IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 16, 2019, accepted June 9, 2019, date of publication June 14, 2019, date of current version July 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2922842

WiLabel: Behavior-Based Room Type Automatic Annotation for Indoor Floorplan

YONGLE CHEN, QINGHUA YAO, DAN YU, AND YULI YANG
College of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China

Corresponding author: Dan Yu (yudan@tyut.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 61401300, and in part by the Natural Science Foundation of Shanxi Province under Grant 201701D111002 and Grant 201601D021074.

ABSTRACT The growing indoor location based services (LBS) applications enhance the requirement of room type annotation. Existing room type annotations are either depending on additional sensors or prone to privacy disclosure. We proposed a method called WiLabel-based on channel state information (CSI) alone. By analyzing the CSI fluctuation, we adopt the percentage of nonzero elements (PEM) algorithm to classify indoor scene and design a zero prior knowledge behavior recognition method to achieve behavior perception in the fewer-person scene, then, design a behavior-based decision tree classifier to determine the room type. The evaluation results from 84 rooms of the college building and mall show that the WiLabel can achieve an average accuracy of 90.5% superior to others.

INDEX TERMS Channel state information, indoor floorplan, room type annotation.

I. INTRODUCTION

The advances in Location Based Services (LBS) technology have enabled the explosive increasing of intelligent indoor applications, such as service robot, checkout-free supermarket and smart home [1]. Meanwhile, an indoor map with room type annotation is the pre-requisite to support and enrich the LBS applications. The traditional methods [2], [3] to generate indoor maps are mostly based on the building floorplan, absence of the specific room type annotation to represent the room type. For the moment, the room type annotation of the indoor floorplan is still an open issue.

In general, there are two basic approaches to addressing room type annotation: manual labeling and crowdsourcing. The manual labeling identifies each room type in person to enrich the electronic map, which is time-consuming and laborious for a large-scale map. The crowdsourcing [4] gathers volunteer data to identify the room type, depending on a large number of participants, and yet involving the user privacy data, such as camera, microphone, GPS. Therefore, both methods are hard to solve efficient the room type annotation problem of indoor floorplan.

We proposed a behavior-based room type annotation method called WiLabel, which can identify indoor room type automatically by wireless signal resource alone. We utilize the channel state information (CSI) [5] to perceive numbers and activities of people [6], [7], and then design a behavior-based decision tree classification model to obtain the corresponding room type. In detail, the main contributions of this article are the following:

- We design a WiLabel room type automatic annotation method only based on CSI without involving additional hardware cost, which achieve an average accuracy of 90.5% in college buildings and malls.
- We adopt a CSI-based percentage of nonzero elements (PEM) algorithm to real-time classify indoor scenes by detecting the number of people and propose a behavior recognition method without prior knowledge according to the wavelet decomposition feature.
- We design a behavior-based decision tree classification model to confirm the room type, which can be automatic constructed by various room types, to be easily expended to other scenes.

II. RELATED WORK

A. ROOM TYPE ANNOTATION

Most of existing methods identify the room types by crowdsourcing various sensor data. Surroundsense [8] is the first system to identify the bookstore, bar and clothing store by crowdsourcing camera, microphone and other sensor data, which can achieve the recognition accuracy of 87%.

The associate editor coordinating the review of this manuscript and approving it for publication was Min Jia.

Similarly, Semsense [9], CSP [10], AutoLabel [11] and some others [12], [13] also utilize crowdsourcing method to gather various sensor data and obtain indoor room types. In addition, SAP [14] utilizes the user's WiFi access point data (user, location, timestamp), and proposed a dual SVM classifier to get room types. Natalia Andrienko et al. samples a fractional locations from the whole track, then utilize the point of interest(POI) and land use(LU) data to cluster the spatial location points and recognize the room type according to the spatio-temporal index [15]. Transitlabel [16] proposes a crowd-sensing system, which utilizes passengers' mobile sensor data to recognize their activities and enrich the indoor floorplan of transit stations. However, the crowdsourcing methods require a considerable number of participants. Meanwhile, multi-sensor strategy also involves additional hardware cost and privacy disclosure. Therefore, we propose a behavior-based method to identify room type only by WiFi signal.

B. BEHAVIOR RECOGNITION

The traditional behavior recognition methods are mostly based on visual recognition and wearable sensors. They will introduce high deployment cost and privacy divulgence, although a fine-grained behavior recognition can be achieved. Nowadays, many works focus on behavior recognition by extracting CSI data from WiFi. Yan Wang et al. proposed an e-eyes system to perceive behavior by calculating the Earth Mover's distance similarity of the CSI amplitude histogram, which can achieve about the accuracy of 96.2% [17]. Wei Wang et al. proposed a method called CARM to calculate log similarity of the CSI action model, which can obtain the accuracy of 96% [18]. Chang et al. train an SVM classifier based on the image features converted by CSI data, to realize the behavior recognition and even identity authentication, which can also obtain a better accuracy of about 90% [19]. Han et al. design Wifall system to identify the fall movement without pre-training stage, but it is only suitable for the individual to determine the fall movement and just obtains the accuracy of 87% [6]. Most of existing methods require pre-training to ensure the recognition of various activities, which will reduce the efficiency of room identification. Otherwise, these methods just recognize some limited scenarios or simple specific actions.

