AtLAS: An Activity-Based Indoor Localization and Semantic Labeling Mechanism for Residences

Xiaoguang Niu*, Member, IEEE, Luyao Xie, Jiawei Wang, Haiming Chen*, Member, IEEE, Dandan Liu, and Ruizhi Chen*

Abstract—Currently, indoor localization technology and indoor location-based services are becoming increasingly important in the area of mobile and ubiquitous computing. However, the design of an indoor location-based system confronts two challenges: 1) achieving high-precision location recognition and 2) identifying what indoor objects actually are (which is called semantic labeling). In this article, we propose AtLAS, an activity-based indoor localization and semantic labeling mechanism. The key idea is that some objects in an indoor environment, such as doors and toilets, determine predictable human behaviors in small areas, which can be reflected in unique sensor readings. AtLAS leverages this idea to determine a user’s accurate location by identifying users’ activities. Furthermore, we leverage the topological structure of indoor objects to mine the semantic knowledge and label the objects through gained knowledge automatically. To the best of our knowledge, AtLAS is the first attempt to build a system that leverages users’ activities to conduct a high-precision indoor localization and semantic labeling system for the case of residences. The experimental results show that AtLAS can achieve a median localization accuracy of 0.57 m, and the system can localize the landmarks with a median accuracy of 0.43 m on average without 5% worst errors. AtLAS can label the objects semantically with a 5.7% false-positive rate and a 5.8% false-negative rate on average.

Index Terms—Action recognition, indoor positioning, location-based services, mobile sensing, semantic labeling.

I. INTRODUCTION

With the rapid development of smart home technologies, a number of applications require users’ high-precision indoor location information and further semantic knowledge corresponding to the location, objects, and activities, such as home-care applications for the old. Additionally, the operation of localization and acquisition of semantic knowledge should not interfere with the users’ activities and disclose any of their privacy concern. Such localization technology should be low cost and should not require any professional knowledge for usage, especially for the aged using smart home-care applications.

Both industry and academia currently pay much attention to indoor location-based services (LBSs) and make every effort to improve positioning accuracy. Many novel methods have been developed; for example, some indoor localization applications locate users by using their relative locations to environmental physical features [1], [2]. WiFi and Bluetooth-based localization systems offer ubiquitous localization [3]–[7]. Some systems utilize inertial sensors to detect landmarks, such as an elevator and an escalator, which are used for calibrating errors of dead reckoning [8]–[10]. In addition, Google provided an indoor localization service, visual positioning service (VPS), utilizing visual information to locate users with high precision in 2017. However, there are still some limitations and problems in the aforementioned methods. For example, the accuracy of positioning is not high enough or numerous extra equipment is needed. In addition, it is also necessary to take user privacy into account.

An indoor map with semantic knowledge is also crucial for indoor LBSs to improve positioning accuracy and further understand human behaviors. The combination of both types of information, such as location and human behaviors, will enable more intelligent location-based services. Realizing the value of this technology, some systems have been developed to generate an indoor map with semantic labels by designing a unique template for each landmark in advance [10], [11]. For instance, TransitLabel [10] created labels of drink vending machines using the emitted sounds when the machines are dispensing drinks to a customer. Current technologies are only applicable for large-scale buildings, such as shopping malls, but not for residences. First, the precision of WiFi and Bluetooth-based localization systems is typically 2–5 m, which is not adequate for smart home applications. Second, the positioning accuracy of these landmark-based indoor positioning methods mainly depend on the density of landmarks and precision.
of landmark location. However, these methods cannot detect enough landmarks or ensure the precision of landmark location in residences. Third, sound-based and vision-based methods easily suffer from the disturbance of surroundings, and it is nearly impossible for everyone in a family to be equipped with a depth camera at every moment when VPS is used.

In this study, through many experiments and analyses, we determined that human behaviors at home exemplify particular patterns when people are close to or pass through certain objects in a residence. These behaviors commonly occur in small areas. Such areas can be considered landmarks to reset accumulative errors caused by dead reckoning. Therefore, human daily activities associated with certain objects can be used for indoor localization. In addition, some action patterns and some relationships among objects can reflect the semantics of objects when users are moving in an indoor environment. For example, actions occurring at toilets always contain “standing” and “sitting,” and there is a frequent connection between a toilet and a sink because people usually wash hands after using the toilet. By mining the semantics of objects from action patterns and the topological structure among objects, indoor maps can be labeled accurately. These activities recognition and semantic knowledge mining tasks can be conducted using inertial sensors, including an accelerometer, a gyroscope, and an orientation sensor in wearable devices, and operate without users’ attention.

Through the observations above, we propose AtLAS, an activity-based indoor localization and semantic labeling mechanism using a wearable device. The core idea is that when a person conducts a specific activity (e.g., sitting on a chair), the movement of the body comprises a unique sequence of motion features, which can be observed using a set of inertial sensors. AtLAS extracts a set of features from inertial sensor data and then recognizes the users’ behavior patterns. When users who perform the same pattern of behavior are within a small area and their orientations are the same, the small area can be considered a landmark. Then the landmarks are associated with PDR to reconstruct the spatial topological structure of the objects inside the apartment. Finally, a set of indoor objects can be labeled semantically according to the topological structure among indoor objects and the features of the object itself. Specifically, a semantic landmark is denoted by several attributes: its specific pattern of behavior, WiFi fingerprint, orientation, and relative position topology with other landmarks. To the best of our knowledge, this study is the first attempt to leverage the knowledge of human behavior patterns for indoor localization and semantic labeling synchronously in residences.

An evaluation of AtLAS shows that the system can achieve median user localization errors of 0.57 m and median landmark localization errors of 0.43 m after removing 5% worst experimental results. The system has a higher detection rate of landmarks than the other systems in residences. In addition, it can label the objects accurately, with a 5.7% false-positive rate and a 5.8% false-negative rate on average.

In summary, our contributions can be summarized as follows.

1) We conduct the first feasibility study to utilize the user’s action sequences, which inherently occur in a small area inside residences, to determine a user’s location with the accuracy of half a meter.

2) We recognize a set of objects in a residence and mine their semantic knowledge using the combined data of multiple sources. These objects and their semantic knowledge are used for labeling an indoor map of the residence, which are significant for the development of intelligent location-based services.

3) AtLAS is implemented on wearable devices in two residences having different layout. 20 participants collect real-time data to quantify its precision and robustness.

The remainder of this article is organized as follows. We present the architecture of the AtLAS system in Section II. Section III provides the details of the AtLAS system. We evaluate the performance of AtLAS in Section IV. Related work is discussed in Section V. Finally, our study is concluded in Section VI.

