi fingerprint technology to rough location and image-matched technology for the accuracy location.

We observe an opportunity for an easy-to-deploy indoor localization system, which preserves the sub-meter accuracy, and also significantly reduces the overhead involved in pure CV-based image database construction and fingerprint database construction. The key idea is to leverage minimal images to construct WiFi marked sub-area CV model which is implemented by the WiFi site survey and the image optimizing matched method. Based on the above intuition, we propose iStart: a low cost deployable and highly accurate indoor location system which only need smart phone without extra hardware. Users can acquire their accurate position information by app-iStart in complex indoor environment.

In designing iStart into a practical localization system, two main challenges need to be addressed:(1) How to improve the accuracy of three-dimensional(3D) coordinates for CV model with minimum data. (2) How to improve system efficiency by using the combined method(fingerprint matched and image matched methods). The accuracy location requires the high-quality CV model which is composed of feature coordinates. To construct CV model, some extra hardware or a huge amount of images are needed. In terms of the first challenge, we utilize the Fundamental Matrix [Nistér 2004] to improve the proportion of feature points for correct matches between photo pairs which help generating accurate 3D coordinate of feature points. To reduce labor and time costs, we design a set of optimization techniques to derive high-quality CV model. For the second challenge, we design the strategy of WiFi marked sub-area which partition the feature space into sub-area. To improve the efficiency of our system, we implement a image-based location method which may relieve the data transmission of network and service workload.

We fully prototype iStart on Android platform and conduct extensive experiments in a larger complex indoor environment. Evaluations demonstrate that iStart achieves the average location error 0.568m, which preserves the accuracy of image-based localization schemes. In addition, iStart greatly reduces the workload to construct the database and achieves the process of image-based accurate location by client.

The key contributions are summarized as follows.

- We implement a prototype system – iStart evaluates the feasibility of building an indoor localization system. The experimental results demonstrate that iStart works properly and has good performance. The experiments are carried out in the real buildings totally covering around $800m^2$, in most case, iStart can locate a user with 4 seconds, with an average location error of less than 0.6 meters and a facing direction error of less than 6 degrees.
- Based on the combination of fingerprint -matched and image-matched technology, iStart implements the processing of fingerprint-marked image location which speeds up the positioning and improves system efficiency.
- It presents a low-cost solution which is calculated by Fundamental Matrix to construct a CV model.
- It demonstrates a edge computing-based working system and evaluates it in a real-world large-scale deployment.

2 RELATED WORK

There are many applications for indoor localization system. Some systems, such as Cricket [Priyantha et al. 2000] and PinPoint [Youssef and Agrawala 2005], allocate dedicated hardware deployment to achieve high position accuracy. Many methods such as Radar [Bahl and Padmanabhan 2000] and Horus [Youssef and Agrawala 2005] leverage existing WiFi infrastructure. Some recent works explore the magnetic field for indoor localization[Chung et al. 2011; Riehle et al. 2012]. All these methods require labor-intensive site surveys. Compared with its counterpart(e.g. wireless and inertial-based), image-based localization can easily yield sub-meter accuracy [Mautz and Tilch 2011]. The high accuracy makes image based approaches a fit for robot localization and navigation[Bonin-Font et al. 2008]. Adopting pure image-based localization method to smartphones incurs two limitations. (1) Hundreds of photos are required to construct the high-quality 3D CV model for location. (2) The process of image-based location is slow due to heavy computation. iStart aims to overcome these limitations and enable easy-to-use image-based localization on unmodified smartphones.

Conventionally, 3D models for indoor mapping are generated from the data captured by laser scanners [Okorn et al. 2010] and/or depth cameras through war-driving [Newcombe et al. 2012]. Google Cartographer and Xsens Scannect [Chow 2014] build the indoor map by Simultaneous Localization and Mapping(SLAM) based tool, which is equipped with inertial measurement units in addition to laser scanners and deep cameras. SfM-based methods enable 3D modeling of indoor environment using unordered 2D photos, which are taken by ordinary devices like smartphone cameras. Based on SfM technology, Agarwal [Agarwal et al. 2011] constructed 3D models of Rome city on 150K photos from Internet. iMoon[Dong et al. 2015] constructs a real environment by thousands of photos from the user. In order to meet the needs of accuracy location, a large number of photos are processed in bundle adjustment of SfM. Instead of constructing the whole 3D model of a building, iStart aims to reduce the overhead of image database construction for localization. Therefore iStart focuses on sub-area 3D model and tries to combine with the technology of WiFi fingerprint.