III. WILABEL DESIGN

The WiLabel framework is shown in Figure 1. WiLabel includes two phases: room occupancy classification and room type identification. In room occupancy classification section, we adopt the PEM method to estimate the number of people [20]. Accounting the number of people in the room, we divide the indoor scenarios into none, fewer-person and multi-person scenes. None scene is an empty room. Fewer-person scene is perceived by the CSI behavioral recognition method, while multi-person scene can be perceived through the people traffic and time stamp. In room identification section, we first propose a zero prior knowledge

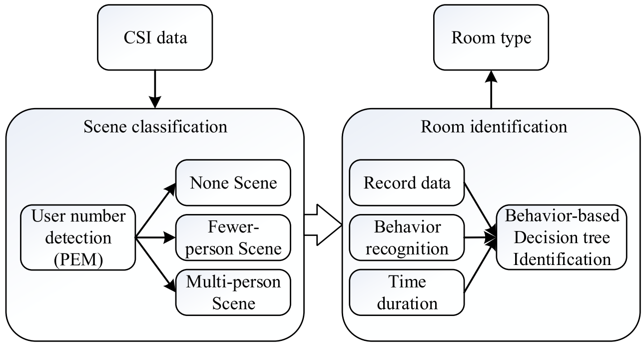


FIGURE 1. WiLabel framework.

behavior recognition algorithm to obtain various behavior attributes, and then design a behavior-based decision tree classification method to classify room type by training the number of users, behavior attribute and time duration.

A. ROOM OCCUPANCY CLASSIFICATION

The pictures in Figure 2 respectively demonstrate the variation of CSI amplitude of 30 subcarriers when 1, 2, 3, 4 people are walking in the room. Figure 2 clearly indicates that CSI values distribute more widely and change more drastic when there are more moving people. If we find a proper quantifiable index to characterize the variation of CSI measurement, it becomes possible to use CSI to classify room occupancy.

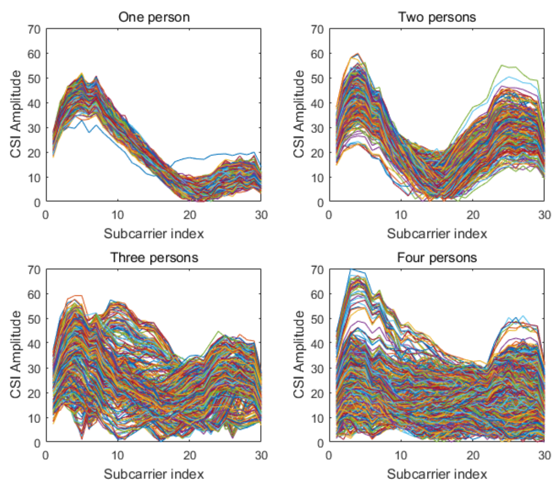


FIGURE 2. CSI amplitude influence with different number of people.

In order to classify the room occupancy with different number of people, we need to design a quantified metric to detect the number of people in the room. In this paper, we adopt the PEM algorithm to solve this problem. Due to the original CSI data is sensitive to the signal fluctuation caused by ambient temperature and air pressure, the noise is inevitable. A moving average filter is added for the amplitude value in matrix. When the amplitude value A=[A1,A2,…, Aj] is computed, Ak is compute by the first t amplitude values, where t is set as an experimental value, no more than the total of CSI subcarrier.

$$A_k = \frac{t * A_k + (t-1) * A_{k-1} + \ldots + A_{k-t+1}}{t + (t-1) + \ldots + 1} \quad (1)$$

Meanwhile, the original PEM method use the data during a duration to compute PEM value, ignoring that the number of people probably have changed. The average PEM value is not accurate. Therefore, we calculate the amplitude values each second as formula (2), where Tr_v is the AP transmit rate.

$$Pro\_data = A\,(1 + (t-1) * \mathrm{Tr\_}\, v : t * \mathrm{Tr\_}\, v, :) \quad (2)$$

Moreover, the original PEM uses a square magnification factor D to amplify the elements around the signal. The square D interior value is set to 1 after normalization. For identifying the behavior accurately, we set the transmitting rate to 700pkt/s [6]. But the amplified elements around the signal will overlap greatly, thus the accuracy of the PEM algorithm will reduce severely. We tune the rectangular magnification factor to 30*10 for reduce the overlap values.

When the signal waveform changes gently, the overlap of elements in matrix will increase which means the less human activities. When the waveform changes severely, the overlap of elements in matrix will reduce which means the more human activities. Therefore, we can easily distinguish the scenarios with different number of people.

B. ROOM TYPE IDENTIFICATION

1) ZERO PRIOR KNOWLEDGE BEHAVIOR RECOGNITION

In this section, a lightweight zero-prior knowledge behavior recognition method is proposed to reduce the cost of training CSI behavior model, which can recognize some basic actions even in two-person scene. In figure 3, our algorithm is divided into two phases, signal processing phase and feature extraction phase.

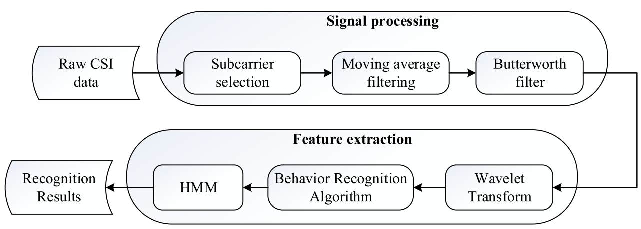


FIGURE 3. Zero prior knowledge behavior recognition flowchart.

a: SIGNAL PROCESSING

SUBCARRIER SELECTION

Due to the frequency diversity of CSI, the 30 subcarriers can be regarded as the different lanes heading to the same destination. The different subcarrier has a different sensitivity to environment variations since each subcarrier has different central frequency and wavelength. Influenced by multipath and shadow problem, CSI has different amplitude in different sensitivity to environment variations, that is, the greater the variance, the more sensitive the subcarrier is to ambient environment. However, the experiment reveal the maximum of variance is probably the error value caused by the overlap during multipath propagation [21], so we will filter out the subcarrier with maximum variance, and select from the second maximum subcarrier.