II. DESIGN OVERVIEW

The purpose of AtLAS is to locate users by analyzing the massive amount of sensor data of users’ daily lives and attaching labels to indoor objects in residences. AtLAS system consists of four main units: 1) sensor data collection; 2) users’ patterns of behavior identification; 3) landmarks detection; and 4) localization and semantic labeling. As shown in Fig. 1, sensor data are collected and preprocessed first. Then low-level actions can be identified from sensor data and high-level patterns of behavior can be recognized from action sequences. So the results and some other information will be used to detect landmarks for localization and semantic labeling. In the rest of this section, we provide an overview of the system architecture and postpone the details until Section III.

A. Sensor Data Collection

The data collection is based on a smartphone or wearable device. Generally speaking, using commercial off-the-shelf
TABLE I
TYPE OF BASIC ACTIONS

<table>
<thead>
<tr>
<th>State</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>a₁</td>
<td>Walking</td>
</tr>
<tr>
<td>a₂</td>
<td>Sitting down</td>
</tr>
<tr>
<td>a₃</td>
<td>Standing up</td>
</tr>
<tr>
<td>a₄</td>
<td>Slight turning left (heading change: −90°&lt;θ&lt;0°)</td>
</tr>
<tr>
<td>a₅</td>
<td>Slight turning right (heading change: 0°&lt;θ&lt;90°)</td>
</tr>
<tr>
<td>a₆</td>
<td>Sharp turning left (heading change: −180°&lt;θ&lt;−90°)</td>
</tr>
<tr>
<td>a₇</td>
<td>Sharp turning right (heading change: 90°&lt;θ&lt;180°)</td>
</tr>
<tr>
<td>a₈</td>
<td>Slight stooping (orientation: 0°&lt;λ&lt;45°)</td>
</tr>
<tr>
<td>a₉</td>
<td>Squat</td>
</tr>
<tr>
<td>a₁₀</td>
<td>Slight reclining (orientation: −45°&lt;λ&lt;0°)</td>
</tr>
<tr>
<td>a₁₁</td>
<td>Sharp reclining (orientation: −90°&lt;λ&lt;−45°)</td>
</tr>
<tr>
<td>a₁₂</td>
<td>Pause</td>
</tr>
</tbody>
</table>

TABLE II
TYPE OF PATTERN OF BEHAVIOR

<table>
<thead>
<tr>
<th>Category</th>
<th>State</th>
<th>Definition</th>
<th>Standard template</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walk-through(WT)</td>
<td>w₁</td>
<td>Left axis, push</td>
<td>a₁ a₂ a₃ a₄ a₅</td>
</tr>
<tr>
<td></td>
<td>w₂</td>
<td>Right axis, push</td>
<td>a₁ a₂ a₃ a₄ a₅</td>
</tr>
<tr>
<td></td>
<td>w₃</td>
<td>Left axis, pull</td>
<td>a₁ a₂ a₃ a₄ a₅</td>
</tr>
<tr>
<td></td>
<td>w₄</td>
<td>Right axis, pull</td>
<td>a₁ a₂ a₃ a₄ a₅</td>
</tr>
<tr>
<td>Goal-oriented(GO)</td>
<td>g₁</td>
<td>Sitting</td>
<td>a₁ a₂ a₃ a₄ a₅ a₆ a₇ a₈</td>
</tr>
<tr>
<td></td>
<td>g₂</td>
<td>Standing</td>
<td>a₁ a₂ a₃ a₄ a₅ a₆ a₇ a₈</td>
</tr>
<tr>
<td></td>
<td>g₃</td>
<td>Reclining</td>
<td>a₁ a₂ a₃ a₄ a₅ a₆ a₇ a₈</td>
</tr>
<tr>
<td></td>
<td>g₄</td>
<td>Lying</td>
<td>a₁ a₂ a₃ a₄ a₅ a₆ a₇ a₈</td>
</tr>
<tr>
<td></td>
<td>g₅</td>
<td>Opening</td>
<td>a₁ a₂ a₃ a₄ a₅ a₆ a₇ a₈</td>
</tr>
<tr>
<td></td>
<td>g₆</td>
<td>Stooping</td>
<td>a₁ a₂ a₃ a₄ a₅ a₆ a₇ a₈</td>
</tr>
</tbody>
</table>
have a confidence level. Only the result with the confidence level that exceeds the threshold can be used as the recognition result. Table I shows twelve types of basic actions as $A = \{a_i| 1 \sim 12\}$. After recognizing basic actions using sensors data, we can obtain a sequence of actions.

2) **High-Level Patterns of Behavior Recognition:** Fig. 4 shows the sensor readings of an action sequence. According to the marks in Fig. 4, we can determine two subsequences: 1) walking, turning left, sitting down, standing up, turning right, and walking and 2) walking, stooping, and walking. According to Table II, these two subsequences can be recognized as two patterns of behavior: 1) sitting and 2) stooping. Considering the variants of action sequences, we apply fuzzy pattern recognition to reduce the influence of variants on the action sequence. In this article, the patterns of behavior include two categories: 1) goal oriented (GO) and 2) walk-through (WT). Patterns of behavior in GO mean going to a certain place with a purpose; the WT type in an indoor environment only includes the patterns of walking through a door. Before identification, we need to extract the action sequences occurring most often in users’ activities from the massive amount of data and adjust them to the standard templates. Each pattern of behavior corresponds to a standard template.

Table II shows the patterns of behavior $\{\omega_j|1 \sim 10\}$ of WT and GO based on Table I. WT only includes the patterns of walking through a door. Because the patterns of walking through a door depend on two axes of the door and pushing or pulling the door, WT has four patterns: 1) left axis and push; 2) right axis and push; 3) left axis and pull; and 4) right axis and pull. In GO, patterns of behavior include sitting, standing, reclining, lying, opening, and stooping. A pattern of behavior corresponds to one or more specific behaviors. For example, the pattern of opening includes opening a refrigerator, opening a wardrobe, etc.

In a standard template, we define the actions before the action $a_i$ as precursors of $a_i$ and the actions after $a_i$ as successors of $a_i$. For example, in the standard template of pattern $\omega_1$, the precursors of action $a_8$ are $a_1$ and $a_{12}$; the successors are $a_5$ and $a_1$. For action set $A = \{a_i|1 \sim 12\}$, $\omega_j(a_i)$ denotes the degree of membership that action $a_i$ belongs to the pattern of behavior $\omega_j$. The range of $\omega_j(a_i)$ is from 0 to 1. The closer the $\omega_j(a_i)$ is to 1, the higher the degree of membership will be.

