2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 4-7 October 2016, Alcalá de Henares, Spain

A Hybrid Method for Indoor Positioning Based on Wireless Network and Smartphone Camera

Jichao Jiao, Fei Li, Zhongliang Deng, Wen Liu
Laboratory of Intelligent Communication, Navigation and Micro/Nano-Systems
Beijing University of Posts and Telecommunications
Beijing, 100876 – China
[e-mail: li-fei@bupt.edu.cn]

***Abstract*—In this paper, we propose a novel smartphone-based indoor localization algorithm by deeply combining wireless signals and images, which leads to improving the localization performance. Although the measured signals are noisy, we demonstrate that the combination of visual and wireless data significantly improves the indoor positioning accuracy. Different with common wireless-based indoor positioning methods, the wireless signals that received in a certain time are transformed into frequency domain, and an image named W-image is created by using our proposed approach. Then, SIFT features that are extracted from the W-image is deeply combined with the LBP features that are extracted from the smartphone camera-based images. Moreover, the hybrid features from the smartphone camera images and images of reference database are matched which are used to determine correspondences indoor positioning points. In order to reduce the computation complexity of our proposed method, the wireless signals are illustrated to estimate the coarse positioning points for reducing the search space of image matching. By leveraging wireless signals and images, we are able to achieve almost a 0.86 in mean average precision and 57.65ms in mean average running time. It can be widely applied to the smartphones installed RGB cameras to offer the high-accuracy location-based services.**

Keywords*—indoor positioning; smartphone camera; wireless signal visualizing; feature matching

I. Introduction

Considering that people spend their majority of time indoors, the demand for an indoor positioning service has also been accelerated. Meanwhile, with the rapid growth in the use of smartphones integrated powerful RGB-camera, they have been efficient platforms for indoor positioning and navigation. However, there are two problems with these monocular views that include: (1) limited computational and memory resources of the smartphone; (2) the high personnel density of users moving in large buildings. Moreover, the indoor positioning based on single-sensor is difficult to be obtained high accuracy result. For one hand, the RGB-based methods accurately locate individuals in the absence of occlusion, but their performance will deteriorate in the crowded scenes. On the other hand, the wireless signal data does not suffer from the occlusion problem during calculating the indoor positioning information. To achieve these challenges, we propose a novel method that complements RGB data with wireless signals emitted by cell phones.

Using a single sensor model for indoor positioning has some shortcoming in accuracy, robustness and adaptive etc. Multi-sensor combination positioning method can make up for the lack of different types of haptic device and get the higher accuracy than the single sensor positioning. In [1], Gallagher and his colleagues employed the particle filter methodology to combine relative motion information based on step detection with wireless signal strength measurements, and the positioning accuracy is more than 5 meters.

In [2], Vintervold presented an integrated camera system/INS algorithm for estimating the position, orientation, liner and angular velocities. This method was performed as part of a larger system, and the estimation results were not optimized. In [3], Gallagher et al. presented a system that ran locally on a mid-range smartphone, and a Kalman filter was used for fusing the information of all the sensors of a smartphone. In [4], the authors proposed an enrichment of RGB data with the wireless signals emitted by personal cell phones. They introduced a novel image-driven representation of wireless data, which embedded all received signals in a single image. Besides, this single image is combined with RGB image for locating and tracking persons within a sparsity-driven framework.

Therefore, we propose a new algorithm to fuse the W-image features and RGB-image features together to estimate the indoor location with high accuracy in the crowded scenes. Fig.1 displays the process diagram of the proposed system.

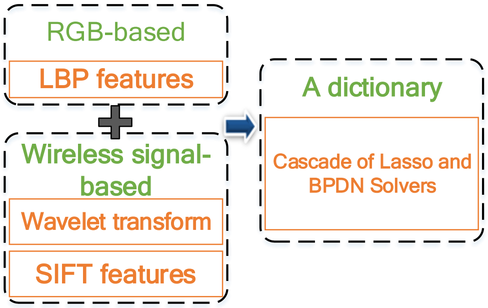

Fig. 1. Process diagram of the proposed system

II. Wireless Data Visualization

In order to locate individuals with the combination of wireless signals data and RGB image data, we have to formulate a relevant representation of the wireless signals to be efficiently fused with RGB data. Therefore, we leverage wavelet to transform the received signal intensity to frequency domain as an image called W-image (Wireless image), then SIFT features are extracted from the W-image.