MOVING AVERAGE FILTER

In fact, the indoor environment contains many interference factors. the original signal includes a multitude of noise and its waveform is rougher, thus we still adopt the filtering method in formula (1) to smooth the waveform.

BUTTERWORTH FILTER

According to the relationship among wavelength, frequency and speed, when the frequency of the router is 5.32GHz, its wavelength is about 0.0564m. Supposing that the cosine function frequency in the channel frequency response (CFR) power is 200Hz [17], the moving speed of the pedestrian is computed as approximately 5.64m/s, which is far over the normal speed of activity. That is because the high-frequency data is often noise effected by a variety of external factors. Therefore, we utilize 9-order Butterworth low-pass filter to eliminate these high-frequency noises. Low pass frequency band ranges in 3Hz to 200Hz. Besides, the stopband characteristics of the Butterworth filter has the slow descent gain, which doesn't make the characteristic of human activity distortion, superior to other low pass filters.

b: FEATURE EXTRACTION

WAVELET TRANSFORM

We analyze the characteristics of different behaviors from time domain and frequency domain. Short Time Fourier Transform and Wavelet Transform are two essential methods to analyze signal time-frequency domain. The former adopts the slide window method to decompose the signal, without analyzing each frequency band, while the latter has better tradeoff in time and frequency resolution. Therefore, we adopt the wavelet transform method in feature extraction phase.

The Butterworth Low-pass filter has filtered the data frequency within 200Hz, however, there are still some noise data in the low frequency data due to its slow descent of stopband characteristics. We adopt the Harr wave to decompose the signal into 12-layer wavelet and filter the noise out by carrying out the feature extraction for each layer. Then, we get the wavelet decomposition coefficient matrix, where the higher coefficient of the decomposition layer represents the higher corresponding frequency. The bottom two images in figure 4 are wavelet decomposition matrices, the darker color represents the larger coefficient. The bottom-left figure shows that the coefficients at the top of 12th layer decomposition is larger, which means that the activity is moving. While the largest coefficient layer in the bottom-right figure is the 9th layer, and the lower coefficient means the static activity.

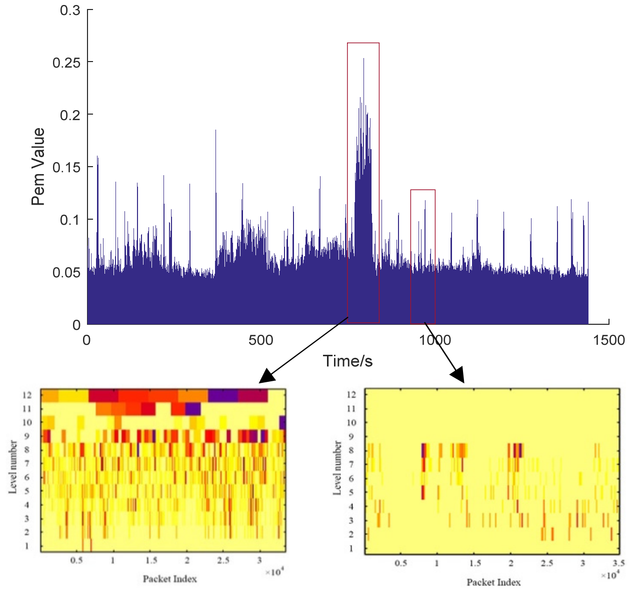


FIGURE 4. Signal processing in the room with single person.

BEHAVIOR RECOGNITION ALGORITHM

According to the unique characteristics of different behavior, we design an algorithm to identify six main actions in daily life, i.e. walking, siting, standing, bowing, upstairs/downstairs and gesture. We design three indicators to characterize the behavior, the highest layer coefficient of wavelet decomposition, action duration, the previous recognition behavior.

Maxlevel is the highest layer coefficient of wavelet decomposition, *Tre* represents *Maxlevel* action duration, Tre_irepresents the time threshold of action i, and$Last_b$ represents the previous action recognition result. In fact, the recognition of upstairs/downstairs is different from the other five actions in the room scene. The upstairs/downstairs actions occur in a separate space without the interference of sitting, standing and other actions. When the highest level of wavelet decomposition is 11 or 12 layers, the other four actions except walking do not occur, and the action duration is greater than the upstairs/downstairs threshold, the output is upstairs/downstairs action, which is identified more accuracy than the other actions due to the unique character of upstairs/downstairs.

When an action characteristic is not inadequate unique. the output of algorithm is unknow. Thus, we design a first-order Hidden Markov Model (HMM) to infer current action by previous recognition behavior. Due to action continuity, each action is only related to the previous action. Therefore, there are just two action states, observation state and hidden state. The probability of next action occurring is inferred by the transfer matrix. Based on the experimental and empirical data, we can obtain the transfer probability of each state, and then establish a state transition matrix to improve the accuracy of the system.