Fig. 5 shows the process of identification. In Fig. 5, the action sequence is abcdedegadhec. We define three patterns of behavior: 1) $\omega_A = abcde$; 2) $\omega_B = adec$; and 3) $\omega_C = acefd$. First, we use the action of walking to split the action sequence into two subsequences, because after users finish the current pattern of behavior, they need to walk to the next place for the next pattern of behavior. Then, for each subsequence, the degree of membership of each action is computed. The degree of membership function is provided in the following equation:

$$\omega_j(a_i) = \frac{p_{ai} + q_{ai}}{n_j - 1}. \quad (1)$$

In the above equation, $p_{ai}$ and $q_{ai}$ denote the number of precursors of $a_i$ and the number of successors of $a_i$, respectively. $n_j$ denotes the number of actions in the standard template $\omega_j$. In Fig. 5, for example, the action $c$ of $X_{B1}$ has one precursor
and no successor, and the number of actions in $\omega_B$ is four, so its degree of membership to $\omega_B$ is 0.33. Note that the degree of membership of the first action is 1, specifically.

Finally, we compute the probabilities that this subsequence belongs to corresponding patterns of behavior. The following equation shows the calculation of the probability:

$$\sigma(X_j) = \frac{1}{n_j} \sum_{k=1}^{n_j} \omega_\ell(a_k).$$  \hspace{1cm} (2)

In (2), $n_j$ denotes the number of actions in the standard template $\omega_j$. We set a constant value $\varepsilon$ to evaluate the probabilities. If $\sigma(X_i) = \max_{j=1}^{n} \{\sigma(X_j)\} > \varepsilon$, we consider the subsequence as a pattern of behavior $\omega_i$. Two subsequences in Fig. 5 are identified as patterns of behavior $\omega_A$ and $\omega_B$, respectively. Obviously, the variants of extra actions $f$ and $h$ in Fig. 5 have no influence on the patterns of behavior identification.

### C. Landmarks Detection

Most indoor objects can force users to conduct specific patterns of behavior in a small area. Such an area is called an activity area, e.g., the area in front of a toilet. During a pattern of behavior, its WiFi fingerprint and the orientation with the longest duration can also be obtained. The orientation is classified into four groups according to the $z$-axis data of the orientation sensor: 1) 1–90; 2) 91–180; 3) 181–270; and 4) 271–360. The relative positions between the different activity areas can be computed by a dead-reckoning filter, which will be shown in Section III-C2.

Fig. 6 illustrates the processes of landmarks detection. Such landmarks include activity areas of toilets, chairs, wardrobes, etc. From the massive amount of data, we extract records each of which corresponds to an activity area in a user’s trace. Each record includes its pattern of behavior, WiFi fingerprint, orientation, and relative positions relations with the last and the next activity areas. The task of discovering landmarks is rooted in 1) detecting indoor activity areas and 2) testing whether an indoor activity area can be considered a landmark.

1) Processes of Landmark Detection: As shown in Fig. 6(a), all the records are gathered in a matrix: matrix $\left[\begin{array}{c} i \\ j \end{array}\right]$ $(0 \leq j < 4)$ denotes the $j$th attribute of the $i$th record. We need to classify these records into some groups according to three of their attributes: 1) pattern of behavior; 2) WiFi fingerprint; and 3) orientation. The classification basis relies on the premise that the pattern of behavior and orientation are the same, and Wi-Fi is similar. The calculation of the similarity of WiFi at two locations refers to [9]. We choose a similarity threshold of 0.8 in our system. Fig. 7 shows the determination of thresholds according to the WiFi number after merging similar WiFi. We observed that 0.8 was a reasonable value to obtain accurate WiFi groups.

The purpose of classification is to find the same activity area that different records represent. However, there is a special case in the detection: a WiFi area may contain different activity areas with the same pattern of behavior and orientation (e.g., areas of two adjacent chairs), which leads to the result that these different activity areas are classified into one group. Therefore, each group needs to be classified further. Fig. 6(b) and (c) shows the process of further clustering using

---

**Fig. 4.** Sensor readings of action sequence.

**Fig. 5.** Fuzzy pattern recognition.
Fig. 6. (a) Matrix of records consists of several rows of records, representing by the pattern of behavior, Wi-Fi RSSI, orientation, and relative position. Classify the records into groups according to the first three attributes of records. (b) Recognize the seed points (door) from the groups according to the unique pattern of behavior and Wi-Fi. Based on seed points, use PDR to compute the positions of each record. (c) Cluster the positions of records in each group for further classification using DBSCAN. (d) Qualify the candidates as landmarks using the median distance of each cluster between the center and all members.

DBSCAN based on relative positions. To calculate relative positions, we need to recognize one or more points as references, which are called seed points, as shown in Fig. 6(b). In our system, we choose doors as seed points for the following reasons. First, doors frequently connect to other indoor objects. Second, the behavior of opening a door occurs at a small area. Finally, doors can be recognized easily because its pattern of behavior is the WT type, and WiFi RSSIs inside and outside the door are distinctly different. Note that a door has two activity areas: inside and outside the door. Based on the seed points, the relative coordinates of members in the other groups can be computed through dead reckoning. Then we rely on a DBSCAN clustering algorithm [14] to further classify each group based on group members’ relative coordinates with seed points. DBSCAN is density based and has two advantages: 1) the number of clusters is not required before conducting clustering and 2) outliers can be detected. The DBSCAN parameter $\text{MinPts}$ is used in order to reduce outliers and $\text{Eps}$ specifies the radius of each cluster. In our system, we set the $\text{MinPts}$ to 10. Because the error of SmartPDR is approximately from 0.3 m to 0.7 m, and the radius of each activity area is approximately 0.3 m, the other parameter, $\text{Eps}$, in this article is from 1.2 to 2.0. We decrease $\text{Eps}$ by 0.1 per time to conduct multiple clustering. If the center of a cluster of each clustering remains almost unchanged, we think this cluster is effective. As shown in Fig. 8, the clustering achieves the best effect when $\text{Eps}$ is 1.4. Each output cluster represents an indoor activity area. The location of the activity areas is the center of mass of all samples within the output clusters.

All activity areas are detected as landmark candidates. However, some landmark candidates with a large area cannot locate users precisely. To qualify the candidate cluster as a landmark, it must be confined to a small area. Fig. 6(d) shows the qualification from candidates to landmarks. Based on the result of the DBSCAN clustering algorithm, the members’ locations of a cluster are known. The core idea is that the median distance between the center and all members of a cluster can reflect the distribution of members, which also indicates the area of the cluster. We need to determine whether the median distance of a cluster exceeds a threshold. If it does, this candidate is a line-activity area, and it cannot be qualified as a landmark. Otherwise, it is a point-activity area and can be used to conduct indoor localization.

We set different thresholds to test the accuracy of distinguishing point-activity areas and line-activity areas. The result shows that when the threshold is less than 0.3, all candidates are recognized as line-activity areas. When the threshold is more than 1.5, all candidates are recognized as point-activity areas. When the threshold is 0.9, AtLAS can distinguish point-activity areas and line-activity areas with the best accuracy of 96.2%. Therefore, we choose the threshold of 0.9.