A sorting method is used to find three signals with the maximum signal strength from all of the received signals. The labeled data is automatically collected when a single person walks around the indoor scene with a smartphone.

A. Received Signals Selection

First, for any individual i with a smartphone, we observe the $RSSI$ streams of wireless hardware inside the area[8]. Then, we calculate the cumulative mean value of $RSSI$ streams of the data points in a period time. Then we can get the following information in a given time frame t in a space:

$$W_i^{(t)} = \{RSS_1, ..., RSS_j, \text{phone MAC}\}^{(t)} \tag{1}$$

where RSS_j is the received signal strength from the j_{th} access point （AP_j）. The maximum strength signal of the three received signals is selected. $[RSS_x^i, RSS_x^j, RSS_x^k]$ is the wireless signal strength at the location x .

B. Wavelet Transform

Our goal is to convert wireless data into an image. In reality, RSS is noisy and anisotropic, and the noise can create some outliers[5] [6]. Therefore, we can receive a curve of data points of the RSSI streams[7] in a period that is 1ms in our paper. One of the reasons for this choice is that the capability of wavelet transform for locating both in time and frequency, and which can be used for noise smoothing[11]. The wavelet transform is introduced as a projection on the basis of the scaled and time-shifted version of the original wavelet [12].

The continuous wavelet transform of a received signal $s(t)$ is defined as follows:

$$CWT(a,\tau) = \int_{+\infty}^{-\infty} s(t)\psi^{*}_{a,\tau}(t)dt \tag{2}$$

where a is the scale factor and $a > 0$, and τ is the translation weight. Haar wavelet is chosen as to be the mother wavelet and it is given by the following function:

$$\psi(t) = \begin{cases} 1 & if \;\; 0 \le t < T/2 \\ -1 & if \;\; T/2 \le t < T \\ 0 & otherwise \end{cases} \tag{3}$$

The main purpose of the mother wavelet is to provide a source function to generate $\Psi_{a,\tau}(t)$ that is shown as follows:

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-\tau}{a}) \tag{4}$$

Finally, we will get the W-image where salient features are extracted.

C. SIFT Features Extraction

In order to fuse information with the RGB image vector for improving the localization performances in both accuracy and speed, the SIFT features are extracted from the W-image [9].

III. RGB Image Feature Extraction

In the proposed method, LBP features are extracted from RGB data [16]. LBP is an excellent texture descriptor for its invariance in gray-scale and rotation. It has been successfully applied to the object detection.

In our paper, we use the $LBP_{8,1}^{2}$ uniform pattern to calculate the histogram of each block. Generally, the notation $LBP_{P,R}^{u}$, which denotes that P sampling points $g_p(p = 0,1,...,P)$ with radius r for each pixel, and the number of 0 and 1 transition is no more than u . The uniformed local binary pattern we used is defined by the following function:

$$U(LBP_{P,R}^{u}) = \sum_{i=0}^{P=1} \left| s(g_i - g_c) - s(g_{i-1} - g_c) \right| \tag{5}$$

$$LBP_{P,R}^{u} = \begin{cases} \sum_{p=0}^{p-1} s(g_p - g_c) & if \;\; U(LBP_{P,R}) \le 2 \\ & otherwise \\ P+1 & \end{cases} \tag{6}$$

where

$$s(t) = \begin{cases} 1 & if \;\; x \ge 0 \\ 0 & if \;\; x < 0 \end{cases}.$$

Then we vote pixels in the block with different bins, and all of the non-uniform patterns are combined into one bin. The number of uniforms is 58, and other models are utilized as a class. Eventually, we can get a 59-D vector.

IV. The Pedestrian Localization Framework

Localization by the combination of two types of features is more powerful than one single method. In this section, we will show how to naturally fuse the W-image and RGB image to infer the ground plane occupancy of individuals indoor scene. We formulate the task as an inverse problem by using a dictionary and a cascade of convex solvers.

A. Problem Formulation

We aim to infer the location of individuals on the ground given LBP features from the RGB as well as the W-image features. We frame this as the best subset selection problem:

$$\arg\min_{x} \|x\|_0 \quad s.t. \quad \mathrm{A}x + n = b \tag{7}$$

where x is the discretized ground plane points, b is the observed data at a given time, A is a dictionary representing for each ground plane point that obtained at the ideal expected observation, and n is the noise. This function is to find a sparse occupancy vector x that can reconstruct the observation b .