By implementing multi-modal localization incorporating Manifold Alignment (WiFi and floorplan) and Trapezoid Representation, ClickLoc[Xu et al. 2016] speeds up the image location process. iMoon[Dong et al. 2015] uses WiFi fingerprint to acquire the coarse location of user, and selects the partitions that cover these coarse locations. Different partition of 3D model(Geographical space), iStart defines the sub-area model by feature space partition.

3 SYSTEM OVERVIEW

Our system iStart is built on a client/server architecture. iStart provides a friendly 3D map for better reading. Based on the construction of the sub-area CV model, the total data transmission between the server and the client is reduced, while the system utilization is improved. iStart is achieved as a high efficiency, low consumption and accuracy location system.

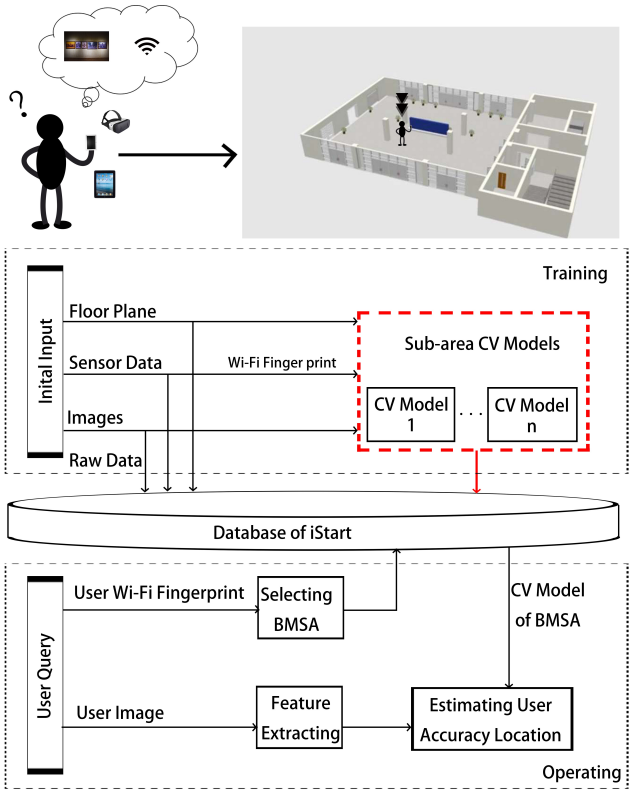


Figure 1: Overview of iStart system, a mobile image-based indoor location system for smartphones.

The Figure 1 shows the structure of iStart. At beginning, iStart divides the indoor area into multiple sub-areas, and obtains the raw data (including WiFi fingerprint and photos) on the collection point (CP) in each sub-area. During the training steps, the WiFi fingerprints of all sub-areas constitute the WiFi fingerprint database. Meanwhile, the photos are calculated to build the CV model for each sub-area. And, these sub-area CV models comprise the CV model database. When a user makes a request for localization, the fingerprint and the query photo are obtained by user's phone. The user's fingerprint is sent to server and matched with fingerprint database to obtain the best matched sub-area (BMSA). Then, the CV model of BMSA is sent to client. Meanwhile, the user's photo is processed by image feature extraction method. After that, the extracted features from user's image are matched with BMSA's CV model to estimate user's accuracy location and facing direction.

In architecture of iStart, the client is responsible for the following two tasks: (1) providing a photo and real-time WiFi fingerprint. (2) executing the process of image-based location.

The server has three main functions: (1) building CV models of sub-area. (2) constituting CV model database and associated fingerprint database. (3) executing WiFi fingerprints-match process.

In the location process of iStart, only a small amount of data has to be transmitted between the server and the client. And the server and the client process data respectively. The working process of iStart not only reduces server pressure but also significantly improves the system efficiency.

4 TRAINING

This section describes how we build CV model utilizing fewer data. Firstly, we explain that how to collect data by smartphone on indoor environment in section 4.1 In section 4.2, we describe how we build CV model and fingerprint database with collected data.

4.1 Data Collection

Building a CV model of indoor environment is a preliminary work for location system. The indoor environment is very complex, which includes two types of scenes, room-level environment and large open environment. We believe that room-level environment seems like the office or school laboratory, which is divided into multiple small rooms or cells by obstructions (like walls or high bookshelf). While large open indoor environment seems like museums or exhibition centers, which is no or almost no obstructions (walls) in the whole region.

In order to get complete data, we usually collect photos at the centers of small rooms. However, the large open indoor environment usually has a large open area with sparse indoor infrastructures, which indicates that most of region is accessible to people and the collocation of measurement points lack dependable boundary conditions. In addition, the people density is usually high in large open indoor environment, which indicates extensive requests for localization. In our view, dividing the open space into multiple sub-areas and building model for each sub-area can effectively improve the accuracy of localization and relieve the server stress.