After detecting all indoor landmarks, each landmark needs its specific features to conduct the indoor localization. According to the result of DBSCAN, if there is only one landmark in an output cluster, its feature is the vector [pattern of behavior, Wi-Fi RSSI, orientation]; otherwise, we need to take methods to further distinguish different landmarks of the same group. In this article, relative positions are used to accomplish this. For example, point A and B are different landmarks of the same group. The coordinate of the close point C is $(c_x, c_y)$. According to the point C, the coordinates of points A and B are $(c_x + \Delta x_1, c_y + \Delta y_1)$ and $(c_x + \Delta x_2, c_y + \Delta y_2)$,
respectively. Such landmarks that depend on other landmarks are conditional; otherwise, they are independent landmarks.

AtLAS will update all landmarks all of the time. The update of landmarks includes the appearances of new landmarks, the disappearances of old landmarks, and the movements of landmarks. When we discover the movement of a landmark, we can essentially interpret it as the disappearance of a landmark and the appearance of a new landmark. Therefore, AtLAS will detect new landmarks consistently. When some landmarks have never been used to locate users’ positions within a month, they will be placed on a clipboard. If they are not used in the next month, they will be removed completely; otherwise, they will be recovered from the clipboard.

2) Dead-Reckoning Filter: We choose SmartPDR [15] for our implementation after comparing several algorithms of PDR. SmartPDR computes the displacement using step event detection, heading direction estimation, and step length estimation. The error of SmartPDR is mainly caused by the step length, which can be evaluated by accumulating the error of each step. Therefore, we design a method to filter the traces with a large error. Considering the effect of users’ heights on step length, we perform experiments on people with different heights to measure the error of each step. Fig. 9 shows the relationships between step length and the error of users with different heights. We find that the error of a normal step length is lower than other step lengths. Therefore, we divide step length into five levels, and each level corresponds to an error. Table III shows an example of these five levels. If a user cannot correspond with Table III, his ranges of five levels and corresponding errors need to be changed proportionally. For example, if a user’s normal step length is $\theta$ is not within $0.6–0.9$ m, all of his ranges of levels and errors need to be multiplied by $\theta/0.75$.

Therefore, we can evaluate each trace of users according to

$$f = 1.5 - \sum_{i=1}^{n} error_i.$$  \hspace{1cm} (3)

In the above equation, $i$ denotes the $i$th step. $error_i$ denotes the error of the $i$th step. $f$ denotes the evaluation of the trace. We need to set a constant threshold, which is shown in Section IV. If the output $f$ is less than the chosen threshold, we regard this trace as invalid in the landmark detection and filter it out. In the process of landmarks detection, all indoor traces must be evaluated, which can make the detection more accurate.

The problem is that SmartPDR cannot recognize which step is normal due to the different heights of users. Therefore, it is necessary to measure a user’s normal step length. The solution is using two definite landmarks to calculate a user’s normal step length through

$$l = \frac{1}{k} \sum_{i=1}^{k} S_{nk}.$$  \hspace{1cm} (4)

In the above equation, $S$ denotes the distance between two definite landmarks. $k$ denotes the number of traces through these two landmarks. $nk$ denotes the number of steps in the $k$th trace. $L$ denotes the normal step length of a user. The users need to measure the distance $S$ manually. After calculating a user’s normal step length, we can evaluate all of the traces of the user according to Table III and (3).

D. Indoor Localization and Floorplan Generation

After landmarks detection, AtLAS generates a database of landmarks and each landmark has its own coordinate. AtLAS begins to locate users as soon as users’ patterns of behavior are detected. According to the database, when the user meets the requirements of an independent landmark, he will be located with that landmark. For example, a user’s patterns of behavior, orientation, and WiFi RSSI are known. If AtLAS can find only one independent landmark from the database according

<table>
<thead>
<tr>
<th>Level</th>
<th>Range</th>
<th>Error (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$&lt;0.4m$</td>
<td>0.15</td>
</tr>
<tr>
<td>2</td>
<td>$0.4m–0.5m$</td>
<td>0.10</td>
</tr>
<tr>
<td>3 (normal)</td>
<td>$0.6m–0.9m$</td>
<td>0.05</td>
</tr>
<tr>
<td>4</td>
<td>$0.9m–1.1m$</td>
<td>0.10</td>
</tr>
<tr>
<td>5</td>
<td>$&gt;1.1m$</td>
<td>0.15</td>
</tr>
</tbody>
</table>
to this information, the user will be located with it. Otherwise, AtLAS uses features of actions, which are designed in the process of landmarks detection, to further locate uses. When there are no independent landmarks that can be matched, we use the last landmark to compute a relative position of the current location. If a current location and a pattern of behavior match a conditional landmark, the user will be located with this conditional landmark. Users’ locations will be shown on the generated floorplan.

Before semantic labeling, we need to detect all indoor objects to generate floorplan. In the section above, we have described point-activity areas and line-activity areas. In this unit, we present how to generate blank areas where people described point-activity areas and line-activity areas. In this study, we only try to label twelve objects, including door, bed, wardrobe, toilet, refrigerator, washer, sink, armchair, sofa, dining table, chopping board, and cooking bench. Other objects can also be labeled according to the corresponding features.

Before the division of functional areas, we first use a priori knowledge to label two kinds of objects that are easy to label: doors and beds.

1) Door: Doors can be recognized easily because its pattern of behavior is the WT type and the WiFi RSSI on both sides of the door is different.

2) Bed: Through analyzing the data, we find that people always sleep on the bed for a long time at night, which can be reflected by the inertial sensors. Therefore, the bed can be detected by the orientation sensor and the duration at night.

1) Functional Area Identification: We use doors to divide indoor functional areas according to the method in [16]. In this article, the functional areas mainly include bedroom, bathroom, kitchen, dining room and other. We use a library for support vector machine (libSVM) SF, which supports multi-class classification, to train and classify these functional areas. Considering the time and frequency of a behavior in each functional area is different, the feature vector of each functional area is defined as \([\text{PoB}_1, \text{Time}_{f1}, \ldots, \text{PoB}_n, \text{Time}_{f_n}]\) and the label is the type of the various functional areas. \(\text{PoB}_n\) denotes the \(n\)th kind of pattern of behavior occurring in this functional area. Here, patterns of behavior include sitting, standing, reclining, lying, opening, and stooping. We divide the time into eight parts from 0 to 24 on average. \(\text{Time}_{f_n}\) is a \(1 \times 8\) vector which denotes the frequency of the \(n\)th pattern of behavior in each part of time. We extract large amounts of feature vectors of 116 families from questionnaires and train the libSVM model, which can be used to identify functional areas universally.