B. A Multi-modal Dictionary

We first construct a dictionary A [10], where each atom represents the expected SIFT and LBP feature vector. It is a possible set of "ideal" observation of an individual occupying a ground plane point. $n \times m$ is the size of the dictionary A , where n is the size of an atom and m is the number of ground plane points. For one thing, we know that the RGB images can't match the observed data in the linear operation $\mathrm{A}x$, but the W-images are correctly modeled as a linear operation. So they indeed sum up to match the observed data. Finally, the multi-modal nature of the atoms can locate individuals with better location accuracy. Moreover, the RGB data is sparse, i.e. the person could only take one picture. Therefore, several different RGB images are captured from different orientation at one place, which results in obtaining several atoms (at least four orientation in dictionary A for each ground plane point.

C. Representing the Observation Vector

The observation vector b is the output of LBP feature vector augmented with the SIFT feature vector:

$$b = \left[-S \; -W\right]^T \tag{8}$$

where S is the SIFT feature vector, and W is the W-image feature vector. The sample is shown by Fig. 2.

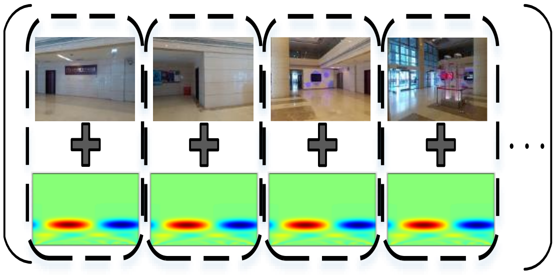

Fig. 2. Illustration of dictionary A. The top row is the RGB image captured by the smartphone, and the bottom row is the W image. Each column is made of the ideal W image observation concatenated with the RGB image for different orientation.

D. Cascade of Lasso and BPDN Solvers

Equation (7) is an NP-hard problem, so we have to relax it by leveraging the multi-modal nature of our data. In this section, the W-image can provide additional prior on the desired solution such as the lower bound of the number of individuals to be located, namely the coarse location. We propose a cascade of solvers. Finally, we change the NP-hard problem into a Basis Pursuit De-Noise (BPDN) problem:

$$x^* = \arg\min_{x} \frac{1}{2}\|b - Ax\|_2^2 + \lambda\|x\|_1 \tag{9}$$

where λ is the trade-off between sparsity level and reconstruction fidelity.

V. Experiments

A. Study Materials and System Configuration

In order to fuse the image data and wireless information to achieve high precision positioning in the crowded environment, the impact of complementing RGB image with wireless signal data for locating people in crowded scenes is studied. We conducted experiments at the New Research Building locating in Beijing University of Posts and Telecommunications (BUPT). First, a dataset of RGB-W vector from crowded scenes indoor is built. Each person with localization request is equipped a smartphone that broadcasts the RSS value to a server. The key technical parameters of the smartphone and APs are shown by TABLE I and TABLE II, respectively.

TABLE I. The key technical parameters

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

TABLE II. The initial condition

Name	Value
APs	4
Sampling period	2.0 sec
The initial RSSI value at one-meter	3.0 dB/m
image resolution	2048 × 2048 pixels
Reference point density	1 m

B. Localization Results

Root mean square error (RMSE) is introduced to evaluate the performance of the proposed algorithm. The positioning accuracy in RMSE can be computed between the real ground positioning and its estimated positioning results. The comparisons of the estimation accuracy are listed in TABLE III.

In TABLE III, we illustrate the performance of the different method. We evaluate the performance of wireless signals by fingerprinting to get more insight on the localization error of the W data. The RGB image method is introduced by the sparsity-driven formulation without the wireless signals. From the comparison results, we can find that the proposed algorithm is improved outperforming the methods of wireless signal-based and RGB image-based. Our positioning algorithm is highly robust and can achieve an accurate estimation with RMSE of 0.86 meters. The comparison result indicates that the fusion method could obtain better accuracy than that based on single-sensor.

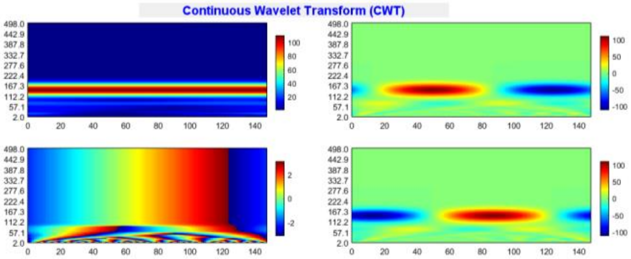

Fig. 3. W image. The figure shows the W image captured by CWT of an RSSI stream.