Usually, the CPs are collocated in the center of the sub-areas (on a two-dimensional plane). On each CP, we acquire one WiFi fingerprint and eight photos (take pictures with every 45 degrees). In order to construct the sub-area feature space, we set auxiliary collection points for every CPs at a distance which is about 10cm. Eight photos are collected with the same direction of the CP for each ACP. Two photos with the same angle (from the CP and its ACP) as a photo pairs, which are processed by iStart to obtain the three-dimensional coordinates of the feature points. We used a self-built toolkit, as shown in Figure 2, to collect same direction photos. At each CP, we took photos and collected WiFi fingerprints using the Android phone(Redmi 5, MI) placed on the tripod.

4.2 WiFi Fingerprints Database Construction

A WiFi RSS fingerprint is gathered on each CP via phone. On each CP, we collected WiFi RSS within 60s and averaged data as a WiFi fingerprint. One of WiFi fingerprints is associate with a sub-area.

Figure 2: Photo collection device. The collection device only requires a mobile phone and a tripod. This greatly reduces the cost of collecting data.

All the sub-area WiFi fingerprints and the position of CPs are stored in the database.

4.3 Sub-area CV Model Construction

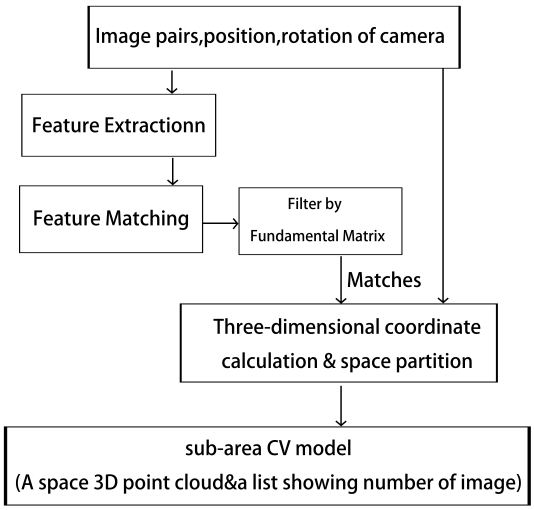


Figure 3: Sub-area CV module construction location

Different from bootstrap-based method which applies SfM structure to construct a CV module, iStart builds sparse 3D point cloud of the real environment by sub-areas using a series ordered photo pairs. A photo pair includes two photos which are taken with the same direction from a CP and its related ACP. On each sub-area, there are 8 photo pairs to construct a sub-area CV model. The process of construction of sub-area CV model is shown in Figure 3. The specific steps are as follows:

1) Feature extraction. In this step, the highly distinctive and invariant features are extracted from images. iStart extract image features which are described with Oriented FAST and Rotated BRIEF (ORB) algorithm [Rublee et al. 2011]. We adopt the ORB algorithm as it is orders of magnitude faster than SURF and SIFT [Lowe 1999] and can extract image features in real time on smart devices.

2) Feature matching. iStart tries to match the features between two photos of photo pair. The main information of features include position,neighborhood diameter,feature direction,response intensity,milt-scale information and classification. The implementation of feature point matching is to match the information of the feature points one by one.In order to improve the efficiency of matching, we employ Milt-Probe LSH approach to KNN match algorithm.This method is optimized for basic LSH method to query a large number of hash tables to ensure the shortcoming of search quality. The core idea is to use a strictly selected probe sequence to detect multiple buckets that may contain nearest neighbors. According to the nature of LSH, if the p is the neighbour of query point then p is likely to be in the bucket near the bucket where query is located. The purpose of milt-probe LSH is to find these adjacent buckets and increase the probability the neighbours of a given query object q. Then we use KNN to match the features between two photos of photo pairs.We have found that there can be some possible errors while matching which may affect the result. So we filter the mismatch point pairs by Euclidean Distance, furthermore, istart uses Fundamental Matrix to reduce the wrong matches. Fundamental Matrix contains rotation information(R) and transform information(T) of two cameras, as equitation (1) displays where M_r, M_l are the internal matrix of left and right cameras,R, T is the rotation and transform matrix.

$$F = {M_r}^{-T} R T {M_l}^{-1} \tag{1}$$

T and R describe the relative position of one camera to another camera in global coordinate system. Besides Fundamental Matrix also contains the internal parameters of two cameras so that it can make connection between two cameras in pixel coordinate. So good matches which provide correct estimation is the basis for building high quality CV model of the sub-area.