Indeed, we do not have to deal with all the action CSI data due to the high cost. Due to my improved PEM algorithm have distinguished different indoor scenarios, we just deal with the data in fewer-person scene. In figure 4, we collect 1480 seconds of data in a single person scene, then execute the behavior recognition algorithm. According to our behavior recognition algorithm, we can find *Maxlevel* is 12 at the 750 seconds and the identification result is walking as shown in bottom-left figure, while the *Maxlevel* is about 8 at the 1000 seconds and the recognition result is a small gesture as shown in the bottom-right figure.

2) BEHAVIOR-BASED DECISION TREE IDENTIFICATION

In this section, we design a behavior-based decision tree algorithm to identify the room type of room. this is a supervise decision tree, which requires the user to learn the behavior characteristics of different rooms. In figure 5, we divide the behavior into three attributes to construct our decision tree as classifier attribute, i.e. PEM value, time duration and behaviors, where the behavior also includes six actions.

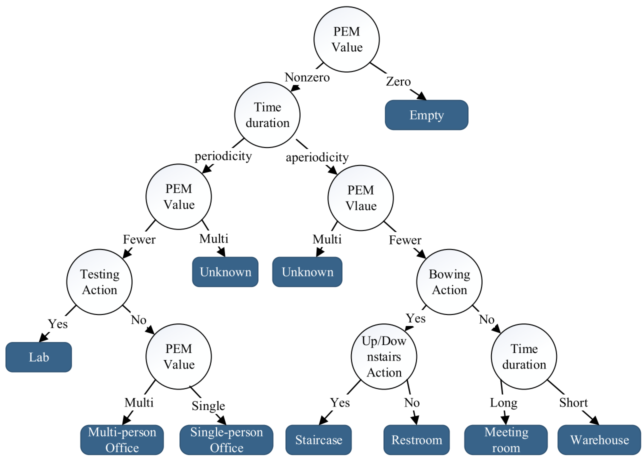


FIGURE 5. Decision tree diagram of our college building.

Before constructing decision tree, we utilize the principal component analysis (PCA) method to extract the main component from six actions to characterize the behavior of each sample room. The information entropy [22] is used to determine how to select the best attribute value [23]. The information entropy can be computed by the formula (3), ε is the sample and ε_i is the probability of thei_{th} attribute appearing. In formula (4), we calculate the information gain by subtracting the conditional information entropy of ε, where A represents the sample attributes.

$$Entropy(\varepsilon) = -\sum_{i=1}^{n} \varepsilon_i \log_2(\varepsilon_i) \tag{3}$$

$$InGa(\varepsilon, A) = Entropy(\varepsilon) - \sum_{j=1}^{J} \frac{|\varepsilon_A|}{\varepsilon} * Entropy(\varepsilon, A) \tag{4}$$

The information gain is not used as the metric to classify the attribute. We use information gain rate to normalize the information gain by formula (5), where SpIn(ε, A) represents

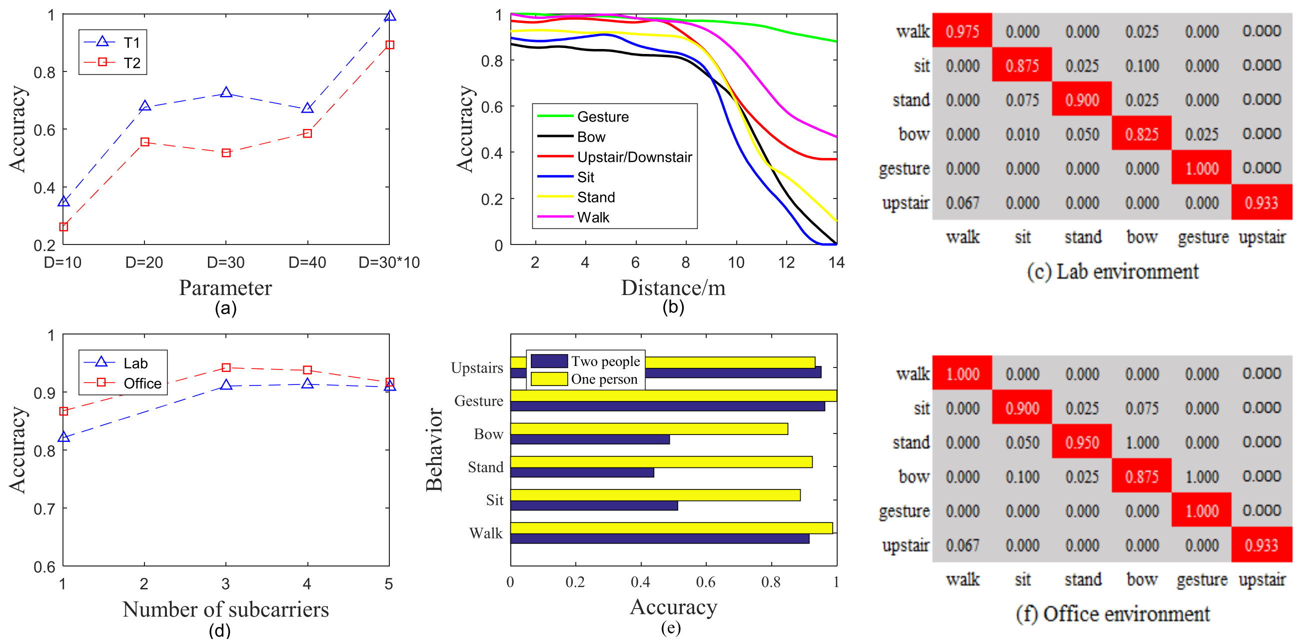


FIGURE 6. **Behavior recognition experiment results. (a) Comparison between raw PEM and our PEM. (b) Impact of distance. (c, f) Recognition accuracy in two different environments. (d) Impact of subcarrier numbers. (e) Performance influence by people number.**

the "splitting information" value [24] in by formula (6). That is an attribute entropy. We decide the attribute to classify decision tree by selecting the maximum value of the information gain rate.