E. Semantic Labeling Mechanism

After indoor objects detection, the distribution and connection of indoor objects can be represented by a directed graph. The nodes in the graph denote indoor objects, and the edges denote that there have been trajectories between objects. Residents always have some functional areas, such as a bedroom, kitchen, etc. As shown in Fig. 10, each node belongs to one or more functional areas (e.g., there is an overlap between two functional areas). Through many experiments and analyses, a key observation is determined regarding the information of an object and the functional areas related to this object: they are analogous to fingerprints for labeling this object. We find that some objects belong to one functional area in a fixed manner, and patterns of behavior that occurred with the objects and the time have mainly occurred in a regular manner, which allow them to be used as features. Therefore, we use a template matching method to label indoor objects. In this study, we only try to label twelve objects, including door, bed, wardrobe, toilet, refrigerator, washer, sink, armchair, sofa, dining table, chopping board, and cooking bench. Other objects can also be labeled according to the corresponding features.

Before the division of functional areas, we first use a priori knowledge to label two kinds of objects that are easy to label: doors and beds.

Fig. 10. Functional areas of a residence.
we reuse the libSVM to identify these functional areas. If there still is a small difference among the current results of libSVM, we label this area as all possible results of the libSVM.

2) Indoor Object Label: So far, we have identified the functional area each object belongs to. We set standard templates for each kind of object in advance according to data from the questionnaires of 116 families.

As shown in Fig. 11, a template includes the features of the object itself and some features about the functional area linked to this object. FA denotes function area. The format of the template of any kind of object is as follows: \{FA itself, PoB itself, Time f1, Feature FA\}. In the template, FA itself is a 1 × 5 vector which denotes the functional area to which a certain kind of object belongs. FA itself needs to be the same as the template but not similar. PoB itself is a 1 × 6 vector [sitting, standing, reclining, lying, opening, stooping], which denotes the patterns of behavior occurring with an object. The vector [0, 0, 0, 0, 1, 1], for example, represents that the patterns of behavior that occur with this object are opening and stooping. We divide the time into eight parts from 0 to 24 on average. Time f1 in the template denotes the frequency of the nth pattern of behavior in each part of time. For instance, the vector Time f1 [0, 0, 0, 0.5, 0.5, 0, 0] represents that the frequencies of sitting that occurs on an object are 0.5 in time period 12:00–15:00 and 15:00–18:00. Feature FA is a 5 × 18 2-D matrix. Each row corresponds to a functional area, which can be represented as [PoB FA, fFA IN, fFA OUT]. PoB FA is a 1 × 6 vector, denoting the patterns of behavior that occur in a functional area that links to the object which is ready to be labeled. fFA IN denotes the frequency of connection from objects corresponding to PoB FA to the object to be labeled. fFA OUT denotes the frequency of connection from the object to be labeled to objects corresponding to PoB FA. Note that the low frequency will be considered to be 0, which is chosen to improve the recognition accuracy. If there are several same functional areas, such as two bedrooms, their data need to be merged.

Because these features are all vectors, we use Jaccard distance and Pearson distance to calculate the similarity between the actual situation and the template. The similarity \( S_j \) of vectors FA and PoB needs to be calculated by Jaccard similarity coefficient because it is more reasonable for vectors only containing 0 and 1. The similarity \( S_j \) of vectors Time, \( f_{\text{in}} \), \( f_{\text{IN}} \) and \( f_{\text{OUT}} \) can be calculated by PPMCC. The total similarity is calculated in the following equation. Through experiments, we find that different objects correspond to different \( \mu \) and \( \nu \).

\[
S = \mu S_j + \nu S_p. \tag{5}
\]

When the result is closer to 0, the actual situation is more similar to the template. In addition, the recognition result should meet (6) and (7), which is to ensure that the result is not misidentified

\[
\min_{i} (S_i) < \gamma \tag{6}
\]

\[
\min_{i} \left( S_i - \min_{i} S_i \right) < \theta. \tag{7}
\]

In the above equations, \( S_i \) denotes a set of results that match all the templates. \( \gamma \) and \( \theta \) denote a constant threshold.

According to the specific habits of household items placement, we set up some rules that the objects that have been labeled can help to label other objects, which can help the system label objects quickly and more accurately. For example, there are frequent connections between objects A and B, and A has been labeled. The object B can be labeled “name1” or “name2.” According to observation, “name1” has never been connected with the object A. Therefore, the object B can only be labeled “name2.” Therefore, there are also some rules limiting connections among objects, which is constructed to reduce the range of the object to be labeled.

The construction of a feature vector in an actual situation is similar to the construction of templates. The specific matching process is as follows.

1) We need to find popular objects that can be labeled by the template matching method accurately according to the frequency of connections. Popular objects can be labeled at a higher accuracy because of sufficient data.

2) AtLAS starts a breadth-first search (BFS) algorithm from the objects that have been labeled to label other indoor objects. If the result of template matching cannot satisfy (6) and (7), the corresponding object will be skipped temporarily. In addition, the labeled objects can reduce the range of unlabeled ones.

3) The step 2) is repeated until the labels of indoor objects no longer change.

Algorithm 1 shows the detailed steps for semantic labeling. Lines 1–13 show the main steps for functional area identification. Lines 14–24 show the main steps for semantically labeling objects.

IV. PERFORMANCE EVALUATION

In this article, we first need to set the standard template of fuzzy pattern recognition and semantic labeling in advance, which can be obtained by analyzing the massive of data. We have invited six families containing 20 volunteers (almost all members of each family participated), of different ages (12 in their 40 s, five in their 20 s, and three in their teens) and gender (8 males and 12 females), to collect data in residences.
These residences include three kinds of house types and indoor layouts (90, 120, and 160 m²), and for each type, we have two families to collect data. All the collection activities continued for approximately a month. All data is used to generate the standard template of fuzzy pattern recognition on the computer, which is similar to the training process. In addition, we invited 116 families to fill in the questionnaires. This is the link to the questionnaire: https://www.wjx.cn/jq/34525387.aspx.

In this section, we conduct the experimental evaluation to test the correctness and effectiveness of AtLAS. AtLAS is implemented in three families with different layouts, 90, 160, and 300 m², respectively, in Wuhan. Fig. 12 shows the floor-plan of 160-m² residence. We do not show the layout for each testbeds due to space constraints. We embed sensors into necklaces. The inertial sensors as well as WiFi information are sampled and sent to smartphones and the server for processing in online or offline mode. Users of different families, wearing the necklaces, collected the data of their daily life in the testbeds for 24 h every day in a month. In addition, we installed some cameras on each testbed and fixed some sensors on participants’ feet, which is to compute all participants’ ground truth of locations.

In this section, we evaluated the action identification accuracy, landmarks detection (including activity area detection), users’ localization accuracy, and semantic labeling accuracy.

### A. Pattern of Behavior Recognition Accuracy

We evaluate AtLAS on patterns of behavior recognition accuracy. We performed nearly 1000 times for each pattern of behavior and evaluate their accuracy, mistake rate, and miss rate. Table IV shows the result of the action identification. The result shows that the identification accuracy of AtLAS is 91.4%. The mistake rate is less than miss rate, which is what we expect to see.