3) Three-dimensional coordinate calculation. When matched features are found, their 3D coordinates are calculated through camera poses (position of CPs and ACPs), focal length and rotation (gyroscope). istar uses pts1, pts2 which are the match points of two photos and internal matrix to calculate the essential matrix. The defining equation for the essential matrix is

$$x'^{T} E X^{T} = 0 \tag{2}$$

in terms of normalized image coordinates for corresponding points x to x'. Then recover relative rotation and translation from an essential matrix and the corresponding points in two images, using cheirality check. Returns the number of in-liners which pass the check. Then we use triangulation to evaluate the depth of the image. Triangulation refers to the observation of the angle between the same point through two places to determine the distance of

the points.We use triangulation to estimate the pixel distance. Triangulation is solved by least-square method. First we input the projection matrix and internal matrix of camera, then convert to non-homogeneous coordinates, at last verify the relationship between the triangle and the re-projection of the feature points. As following equations display, s indicates vertical distance of the camera image plane, x indicates the pixel position of the point X, K indicates the camera internal matrix, R and T indicates the rotation and transform matrix.

$$sx = K(RX + T) \tag{3}$$

There are two unknowns in this equation,s and X.Do the outer product on both sides of the equation and you can eliminate s.

$$0 = \widehat{x}K(RX + T) \tag{4}$$

Finishing up can get a linear equation about the space coordinate X.

$$\widehat{x}KRX = -\widehat{x}kt \tag{5}$$

The above equation cannot be solved directly, so it is transformed into homogeneous equation equation.

$$\widehat{x}K(R\ \ T)\begin{pmatrix} X \\ 1 \end{pmatrix} = 0 \tag{6}$$

Use SVD to find the zero space of the X left matrix, and then normalize the last element to 1, we can find X.

4) Sub-area feature space partition. We set the feature space of sub-area as a circle. Here, the radius of the circle is denoted by d. The distance of feature is Df, which is calculated below.

$$Df = \sqrt{(xf - xm)^2 + (yf - ym)^2 + (zf - zm)^2} \tag{7}$$

where xf, yf, zf are the three-dimensional coordinates of CP. The xm, ym, zm are three-dimensional coordinates of feature point. The value of d is usually four or five times of P.P is the average of Df. The features which Df are too larger than d will be abandoned.

A sub-area CV model is a 3D point cloud which made up of feature points. The value and the 3D coordinates of feature points and a list which showing what images and which features in these images are saved in the sub-area CV model. All of these sub-area CV models constitute sub-area CV models database.

5 INDOOR LOCALIZATION

When the user requests localization, a photo and a real-time WiFi RSS fingerprint are acquired by user's phone. The user's WiFi RSS is sent to server matched with fingerprint database to find BMSA. Then the CV model of BMSA is returned to client. The server processes WiFi matching, while the client starts to do feature extracting of the user image. After receiving the BMSA CV model, the client performs feature matching process and calculates the user's accuracy location and facing direction.

The process of localization on server and client as described in Algorithm 1 and Algorithm 2 respectively.

6 BENCHMARK VALIDATION

Previous localization systems like Zee in [Rai et al. 2012] and LiFS in [Yang et al. 2012] can achieve high localization accuracy within meters in room-level environment, where leverage dense indoor constraints and reference points (like an elevator, stairs). Some indoor places like museums and exhibition centre are no or almost no obstructions(walls) in the whole area. In these scenarios, accurate positioning is an urgent need for visitors. Here we validate iStart against open large environment and room-level environment altogether. Validation is done by two experiments, image feature matching and sub-area image-based accuracy location. Validation is done by two experiments, image feature matching and sub-area image-based accuracy location.

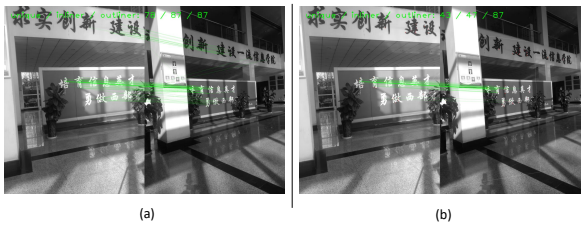
Algorithm 1 Pseudo code of indoor localization algorithm implemented on servicer

Input: WiFi Fingerprints: UWiFi
Output: User's sub-area: Us
1: Lcoarse=k-NN(UWiFi, DBWiFi)
//Apply k-NN to get best matched sub-area based on WiFi fingerprints
//DBWiFi: Database of sub-area WiFi finger-prints
2: find Rphoto
//find reference photos of best matched sub-area