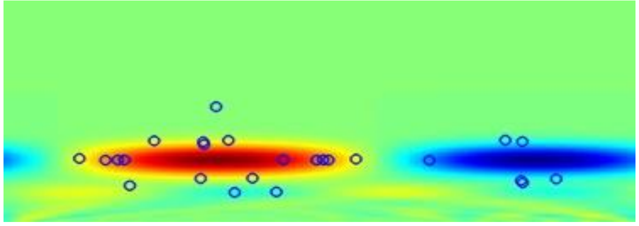

Fig. 4. An example of SIFT features extraction.

We use the percentage of the matching to study the performance of the proposed method in dense environments. During the experiment, several people are moving freely in the scene. According to the TABLE IV, we can find that the proposed method obtained a better performance than the other two methods.

TABLE III. PERFORMANCE COMPARISON IN ACCURACY

Algorithm	**Min error(m)**	**Max error(m)**	**RMSE(m)**	**Running time(ms)**
Our proposed method	0.42	2.83	0.86	57.65
W-based	1.36	5.32	2.16	66.35
RGB only	0.79	3.26	1.56	70.15

TABLE IV. PERFORMANCE COMPARISON IN DENSE AND CROWDED SCENES

Number of people	**Our proposed method(%)**	**W-based(%)**	**RGB only(%)**
3	62.2	59.6	60.5
6	51.3	46.3	49.8
9	40.6	31.8	37.2
12	29.4	21.5	25.1

VI. CONCLUSION

This paper has presented a smartphone-based indoor positioning method. In this algorithm, the vision information and the imaging information of wireless signals are integrated to achieve a high accuracy indoor positioning result in a crowded environment scene for humans. The experimental results showed that visualizing data, which is transformed from the wireless signals, is complementary information for assisting smartphone camera-based indoor positioning. We observed both speed and accuracy improvements over baselines based on current state of the art approaches in both generation and retrieval settings. In addition, the results also showed that the positioning became poor in crowded scene with more than 12 persons. Therefore, the future work that we will do is to investigate on how to improve the performance in crowded scenes liking meeting rooms.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China (NO.61401040), the National High Technology Research and Development Program (“863” Program) of China (NO.2015AA124103), the Beijing University of Posts and Telecommunications Young Special Scientific Research Innovation Plan (2016RC13).

REFERENCES

[1] Hilsenbeck, S., Möller, A., Huitl, R., Schroth, G., Kranz, M., & Steinbach, E. (2012, November). Scale-preserving long-term visual odometry for indoor navigation. In Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on (pp. 1-10). IEEE.

[2] Vintervold Y S. Camera-Based Integrated Indoor Positioning[J]. Remote Sensing of Environment, 2013, 131(8):119-139.

[3] Gallagher T, Wise E, Li B, et al. Indoor positioning system based on sensor fusion for the Blind and Visually Impaired[C]// Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on. IEEE, 2012:1-9.

[4] Alahi A, Haque A, Li F F. RGB-W: When Vision Meets Wireless[C]// IEEE International Conference on Computer Vision. IEEE, 2015.

[5] Lee D L, Chen Q. A model-based WiFi localization method.[C]// 2nf International Conference on Scalable Information Systems, Infoscale 2007, Suzhou, China, June. 2007:1-7.

[6] Tsuda Y, Kong Q, Maekawa T. Detecting and correcting WiFi positioning errors[C]// ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2013:777-786.

[7] Aly H, Youssef M. New insights into wifi-based device-free localization[C]// ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication. 2013:541-548.

[8] Liu H, Yang J, Sidhom S, et al. Accurate WiFi Based Localization for Smartphones Using Peer Assistance[J]. Mobile Computing IEEE Transactions on, 2014, 13(10):2199-2214.

[9] Bicego M, Lagorio A, Grosso E, et al. On the Use of SIFT Features for Face Authentication[C]// Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on. 2006:35.

[10] Gan G, Cheng J. Pedestrian Detection Based on HOG-LBP Feature.[C]// International Conference on Computational Intelligence & Security. 2011:1184-1187.

[11] Mosavi M R, Emamgholipour I. De-noising of GPS Receivers Positioning Data Using Wavelet Transform and Bilateral Filtering[J]. Wireless Personal Communications, 2013, 71(3):2295-2312.

Hassan K, Dayoub I, Hamouda W, et al. Automatic modulation recognition using wavelet transform and neural network[C]// International Conference on Intelligent Transport Systems Telecommunications. 2009:234-238.