$$GaRa(A) = \frac{InGa(\varepsilon, A)}{SpIn(\varepsilon, A)} \tag{5}$$

$$SpIn(\varepsilon, A) = -\sum_{j=1}^{J} \frac{|\varepsilon_A|}{\varepsilon} * \log_2\left(\frac{|\varepsilon_A|}{\varepsilon}\right) \tag{6}$$

Besides, due to the noise and outliers in sample data, there are few branches of decision tree to reflect the anomaly [25]. The pruning method is proposed to deal with the situation, by judging whether error rate is higher than the error rate before pruning. If the mean number of errors in a subtree is greater than the mean and standard deviation number of errors in the corresponding leaf node by formula (7), the pruning is decided, where *Mean* and *Standard* represent mean value and standard deviation of the number of errors respectively.

$$Mean(false\_leaf) < Mean(false\_tree) + Standard(false\_tree) \tag{7}$$

In figure 5, we divide our work scene into eight kinds of room type. The decision tree is generated by above three attributes, PEM value, time duration and behavior, where we extract the main actions into Testing, Bowing and Up/Downstairs action. The same room is probably recognized to different type at different time. For this purpose, we design an adaptive correction strategy to make our method more robust by setting weight. The deeper level of decision tree layers, the more restrictive conditions are set. Therefore, the weight principle is that the lower level weights are greater than the upper level. For example, a room is recognized as empty at first time, but the second time is recognized as staircase. Thus, the room type should be the staircase instead of the empty. In addition, if the misjudgment occurs at the same level, we adopt the current identification result to achieve self-update annotation of room type.

IV. EVALUATION

In the experiment, we adopt a laptop configured with Intel 5300 NIC to gather the CSI data from Netgear WNDR3700 router by Linux CSI tool. The router transmit rate and transmission frequency is by default set to 700pkt/s and 5Ghz separately. We collect 47 sample rooms in our college buildings and 37 business stores in mall including thirteen room types for evaluation.

A. IMPROVED PEM ALGORITHM PERFORMANCE

In this section, we compare the performance between our improved PEM and the original PEM in figure 6(a). The comparison matric is the classification accuracy. We gather 12 hours sample data in the office building with different number of people. We set the square magnification factor D of original PEM to 10, 20, 30 and 40 separately.

In figure 6(a), T1 demonstrates the classification between none and single person in the room, while T2 demonstrates the classification between fewer-person and multi-person. Intuitively, the latter is more difficult than the former, since it requires a more fine-grained distinguishing of the behavior than confirming if anyone in the room. However, we our

TABLE 1. Room type annotation survey in college buildings.

	Toilet	Single	Multi	Lab	Empty	Staircase	Meeting	Warehouse	Accuracy
Toilet	4	0	0	0	0	0	0	0	1
Single	0	13	1	0	0	0	0	0	0.923
Multi	0	0	4	0	0	0	0	0	1
Lab	0	0	0	8	0	0	0	0	1
Empty	0	0	0	0	4	0	0	0	1
Staircase	0	0	0	0	0	3	0	0	1
Meeting	0	0	0	0	0	0	4	0	1
Warehouse	0	0	0	0	0	0	1	5	0.833
Overall									0.957

improved PEM algorithm still achieve the distinguishing accuracy of 98% and 89% for T1 and T2 separately. Both are much better than the original PEM algorithm. This is because the original PEM algorithm lacks the initial smooth processing, just utilizing a fixed square magnification factor to amplify the signal.

B. BEHVAIOR RECOGNITION PERFORMANCE

We compare the performance of our behavior recognition algorithm with other methods in the fewer-person scene. This is because we hard to recognition the characteristic of CSI variety in multi-person scene as shown in figure 2. We first analyze the effect of distance away from AP on the behavior recognition accuracy in figure 6(b). We find that the best distance range for behavior recognition is 1m-8m. When the distance is over 8m, the behavior identification accuracy will decline sharply due to rapid signal attenuation characteristics. In fact, the length of most rooms is not more than 8m, thus we can obtain the better behavior identification accuracy. Once people exceed this range estimated by the received single strength (RSS) value, we will discard these results.

In the single-person scene, we collect 380 actions in the office and lab to verify the performance of our behavior recognition algorithm in figure 6(c,f), where the accuracy of behavior recognition is about 93%, especially the accuracy of walking and gesture is higher. This is because above two actions have unique wavelet decomposition images, and the wavelet decomposition layer are larger than the others. Besides, the three actions of siting, bowing and standing have a higher misjudgment rate due to on the approximate layer in wavelet decomposition. Through computing the duration of the action, we still obtain a better accuracy to identify the behavior.

In the two-person scene, we collect 300 samples to analyze the identification performance in the office and laboratory. In figure 6(e), we find the walking, gesture and upstairs still have a high recognition rate, while the average recognition accuracy rate of sitting, standing and bowing actions is less than 50%. In summary, our method still obtains an accurate behavior recognition without prior knowledge in some scenarios.