Algorithm 2 Pseudo code of indoor localization algorithm implemented on client

Input: User's query photo: Uphoto
Output: User's location: Uloc, User's facing direction: Ufac
1: Ufeature=ORB(Uphoto)
//Apply ORB to extract feature key points of Uphoto, Ufeature
2: Rphoto
//DBphoto:Database of geo-referenced photos features
//Apply best matched sub-area to find reference photos on DBphoto, Rphoto
//Rphoto is defined by feature of reference photos
3: KPpairselect=F(Ufeature , Rphoto)
//Apply Fundamental Matrix to match KPs
4:Laccuracy=(KPpairselect, Camerapositio, Camerarotation)
//Identify use's location and facing direction based on feature maching between Ufeature and Rphoto
// Cameraposition: the position of camera
//Camerarotation: rotation of camera sub-area

6.1 Image Feature Accuracy matching



(a) (b)

Figure 4: image pairs matched by ORB: (a) key points matching result by ORB without filter. (b) key points matching result by using Fundamental Matrix to remove wrong matches.

Building high-quality CV model requires accurate 3D coordinates of the feature points. In reality, we have seen that there can be some possible errors while matching which may affect the result, as shown in Figure 4 (a). To solve this problem, iStart uses Fundamental Matrix (which can select good matches). The good matches provide correct estimation for CV model. To evaluate our method, we conducted a verification experiment. Here, we choose 100 images randomly from experiments, in Section 6. We use RANSAC technology[Nistér 2004] to capture Fundamental Matrix F of image pairs. A minimum of 8 such points are required to find the fundamental matrix (while using 8-point algorithm). More points are preferred and use RANSAC to get a more robust result. When Fundamental Matrix is determined, the wrong match will be removed. The result is shown as Figure 4. Obviously, there are a large number of feature point matching errors, in Figure 4 (a). This will not provide valid data for sub-area CV model construction. The filtered results shown in Figure 4 (b), we can see that the matching feature points to reduce, but the accuracy rate increased significantly.

6.2 sub-area image-based accuracy location

Through feature matching we can get the corresponding points of query image and reference image, the information of corresponding points contains the key points of the query image and 3d position and key points of the reference points. To achieve accuracy location base on these information, we utilize PnP method which can find an object pose from 3D-2D point correspondences. The function estimates the object pose given a set of object points, their corresponding image projections, as well as the camera matrix and the distortion coefficients. As follow equation displays,

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{8}$$

in this equation, f_x, f_y, c_x, c_y are parameters of camera internal matrix, $[r, t]$ is transform and rotation matrix. Solve this equation by the 3d position and key points, we can evaluate the accuracy location of the query photo.

7 EVALUATION

We implements a prototype to evaluate the indoor location system iStart. To verify the effectiveness of our method, we choose two different environments. Here, two different scenes are chosen for testing; room-level environment and open large environment. The two scenes are as shown in Figure 5, where Figure 5 (a) is the dining and leisure(D & L) area, Figure 5 (b) is the office lobby area. In the D & L area (about $300m^2$), there are some tables, chairs and bookshelf. The office lobby area is relatively more empty.

The black points represent test points(TP), in Figure 5. The TPs are randomly selected by 3 volunteers. For each location they chose, they were asked to take 10 to 16 photos with the camera facing different directions. Meanwhile, a WiFi fingerprint was record by smart device. For each photo, the positions of the actual location and the facing direction of the camera were recorded and used as the ground truth for accuracy analysis in data set B.

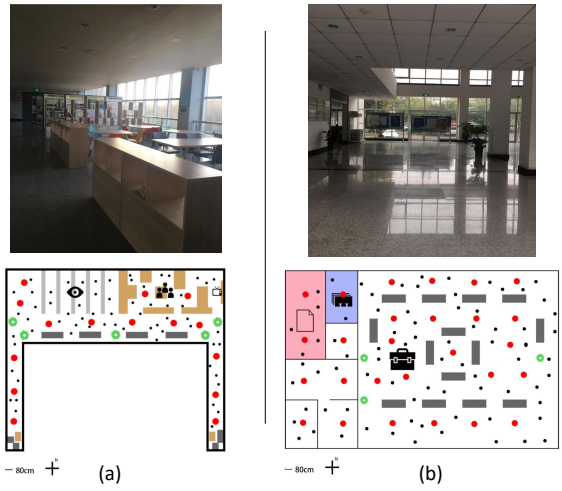


Figure 5: Photos and floor plans of (a) room-level environment and (b) open large environment. The red points respect the position of data collected. And the black points is the position where the test data taken.