### B. Landmarks Detection

In this unit, we evaluate 1) how many activity areas are detected in different residences; 2) the determination of effective PDR traces; 3) the landmark localization errors over
Fig. 13. (a) Number of activity areas in the residence. (b) Activity area detection accuracy and false positive.

Fig. 14. Determination of effective PDR traces.

Fig. 15. Effect of dead-reckoning filter.

1) Determination of PDR Traces: In this article, PDR is used for landmark detection and indoor localization. Therefore, the threshold of PDR traces needs to be set to evaluate whether the traces between landmarks are effective. We evaluate the effect of different thresholds on the convergence time and landmark localization error. The convergence time denotes the time during which the quality and quantity of landmarks tend to be stable. The landmark localization error indicates the difference between the locations of landmarks in the detection method and those in the actual setting. In total, we set five thresholds. Fig. 14 shows the effect of different thresholds. From Fig. 14, we can observe that when the thresholds are 0.4 and 0.5, landmark localization errors are larger than the other three situations, which is because the low threshold cannot filter the traces with large errors. When the thresholds are 0.7 and 0.8, the convergence time is longer than the other three situations due to the sparse data. When the threshold is 0.6, the convergence time reaches a minimum value of ten days on average, and it can achieve 0.43 m of landmark localization errors on average.

In addition, Fig. 14 shows that the error of the landmark locations decreases over time. With the time increasing, the locations of landmarks become more accurate, and finally reach a stable state.

Fig. 15 shows the effect of the dead-reckoning filter. From Fig. 15, we can determine that the error of most traces is within 1 m after the application of the dead-reckoning filter. The maximum error without the dead-reckoning filter is greater than the situation with dead-reckoning filter, which means our method can reduce the error of dead reckoning effectively by filtering the traces with large errors.

2) Density of Landmarks: Fig. 16 shows the distribution of landmarks in the residence. In the residence, AtLAS detected 43 valid landmarks, including 37 independent landmarks and six conditional landmarks. SemanticSLAM [9] detected 12 WiFi landmarks and nine magnetic and inertial sensor landmarks, and an activity-based system [17] detected only 13 landmarks. Through calculation, we get the average distance between adjacent landmarks in trajectories is 3.2 m. However, the average distances of semantic and the activity-based system are more than 5 m, which leads to more
Fig. 16. Comparison of distribution of landmarks in the testbed.

Fig. 17. Average distances between adjacent landmarks in a trajectory in three different residences.

Fig. 18. Effect of different samples.

In addition, the line-activity areas labeled in Fig. 16 are detected accurately, which cannot be used in indoor localization.

Because the average distance between adjacent landmarks in a trajectory is related to different layouts and the areas of residences, Fig. 17 shows average distances of AtLAS and the other two methods in three different residences. Compared to the other two methods, AtLAS can achieve a smaller average distance between adjacent landmarks in a trajectory, which is more beneficial to correct drift error of PDR.

C. Localization Accuracy

In this part, we evaluate how many data can satisfy the demand of AtLAS to estimate a user location accurately as well as user localization precision in different cases.

1) Effect of the Number of Samples: Considering that AtLAS collects data from all family members, more samples will strengthen the system performance. In addition, the effect will not increase with increasing the number of samples after it reaches a peak. Fig. 18 shows the effect of different numbers of samples on the overall user localization precision. The fact is that more samples can detect more accurate landmarks. When the samples are sufficient, even if some landmarks have some outliers, the system can reduce them to estimate their locations accurately. We find that when the number of samples is approximately 40, the performance of AtLAS will reach a peak of a median accuracy of 0.57 m. Therefore, we leverage 40 samples to conduct the rest of the evaluation.

2) User Localization Accuracy: Fig. 19 shows the CDF of the localization error in two modes. The result shows that the offline mode is more accurate than the online mode in localization.

Fig. 20 shows the CDF of the localization error with and without the conditional landmarks. From Fig. 20, the case of using conditional landmarks improves the localization accuracy of AtLAS, which is because it has more landmarks.

3) Comparison With Other Systems: We compare AtLAS, SemanticSLAM [9], and an activity-based system [17] in the residence. SemanticSLAM is based on seed landmarks and organic landmarks to correct the error caused by sensor drift. Seed landmarks include elevator, escalator, and stairs, and organic landmarks contain WIFI landmarks and magnetic and inertial sensor landmarks. The activity-based system can recognize activities, including walking up and down staircases as well as opening and closing doors, which is to detect stairs and doors as landmarks. Fig. 21 shows that AtLAS has a better average error of localization because it uses more landmarks based on patterns of behavior and its density of landmarks.
TABLE V
CONFUSION MATRIX FOR FUNCTIONAL AREA LABELING IN RESIDENCE

<table>
<thead>
<tr>
<th></th>
<th>Bedroom</th>
<th>Bathroom</th>
<th>Kitchen</th>
<th>Dining room</th>
<th>Other</th>
<th>FP</th>
<th>FN</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bedroom</td>
<td>162</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
<td>162</td>
</tr>
<tr>
<td>Bathroom</td>
<td>0</td>
<td>149</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>5%</td>
<td>5.1%</td>
<td>157</td>
</tr>
<tr>
<td>Kitchen</td>
<td>0</td>
<td>0</td>
<td>96</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
<td>96</td>
</tr>
<tr>
<td>Dining room</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>109</td>
<td>6</td>
<td>8.7%</td>
<td>5.2%</td>
<td>115</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>6.7%</td>
<td>9%</td>
<td>166</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.1%</td>
<td>3.9%</td>
<td>696</td>
</tr>
</tbody>
</table>

D. Semantic Labeling

1) Semantic Labeling Accuracy: Table V is the confusion matrix for the result of labeling functional areas. We extract data of functional areas from questionnaires including 116 families. Then, the libSVM model is trained and tested through cross-validation. The result indicates that functional areas can be recognized accurately with a 4.1% false-positive rate and a 3.9% false-negative rate on average. Because the action of lying always occur in bedroom at midnight and actions in kitchen always occur at a fixed time, bedroom’s and kitchen’s false-positive rate and false-negative rate are 0%.

Table VI shows the confusion matrix for the result of semantic labeling objects in testbeds. The result shows that indoor objects can be labeled with a 93.6% accuracy. In addition, AtLAS can achieve a 5.7% false-positive rate and a 5.8% false-negative rate on average.

2) Convergence Rate of Semantic Labels: We randomly selected one family from six families to evaluate how many different types of objects can be labeled according to different number of activity records. We calculate and average the number of types of labeled objects every five days. As shown in Fig. 22, we first divided the members of the family into different groups, including householder and nonhouseholder, and let them walk and do actions consciously and unconsciously. Householders’ activity areas cover every area of family and nonhouseholders’ activity areas mainly contain bedroom, living room, and bathroom. We find that the convergence rate of “householder, unconsciously” is faster than “ungrouped, unconsciously” because nonhouseholders’ data are not comprehensive. The group “ungrouped, consciously” means participants do some activities for a special purpose. We can find the convergence rate of “ungrouped, consciously” is further improved because its data is less redundant.