Algorithm 1 Behavior Recognition Algorithm

```
Input: Wavelet decomposition matrix
Output: Behavior
IF Maxlevel==12 and Tre>Trewalk and Lastb!=sit:
        THEN print walk, break;
 ELIF Maxlevel<=8 and Tre>Tregesture:
        THEN print gesture, break;
 ELIF 10<=Maxlevel<=11 and Tre<Tresit
  and Lastb!=sit:
        THEN print sit, break;
 ELIF 10<=Maxlevel<=11 and Tre<Trestand and
  Lastb!=walk or stand:
        THEN print stand, break;
 ELIF9<=Maxlevel==11 and Tre>Trebow
  and Lastb!=bow or sit:
        THEN print bow, break;
 ELSE:
        print unknown
IFMaxlevel==12 and Tre>Treupstairs and
  Lastb!=sit or bow or stand:
        THEN print upstairs;
```

C. WILABEL PERFORMANCE

We evaluate WiLabel by the CSI data of 47 sample rooms in our college buildings and 37 sample rooms in shopping malls around our school, where there are thirteen room types: single people offices, multiple-person offices, laboratories, toilet, empty rooms, staircases, meeting rooms, warehouse in college buildings and restaurant, coffee/bakery, clothing, entertainment, department store in shopping malls. The reasons we separate the coffee shop from the restaurant is mainly because that there are different crowds and scene features. We found through experiments that the data collected within 3 days, the recognition results are constantly revised, and when the amount of data reaches 3 days, the recognition results converge with great probability, and the data collected more days are basically similar to the 3 days, due to consideration We chose to collect for 3 days to the cost issue. The identification results of the rooms in the malls and the college buildings can be seen in Tables 1 and 2, respectively. The overall recognition average accuracy achieves about 90.5%, superior

TABLE 2. Room type annotation survey in malls.

	Restaurant	Coffee/Bakery	Clothing	Entertainment	Department Store	Accuracy
Restaurant	7	0	0	0	0	1
Coffee/Bakery	0	8	0	0	2	0.8
Clothing (e.g. shoe/cosmetic store)	0	0	8	1	0	0.889
Entertainment (e.g. cinema/theatre)	0	0	1	3	0	0.75
Department Store (e.g. grocery/book store)	0	1	0	1	5	0.714
Overall						0.838

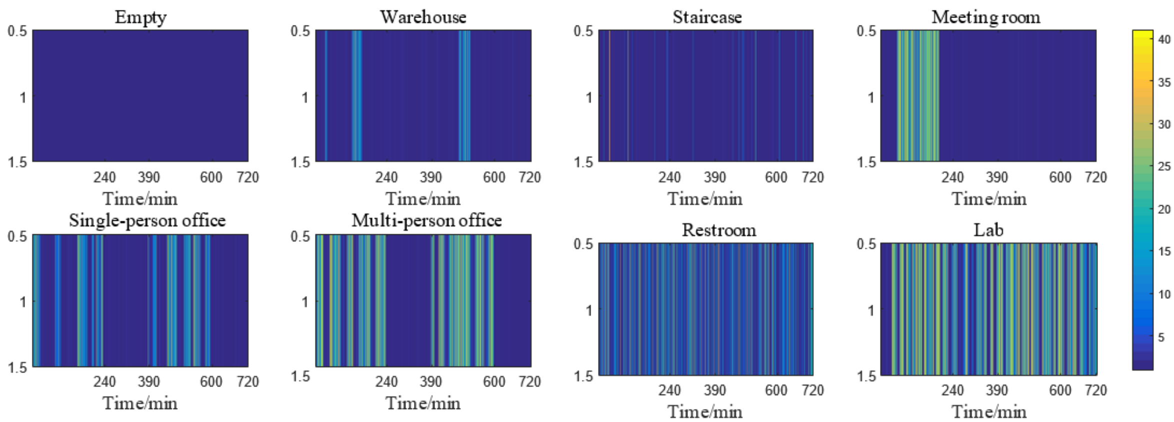


FIGURE 7. PEM values over time.

to the existing Semsense(87%), Surroundsense(87%), Autolabel(87%). In college buildings, we find the laboratory, offices, toilet and staircases are identified more accurate than others. There are just two errors in the results. One is a single-person office is identified to a multiple-person office, which is because there are external people mixed in the office. The other is a warehouse is recognized to a meeting room by mistake, which is because there are many people to stay longer in the room during the identification period. With the time duration increasing, the characteristics of both rooms become more and more distinctive. In shopping malls there are 6 errors in the experiments, part of them is that coffee/bakery shop is identified as a grocery store. It is because some convenience stores have similar environmental characteristics with coffee/bakery shop, it is impossible to clearly define the type of operation of each grocery store. The other part is the clothing store identification for entertainment venues, which may be the fact that some game halls have activity indicators similar to those of clothing stores.

The PEM values of all room types in college buildings within half a day are shown in figure 7. The x-axis represents the period from 8AM to 8PM, where color bar is the PEM value. The bigger the PEM value means the higher the brightness. We also perceive the difference with various amount of people in rooms from figure 7. For example, we can find the PEM values are more concentrated in the morning in figure 7(4), but the other time does not occur the PEM value. Meanwhile, the time duration is longer. Thus, our decision tree method can classify it to a meeting room.