Table 1: The Database for iStart

Data set	Content	Usage	Time spend
A	42 WiFi fingerprints 336 photo pairs	Training	5 hours
B	150 WiFi fingerprints 2026 photos	Evaluation	1 weeks

7.1 System implementation

We also implement an Android-system app running our system. We deployed iStart on a server (Pentium(R) Dual-Core CPU , Memory 4G). In experiment, we used normally configured smart phones to evaluation our system.

When a user in the campus hall, the app can get accuracy location of the user, as shown in Figure 1. Given a query photo and WiFi RSS, the app is expected to return the location where the photo is taken and the direction in which the camera is facing. Once the user's location is identified, the app will display the user's current location on a 3D map.

The floor plans of two scenes are shown in Figure 5, where the red points represent the position of CPs. Because the difference of two scenes, the average distance between CPs are P1=3.2m, P2=3.8m respectively. A WiFi fingerprint and 8 groups of photos are collected on each location of CP, shown in Table 1 data set A. We construct WiFi marked sub-area CV models for these two scenes, according to the method described in Section 3.

7.2 Hit rate

Hit rate refers to the percentage of measurement points that can be located. In our experiment the hit rate of WiFi fingerprint is 100%, in other words, the system can determine the user sub-area correctly. The hit rate of iStart and image-based location is the same. The image-based location method is to obtain the accurate

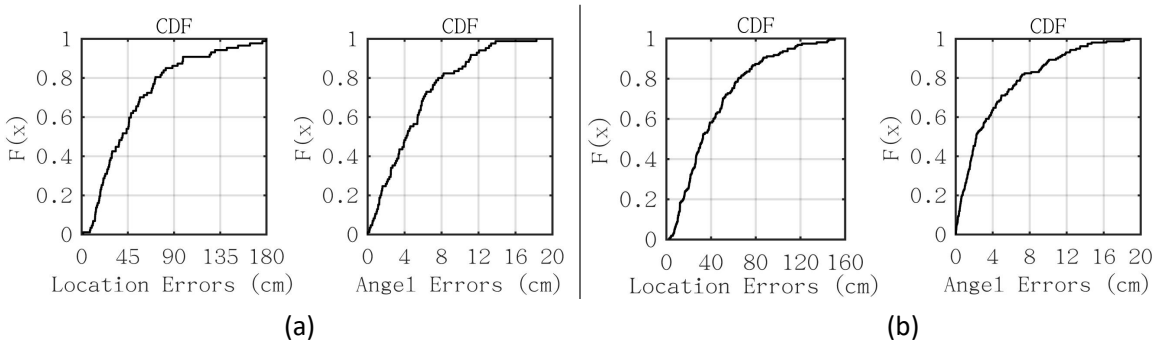


Figure 6: Location error and direction error of (a) scene 1 and (b) scene 2

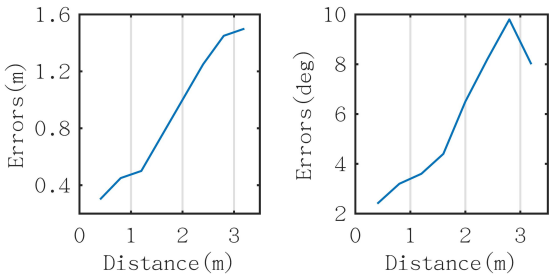


Figure 7: Location and facing direction errors in the case where different distance between CP and TP.

Table 2: Hit rate of istart in differents

Area	WiFi	Image	iStart
Reading Room	100%	93.1%	93.1%
Corridor	100%	95.8%	95.8%
Hallway	100%	97.5%	97.5%

position of the user by matching feature points. When the user's photo is lacking valid feature points, the user's exact location will not be estimated. Therefore, when an user-provided photo lacks valid feature points, such as white walls, the system will ask the user to re-send the photo.

7.3 Accuracy

When the user makes a location request and provides a photo, the system will output the location of the user whose coordinates will be displayed on the map. To evaluate the accuracy of iStart, we compare the actual location and facing direction of the user with the result of iStart.

In Scene 1, the number of test photos is 962, of which 879 are successfully located, with a success rate of 96.13%. In Scene 2, the number of test photos is 1064, while the success rate is 98.53%. In reality, photos taken by different users at the same location may cover different scenes. This is due to the user's direction of face and camera configuration such as focal length, image resolution. In the experiment, the test points are randomly distributed throughout the area where are accessible to humans. And on each test point, 10-16 photos which are facing a different direction were taken by volunteers.