Fig. 23 indicates that different kinds of objects are labeled at different convergence rates. Because doors and beds are the basis of semantic labeling, We merely research convergence rates of other ten objects. We can find that objects like sofas and armchairs have a much faster convergence rate than other objects. It is because users interact with them more frequently and we call them hot objects. In contrast, it takes more time
TABLE VI
CONFUSION MATRIX FOR SEMANTIC LABELING OBJECTS IN RESIDENCE

<table>
<thead>
<tr>
<th>Door</th>
<th>WR</th>
<th>Toilet</th>
<th>RF</th>
<th>Washer</th>
<th>Sink</th>
<th>Armchair</th>
<th>Sofa</th>
<th>Bed</th>
<th>DT</th>
<th>CB1</th>
<th>CB2</th>
<th>FP</th>
<th>FN</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Door</td>
<td>541</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WR</td>
<td>0</td>
<td>392</td>
<td>0</td>
<td>31</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Toilet</td>
<td>0</td>
<td>0</td>
<td>464</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>RF</td>
<td>0</td>
<td>0</td>
<td>565</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5.10%</td>
</tr>
<tr>
<td>Washer</td>
<td>0</td>
<td>0</td>
<td>41</td>
<td>0</td>
<td>401</td>
<td>41</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6.60%</td>
</tr>
<tr>
<td>Sink</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>493</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7.80%</td>
</tr>
<tr>
<td>Armchair</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>461</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.10%</td>
</tr>
<tr>
<td>Sofa</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>407</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11.90%</td>
</tr>
<tr>
<td>Bed</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>471</td>
<td>0</td>
</tr>
<tr>
<td>DT</td>
<td>0</td>
<td>29</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>482</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>CB1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>452</td>
<td>20</td>
<td>2.30%</td>
</tr>
<tr>
<td>CB2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>363</td>
<td>5.30%</td>
</tr>
</tbody>
</table>

| Overall | 5.70% | 5.80% | 5867 |

WR = Wardrobe, RF = Refrigerator, DT = Dining table, CB1 = Shopping board, CB2 = Cooking bench.

Fig. 23. Convergence rate of different objects’ semantic labeling.

for objects such as washers and wardrobes to stabilize the labeling accuracy.

V. RELATED WORKS
A. Patterns of Behavior Identification

Currently, almost all of the machine learning methods on human activity identification learn low-level actions, such as walking, turning, etc. Some systems [18]–[20] extract features from figures to recognize users’ actions, which need to install some cameras in indoor environments and requires a vast supply of figures to train model. The system [21] fixes a wearable depth camera on user’s chest to recognize activities. In addition, Inertial sensors in smartphones or wearable devices have been widely used for human activity recognition (HAR) [22]. Some systems [23]–[25] utilize a deep learning model to mine information from inertial sensor readings, which is in order to recognize users’ activities. These methods cannot recognize high-level activity, such as the pattern of opening, which consists of a set of low-level actions.

A single action cannot reflect what are users doing, but action sequences can. AtlAS recognizes the basic actions by a threshold method and uses action sequences to recognize human patterns of behavior through fuzzy pattern recognition. The advantage of the proposed method is that it can eliminate the variants of action sequences, and it is used to detect indoor landmarks where the same patterns of behavior occur.

B. Indoor Localization

Recently, more systems have been proposed to solve the indoor localization problems. The dead-reckoning results, obtained by smartphone sensor readings, have been used in indoor localization [26]. However, the influence of accumulative error caused by inertial sensor has not been solved. RF-based systems have developed rapidly because of their ubiquitous deployment. Recently, many systems which are based on crowdsensing take the advantage of RF signal and dead reckoning to locate users [3], [27]–[31]. Though these systems have simplified the calibration effort and corrected dead-reckoning errors, the localization precision is relatively lower. The system in [32] uses the Doppler frequency shift to measure the spatial relation between a human and an item. Although its accuracy is high, each item needs to have RFID tags attached. IndoTrack [33] extracts accurate Doppler velocity information from noisy WiFi channel state information (CSI) samples and determines the absolute trajectory of the target by jointly estimating target velocity and location via probabilistic co-modeling of spatial–temporal Doppler and AoA information. ViVi [34] exploits the spatial relationships among the RSS fingerprints of multiple neighboring locations to locate users. Both of IndoTrack and ViVi need to design the location of WIFI device and are limited by the number of AP. Guo [35] takes transformations of received signals as fingerprints and obtains a more accurate localization result.

Some systems leverage sensor readings in specific places, such as a magnetic signature near an electrical service room, an elevator, and WiFi signature to detect landmarks [9], [10], [40]. The system in [36] utilizes the electromagnetic field formed by indoor alternating current to locate users. These systems are applicable for large-area buildings, such as a shopping mall and an engineering building, because they cannot detect enough landmarks with high precision in residences due to a lack of specific places.

Some systems leverage environmental physical features to conduct indoor localization without relying on the RF signature [1], [2], [41]–[43]. In such systems, users’ relative positions to physical features and specific reference points are obtained from videos and images. ClickLoc [37] is an accurate, easy to deploy, sensor enriched, and image-based indoor localization method.
TABLE VII
COMPARISON OF ACTIVITY IDENTIFICATION, INDOOR LOCALIZATION, AND SEMANTIC LABELING TECHNIQUES