D. AP NUMBER INFLUENCE

In this section, we analyze the AP number influence on scene classification, behavior recognition and room identification in figure 8. With the increasing of AP numbers, the accuracy also has an increasing trend. Because the increasing of AP number represents the increasing of the amount of obtained CSI data. The richer CSI data will make the identification

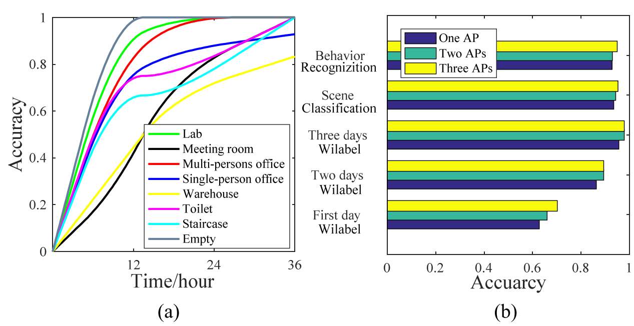


FIGURE 8. Performance with different AP numbers. (a) Accuracy in three days. (b) Impact of AP numbers.

TABLE 3. Comparison with other methods.

Method	Infrastructure	Dataset	Accuracy
Surroundsense[8]	camera、microphone, accelerometer	51 business stores in cluster	87%
Semsense[9]	camera、microphone、Wifi、GPS et al	711 stores in mall in two cities	87%
CSP[10]	camera、microphone、Wifi et al	1241 rooms of various type in 5 cities	69%
AutoLabel[11]	camera、GPS、online data	40 business stores in malls or along street	87%
SAP[14]	Wifi check-in、User Data	Whrrl datasets: 53432 rooms of 199 tags	83%
Transitlabel[16]	accelerometer、Gyroscope、GPS、dead-reckon	8 stations in 2 cities	91%
SISE[26]	camera、inertial data、Wifi	COCO datasets: 152 rooms	80%
WiLabel	Wifi	84 rooms in college buildings or malls	90. 5%

more accurate. But the experimental results also illustrate that the performance gains from AP numbers are limited. Because the most of room size are not over 8m, a single AP also obtain a better performance. Moreover, multiple APs mean more device to deploy. It is not cost-efficient.

E. COMPARISON WITH OTHERS

In this section, we can see in Table 3 the details of WiLabel compared to other existing room type identification methods. Due to different facilities and environments, other methods cannot be verified in the WiLabel scenario, and the accuracy listed in the table is the accuracy rate claimed in the paper. As can be seen from the table, in terms of infrastructure, other methods require more user-side facilities, such as camera, GPS, Wifi, etc., and WiLabel can identify the type of room only by Wifi, which is superior to other methods in power consumption. In addition, due to privacy violations, collecting data using camera or microphone is not allowed in some specific scenarios, and there are limitations in some scenarios. Wifi is basically a public facility, although it can identify the number of people and behavior in the room, it does not pose great threat to privacy. In terms of datasets, only CSP and SAP involve various types of rooms, while several other method scenarios are relatively simple. Autolabel, Surroundsense and Transitlabel only have a shopping mall or station scenario. There are two scenarios of shopping mall and college building in WiLabel. In terms of accuracy, Transitlabel has a high accuracy rate of 91%, while other methods are generally 80%. WiLabel has achieved an accuracy of about 90% through the verification of 84 rooms.

V. CONCLUSIONS

In this paper, we design a behavior-based room type annotation algorithm, Wilabel, to identify the room type. We adopt a CSI-based PEM method to distinguish the number of people in rooms and propose a zero prior knowledge behavior recognition algorithm to obtain behavior attribute, then design a behavior-based decision tree classification model to obtain various room types. The experiment shows we can achieve the average accuracy of 90.5% superior to other methods. However, we still hard to achieve high-accuracy room identification for all scenes only based on CSI information. Because the CSI-based behavior detection of multi-person is hard to obtain the higher accuracy. This is our future work.

REFERENCES

[1] L. Chen, C. D. Nugent, and H. Wang, "A knowledge-driven approach to activity recognition in smart homes," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 6, pp. 961–974, Jun. 2012.

[2] Y. Jiang, Y. Xiang, X. Pan, K. Li, Q. Lv, R. P. Dick, L. Shang, and M. Hannigan, "Hallway based automatic indoor floorplan construction using room fingerprints," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Zurich, Switzerland, Sep. 2013, pp. 315–324.

[3] S. Ma, Q. Liu, and H. Tang, "An overview of location semantics technologies and applications," *Int. J. Semantic Comput.*, vol. 9, no. 3, pp. 373–393, Sep. 2015.

[4] R. Gao, M. Zhao, T. Ye, F. Ye, Y. Wang, K. Bian, T. Wang, and X. Li, "Jigsaw: Indoor floor plan reconstruction via mobile crowdsensing," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, Maui, HI, USA, Sep. 2014, pp. 249–260.

[5] E. Perahia, A. Sheth, T. Kenney, R. Stacey, and D. Halperin, "Investigation into the Doppler component of the IEEE 802.11n channel model," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2010, pp. 1–5.

[6] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-free fall detection by wireless networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 581–594, Feb. 2017.

[7] M. Zhao, F. Adib, and D. Katabi, "Emotion recognition using wireless signals," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, New York, NY, USA, Oct. 2016, pp. 95–108.

[8] M. Azizyan, I. Constandache, and R. R. Choudhury, "SurroundSense: Mobile phone localization via ambience fingerprinting," in *Proc. 15th Annu. Int. Conf. Mobile Comput. Netw.*, Beijing, China, Sep. 2009, pp. 261–272.

[9] M. Elhamshary, and M. Youssef, "SemSense: Automatic construction of semantic indoor floorplans," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Banff, AB, Canada, Oct. 2015, pp. 1–11.