The results of two scenes are shown in Figure 6. The average location error of the two scenes is less than 0.6 meters, while the direction error is less than 6 degrees. In these cases, approximately 30% of test cases are less than 0.2 meters and 60% of test cases are less than 0.5 meters.

Meanwhile we statistics the results of the errors with different distance between CP and TP, which is shown in Figure 6. When the distance is less than 0.5 meter, the error is less than 0.4m. When the distance is 2m, the average location error is about 1m. As the distance becomes larger, the error increases. If the position of a user photo taken is nearly a position of CP, it may get lower error.

7.4 Response Delay

Nowadays mobile phones possess powerful computation and communication capability, and are equipped with various functional built-in sensors. We use the strategy of fingerprint marked sub-area to decrease the computational complexity of the mobile terminal. It also greatly reduces the data transmission between the server and the client. Because of these optimization strategies, iStart can quickly locate by lower-profile phones and servers.

On the experiment described in Section 7, the average location time is no more than 4s. It includes the time of fingerprint location on server, CV model data transmission and sub-area image location on the user phone. Where the average time of fingerprint location is 0.8s, the sub-area image location is 2.9s, and data transmission is approximately 1s.

7.5 Comparsion

A large amount of data collection is a challenge to build indoor positioning systems. Crowdsourced data can built CV model, while collecting crowdsourced data may cost a very long time. The prototype system-iStart we built can quickly build CV model with just smart phone. As the representative indoor positioning systems of the same type, both iMoon [Dong et al. 2015] and ClickLoc [Xu et al. 2016] have excellent precision, but the crowdsourcing data acquisition need much longer collection time, as shown in Table 3. A database that can provides accurate positioning often takes months to complete. Our system uses WiFi marked sub-area partitioning

and image match algorithm can achieve the same effect of crowdsourcing acquisition systems, and just needs about $0.43min/m^2$ collection time.

Table 3: System comparsion

	Average Location error/m	Average degegree error/°	Maximize Location-time/s
iMoon	2	6	4
ClickLoc	1	–	9
iStart	1.3	6	4

8 CONCLUSION

There are more applications of indoor location in reality. AR is one of the most promising fields which indoor location plays a crucial role in. With the development of science and technology, more and more virtual information will gradually become part of our real life, and we humans cannot handle these massive virtual information. But we have developed AI. In the future, with artificial intelligence, we can process large amounts of data and virtual information and make it available to us.Virtual information will become part of our real life. At this time, we will use AR technology to connect these data and information with the real world in a natural way. As a technology to connect the real world, indoor location is important to the AR field. A common complaint they have about indoor location systems is hard to build, which require a large amount of images and labor. The bootstrap-based approach builds the CV model from user data, but this process takes a long time to collect data and bundle adjustment calculations. Also, the lack of data leads to incomplete problems with the CV model.Based on the above situation, we present iStart- an low cost deployable and highly accurate indoor location system which only need smart phone without extra hardware. iStart is built on several existing techniques, e.g. image-match technique and fingerprint-match technique, and address several technical challenges to build a CV model with few images. iStart would be a good primer for any system which considering providing indoor location services.iStart implements accuracy indoor location system and proposes a method to build initial CV model for bootstrap-based location system.

ACKNOWLEDGMENTS

This paper was supported by National Key R&D Program of China (2017YFB0203104) and the Shaanxi Province Natural Science Foundation(2016JQ6077).

REFERENCES

Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz, and Richard Szeliski. 2011. Building Rome in a Day. *Commun. ACM* 54, 10, 105–112.

Paramvir Bahl and Venkata N Padmanabhan. 2000. RADAR: an in-building RF-based user location and tracking system. *Proc IEEE Infocom* 2 (2000), 775–784.

Francisco Bonin-Font, Alberto Ortiz, and Gabriel Oliver. 2008. *Visual Navigation for Mobile Robots: A Survey*. Vol. 53. Kluwer Academic Publishers, Hingham, MA, USA. 263–296 pages.

Jacky C. K. Chow. 2014. Multi-Sensor Integration for Indoor 3D Reconstruction. *University of Calgary* (2014).

Jaewoo Chung, Matt Donahoe, Chris Schmandt, Ig-Jae Kim, Pedram Razavai, and Micaela Wiseman. 2011. Indoor Location Sensing Using Geo-magnetism. In *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services (MobiSys '11)*. ACM, New York, NY, USA, 141–154.

Jiang Dong, Yu Xiao, Marius Noreikis, Zhonghong Ou, and Antti Ylä-Jääski. 2015. Demo: IMoon: Using Smartphones for Image-based Indoor Navigation. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems (SenSys '15)*. ACM, New York, NY, USA, 449–450.