<table>
<thead>
<tr>
<th>Part</th>
<th>Citation</th>
<th>Techniques</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patterns of Behavior Identification</td>
<td>Chen et al. [20]</td>
<td>Kernelized rank pooling, Riemannian optimization</td>
<td>Universality of sequential data</td>
<td>Requires lots of videos</td>
</tr>
<tr>
<td></td>
<td>Voigt et al. [21]</td>
<td>Nonlinear binning strategies, sliding window</td>
<td>Accurate</td>
<td>Needs to wear an uncomfortable device</td>
</tr>
<tr>
<td></td>
<td>Guan et al. [24]</td>
<td>Integration of LSTM networks</td>
<td>Robust</td>
<td>High computational load</td>
</tr>
<tr>
<td></td>
<td>Bianchi et al. [25]</td>
<td>CNN network</td>
<td>Personaliized</td>
<td>High energy consumption</td>
</tr>
<tr>
<td>Indoor Localization</td>
<td>RollCall [32]</td>
<td>Doppler frequency shift, human-item spatial relation measurement</td>
<td>High accuracy, in real time</td>
<td>Each item needs to have RFID tags attached</td>
</tr>
<tr>
<td></td>
<td>InfoTrack [33]</td>
<td>Doppler frequency shift, probabilistic space-time joint trajectory estimation</td>
<td>Accurate</td>
<td>High computational load</td>
</tr>
<tr>
<td></td>
<td>ViVi [34]</td>
<td>Fingerprint spatial gradient, pattern matching</td>
<td>Accurate, robust to temporal instability</td>
<td>Affected by long-term changes</td>
</tr>
<tr>
<td></td>
<td>Guo et al. [35]</td>
<td>Windowing and sliding techniques based on a group of fingerprints, a fusion algorithm called multiple classifiers multiple samples</td>
<td>Robust</td>
<td>Needs to build special receiver platforms</td>
</tr>
<tr>
<td></td>
<td>Lu et al. [36]</td>
<td>Loop closure detection and curation algorithms, dead-reckoning</td>
<td>In real time</td>
<td>Not suitable for residences</td>
</tr>
<tr>
<td></td>
<td>ClickLoc [37]</td>
<td>Semantic information extraction, optimization-based sensor data fusion</td>
<td>Accurate, easy-to-deploy</td>
<td>Needs extra equipment, requires floorplan</td>
</tr>
<tr>
<td>Semantic Labeling</td>
<td>DIMLOC [38]</td>
<td>Vision analysis, scaling factor</td>
<td>High accuracy, robust</td>
<td>Requires extra equipment, may lead privacy leaks</td>
</tr>
<tr>
<td></td>
<td>Jigsaw [16]</td>
<td>Crowdsourcing technology, landmark modeling, augmentation algorithms</td>
<td>Accurate for complex indoor environments</td>
<td>Needs manual image classification, high energy consumption</td>
</tr>
<tr>
<td></td>
<td>TransitLabel [10]</td>
<td>Crowdsourcing technology, a classifier approach based on sensors and sound</td>
<td>Robust</td>
<td>Not suitable for residences</td>
</tr>
<tr>
<td></td>
<td>WiFiMap+ [39]</td>
<td>A two-stream architecture of CSI, a spatial stream generation algorithm</td>
<td>Device-free</td>
<td>Disturbed by multiple subjects, ignores privacy protection</td>
</tr>
</tbody>
</table>

Localization system. These vision-based positioning methods need users to equip a depth camera or set extra specific devices in the indoor environment, and are easily disturbed by the surroundings. In 2017, Google provided an indoor localization service, VPS, which utilizes visual information to locate users. The system in [38] only uses two dimmable LEDs and achieves centimeter precision in positioning.

This article leverages human daily behaviors to conduct high-precision localization, which is not limited by the changes of the environment and requires no calibration work. In addition, AtLAS only needs users to equip themselves with some inexpensive sensors, which is feasible for users in indoor environments.

C. Semantic Labeling

Labeling objects makes a map easier to use. Some systems [9], [44] require labeling the organic landmarks artificially while generating the map, which is tedious. The system in [11], [16], and [21] extracts features of items from videos and figures to discover what the items represent. TransitLabel [10] considers that certain passengers’ activities represent special signatures that are reflected on sensors. WiFiMap+ [39] utilizes WiFi signals to recognize indoor semantics without additional device for users. However, these methods are easily disturbed by the surroundings, such as light, noise, and multiple subjects.

Because there are some relationships between objects in users’ traces, AtLAS uses these relationships and the topological structure among indoor objects to extract features of each object. Then, we label each object by calculating the Euclidean distance between the features of the object and templates. Note that these relationships between objects in users’ traces usually reflect human manners in daily life. Therefore, these relationships will not change with the change of the environment.

Table VII shows some comparison of the aforementioned methods in each of the three parts, with their techniques, advantages, and limitations.

VI. CONCLUSION

In this article, we proposed AtLAS, which is aimed at activity-based indoor localization and semantic labeling in residences. AtLAS identifies users’ patterns of behaviors, which always occur in a small spatial area, to determine users’ high-precision locations. Meanwhile, AtLAS identifies semantics by mining the relationship between the objects in users’ traces, which labels indoor objects semantically to generate a detailed floorplan. The experimental results show that AtLAS can achieve 0.43 m of median landmark localization errors and 0.57 m of median user localization errors without 5% worst errors. In addition, it can label the objects accurately with a 5.7% false-positive rate and a 5.8% false-negative rate on average, including door, bed, wardrobe, toilet, refrigerator, washer, sink, armchair, sofa, and air conditioner. Compared to the state-of-the-art indoor localization systems, AtLAS can provide a higher precision localization and a floorplan with rich semantic knowledge automatically, which are both significant for the development of an intelligent location-based service, e.g., a smart home application.

REFERENCES

Xiaoguang Niu (Member, IEEE) received the B.Sc. degree from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2002, the M.Sc. degree from Nankai University, Tianjin, China, in 2005, and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2009.

He is currently a Professor with the School of Computer Science, Wuhan University, Wuhan, China. He has authored or coauthored more than 30 refereed journal and conference papers. His research interests include wireless networks, indoor positioning, and mobile computing.

Haiming Chen (Member, IEEE) received the B.Eng. and M.Eng. degrees in computer engineering from Tianjin University, Tianjin, China, in 2003 and 2006, respectively, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2010.

He is currently an Associate Professor with the Department of Computer Science, Ningbo University, Ningbo, China. He has authored or coauthored over 50 papers in areas of wireless, ad hoc, sensor networks, and networked embedded computing systems.

Luyao Xie received the B.Eng. degree in information security from the School of Computer Science, Wuhan University, Wuhan, China, in 2019, where she is currently pursuing the master’s degree in computer science, under the supervision of Prof. Niu.

Her research interests include mobile sensing, indoor positioning, and machine learning.

Dandan Liu received the B.Sc. degree in engineering and the Ph.D. degree in engineering from Wuhan University, Wuhan, China, in 2003 and 2008, respectively.

She is currently an Associate Professor with the School of Computer Science, Wuhan University. Her research interests include wireless networks, mobile computing, Internet of Things and communications.

Jiawei Wang received the B.Eng. and M.Eng. degrees of computer science from the School of Computer Science, Wuhan University, Wuhan, China, in 2017 and 2019, respectively. His research interests include information security, sensor and Internet of Things, and indoor location.

Ruizhi Chen received the B.Eng. degree from Wuhan University, Wuhan, China, in 1985, and the Ph.D. degree from the University of Helsinki, Helsinki, Finland, in 1991.

He is currently a Professor and the Director of the State Key Laboratory of Information Engineering in surveying, mapping, and remote sensing with Wuhan University. He was an Endowed Chair Professor with Texas A&M University Corpus Christi, TX, USA, the Head and a Professor with the Department of Navigation and Positioning, Finnish Geodetic Institute, Kirkkonummi, Finland, and the Engineering Manager with Nokia, Espoo, Finland. He has published two books and more than 200 scientific papers. His current research interests include indoor positioning, satellite navigation, and location-based services.