[10] Y. Chon, N. D. Lane, F. Li, H. Cha, and F. Zhao, "Automatically characterizing places with opportunistic crowdsensing using smartphones," in *Proc. ACM Conf. Ubiquitous Comput.*, Pittsburgh, PA, USA, Sep. 2012, pp. 481–490.

[11] R. Meng, S. Shen, R. R. Choudhury, and S. Nelakuditi, "AutoLabel: Labeling places from pictures and websites," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Berlin, Germany, Sep. 2016, pp. 1159–1169.

[12] D. Philipp, P. Baier, C. Dibak, F. Dürr, K. Rothermel, S. Becker, M. Peter, and D. Fritsch, "MapGENIE: Grammar-enhanced indoor map construction from crowd-sourced data," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, Budapest, Hungary, Mar. 2014, pp. 139–147.

[13] H. Aly, A. Basalamah, and M. Youssef, "Map++: A crowd-sensing system for automatic map semantics identification," in *Proc. 11th Annu. IEEE Int. Conf. Sens., Commun., Netw.*, Singapore, Jun. 2014, pp. 546–554.

[14] M. Ye, D. Shou, W.-C. Lee, P. Yin, and K. Janowicz, "On the semantic annotation of places in location-based social networks," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Diego, CA, USA, Aug. 2011, pp. 520–528.

[15] N. Andrienko, G. Andrienko, G. Fuchs, and P. Jankowski, "Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces," *Inf. Vis.*, vol. 15, no. 2, pp. 117–153, Apr. 2016.

[16] M. Elhamshary, M. Youssef, A. Uchiyama, H. Yamaguchi, and T. Higashino, "TransitLabel: A crowd-sensing system for automatic labeling of transit stations semantics," in *Proc. 14th Annu. Int. Conf. Mobile Syst., Appl., Services*, Singapore, Jun. 2016, pp. 193–206.

[17] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, Maui, HI, USA, Sep. 2014, pp. 617–628.

[18] W. Wang, A. X. Liu M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Paris, France, Sep. 2015, pp. 65–76.

[19] J.-Y. Chang, K.-Y. Lee, K. C.-J. Lin, and W. Hsu, "WiFi action recognition via vision-based methods," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 2782–2786.

[20] W. Xi, J. Zhao, X.-Y. Li, K. Zhao, S. Tang, X. Liu, and Z. Jiang, "Electronic frog eye: Counting crowd using WiFi," in *Proc. IEEE Conf. Comput. Commun.*, Toronto, ON, Canada, Apr. 2014, pp. 361–369.

[21] Y. Chen, X. Wang, D. Yu, Y. Yang, and J. Chen, "Localizing access point through simple Gesture," *IEEE Access*, vol. 6, pp. 38870–38880, 2018.

[22] B. J. Strait and T. G. Dewey, "The Shannon information entropy of protein sequences," *Biophys. J.*, vol. 71, no. 1, pp. 148–155, Jul. 1996.

[23] V. Sugumaran, V. Muralidharan, and K. I. Ramachandran, "Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 930–942, Feb. 2007.

[24] H. Schmid, "Probabilistic part-of-speech tagging using decision trees," in *Proc. Int. Conf. New Methods Lang. Process.*, Manchester, U.K., 1994.

[25] R. E. Banfield, L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer, "A comparison of decision tree ensemble creation techniques," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 173–180, Jan. 2007.

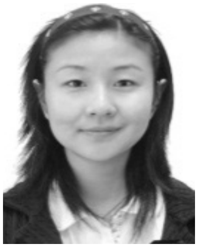
[26] X. Teng, D. Guo, Y. Guo, X. Zhao, and Z. Liu, "SISE: Self-updating of indoor semantic floorplans for general entities," *IEEE Trans. Mobile Comput.*, vol. 17, no. 11, pp. 2646–2659, Nov. 2018.

YONGLE CHEN was born in Weifang, Shandong, China, in 1983. He received the B.S. degree in computer science from Jilin University, in 2007, the M.S. degree in computer science from the Institute of Software, Chinese Academy of Science, in 2009, and the Ph.D. degree in computer science from the University of Chinese Academy of Sciences, in 2013.

From 2013 to 2015, he was an Assistant Professor with the College of Information and Computer, Taiyuan University of Technology, Taiyuan, China. Since 2015, he has been an Associate Professor. His research interests include wireless sensor networks, indoor positioning, and the IoT security.

QINGHUA YAO was born in Jincheng, Shanxi, China, in 1996. He received the B.S. degree in business management from Shanxi University, Taiyuan, China, in 2016. He is currently pursuing the master's degree with the College of Information and Computer, Taiyuan University of Technology.

His research interests include indoor positioning and the Internet of Things (IoT).



DAN YU was born in Taiyuan, Shanxi, China, in 1983. She received the B.S. degree in electronic engineering from the North University of China, in 2007, and the M.S. degree in electronic engineering from the Beijing University of Posts and Telecommunications, in 2013. She is currently pursuing the Ph.D. degree with the College of Information and Computer, Taiyuan University of Technology, Taiyuan, China.

Her research interests include wireless sensor networks and the Internet of Things (IoT).

YULI YANG was born in Yicheng, Shanxi, China, in 1979. She received the M.S. degree in computer science and technology from Guangxi Normal University, China, in 2007, and the Ph.D. degree in computer science and technology from the Taiyuan University of Technology, Taiyuan, China, in 2015.

She is currently a Lecturer with the College of Information and Computer, Taiyuan University of Technology. Her research interests include related with computer network security, cloud computing, and trust management.