Jan Michael Frahm, Pierre Fite-Georgel, David Gallup, Tim Johnson, Rahul Raguram, Changchang Wu, Yi Hung Jen, Enrique Dunn, Brian Clipp, and Svetlana Lazebnik. 2010. Building Rome on a cloudless day. In *European Conference on Computer Vision*. 368–381.

Yifei Jiang, Xin Pan, Kun Li, Qin Lv, Robert P. Dick, Michael Hannigan, and Li Shang. 2012. ARIEL: Automatic Wi-fi Based Room Fingerprinting for Indoor Localization. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12)*. New York, NY, USA, 441–450.

Fan Li, Chunshui Zhao, Guanzhong Ding, Jian Gong, Chenxing Liu, and Feng Zhao. 2012. A reliable and accurate indoor localization method using phone inertial sensors. In *ACM Conference on Ubiquitous Computing*. 421–430.

David G. Lowe. 1999. Object Recognition from Local Scale-Invariant Features. In *Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2 (ICCV '99)*. IEEE Computer Society, Washington, DC, USA, 1150–.

Rainer Mautz and Sebastian Tilch. 2011. Survey of optical indoor positioning systems. In *International Conference on Indoor Positioning and Indoor Navigation*. IEEE, Guimaraes, Portugal, 1–7.

Alessandro Mulloni, Daniel Wagner, Istvan Barakonyi, and Dieter Schmalstieg. 2009. Indoor Positioning and Navigation with Camera Phones. *IEEE Pervasive Computing* 8, 2 (April 2009), 22–31.

Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. 2012. KinectFusion: Real-time dense surface mapping and tracking. In *IEEE International Symposium on Mixed and Augmented Reality*. 127–136.

David Nistér. 2004. An Efficient Solution to the Five-Point Relative Pose Problem. *IEEE Trans. Pattern Anal. Mach. Intell.* 26, 6 (June 2004), 756–777.

Brian Okorn, Xuehan Xiong, Burcu Akinci, and Daniel Huber. 2010. Toward Automated Modeling of Floor Plans. *Proceedings of the Symposium Ond Data Processing Visualization and Transmission* (2010).

Nissanka B. Priyantha, Anit Chakraborty, and Hari Balakrishnan. 2000. The Cricket Location-support System. In *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MobiCom '00)*. ACM, New York, NY, USA, 32–43.

Anshul Rai, Krishna Kant Chintalapudi, Venkata N. Padmanabhan, and Rijurekha Sen. 2012. Zee:zero-effort crowdsourcing for indoor localization. 293–304.

Timothy H. Riehle, Shane M. Anderson, Patrick A. Lichter, Nicholas A. Giudice, Suneel I. Sheikh, Robert J. Knuesel, Daniel T. Kollmann, and Daniel S. Hedin. 2012. Indoor magnetic navigation for the blind. In *Engineering in Medicine and Biology Society*. 1972.

Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. 2011. ORB: An Efficient Alternative to SIFT or SURF. In *Proceedings of the 2011 International Conference on Computer Vision (ICCV '11)*. IEEE Computer Society, Washington, DC, USA, 2564–2571.

Noah Snavely, Ian Simon, Michael Goesele, Richard Szeliski, and Steven M. Seitz. 2010. Scene Reconstruction and Visualization From Community Photo Collections. *Proc. IEEE* 98, 8, 1370–1390.

Sebastian Tilch and Rainer Mautz. 2013. CLIPS:a camera and laser-based indoor positioning system. *Journal of Location Based Services* 7, 1 (2013), 3–22.

Changchang Wu. 2013. Towards Linear-Time Incremental Structure from Motion. In *International Conference on 3d Vision*. 127–134.

Changchang Wu, S Agarwal, B Curless, and S. M Seitz. 2011. Multicore bundle adjustment. In *Computer Vision and Pattern Recognition*. 3057–3064.

Han Xu, Zheng Yang, Zimu Zhou, Longfei Shangguan, Ke Yi, and Yunhao Liu. 2016. Indoor Localization via Multi-modal Sensing on Smartphones. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 208–219.

Zheng Yang, Chenshu Wu, and Yunhao Liu. 2012. Locating in Fingerprint Space: Wireless Indoor Localization with Little Human Intervention. In *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking (Mobicom '12)*. New York, NY, USA, 269–280.

Moustafa Youssef and Ashok Agrawala. 2005. The Horus WLAN Location Determination System. In *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys '05)*. ACM, New York, NY, USA, 205–218.