IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 4, 2019, accepted June 21, 2019, date of publication June 27, 2019, date of current version July 19, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2925460

SELoc: Collect Your Location Data Using Only a Barometer Sensor

HAIBO YE[1,2], LI SHENG[1], TAO GU[3], AND ZHIQIU HUANG[1]

[1]College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211100, China
[2]Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing 211100, China
[3]School of Computer Science and IT, RMIT University, Melbourne, VIC 3001, Australia

Corresponding author: Haibo Ye (yhb@nuaa.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61702261, in part by the China Postdoctoral Science Foundation under Grant 2017M621742, and in part by the Australian Research Council (ARC) Discovery Project Grants under Grant DP190101888 and Grant DP180103932.

ABSTRACT This paper presents the design and implementation of a location data collecting system only using the barometer sensor on the smartphone. It is energy efficient because of using low-power barometer sensor readings to infer the location, and it protects user privacy by providing the non-sensitive location data. To design such a location data collecting system, we make some key technical contributions: 1) a curve fitting-based solution to remove the barometer sensor reading noise caused by weather change; 2) a deep learning algorithm to detect user moving activities based on the restricted Boltzmann machine, and; 3) a clustering-based extraction algorithm for signatures of different locations. The field studies show that the SELoc provides user daily locations with an accuracy of 85%; meanwhile, the average energy consumption is only about 22% compared with GPS.

INDEX TERMS Location data collection, barometer, smartphone sensors.

I. INTRODUCTION

Location based services are becoming more and more widely-used today, and the market size is said to reach 68.85 billion USD by 2023 [1]. Traditional location based services are mostly instant services which make use of the current location of the user, such as navigation and nearby search. Currently, with the help of big data analysis technologies, diverse location based services have been provided based on numerous historical location data. For example, the historical location data contains the behavioral characteristics of the urban crowd, which can help to deal with the major social science issues such as disease transmission, poverty eradication, and urban planning. However, the lacking of the historical location data greatly restricts the development of these services. Currently, the main data sources are the scattered historical locations recorded when people are using instant location services.

There are two major challenges for dealing with the lacking of location data problem. The first challenge is the energy efficiency concern. People are not willing to provide continuous locations because the positioning modules such as GPS and Wi-Fi are power-hungry components on smartphones, and they may deplete batteries within hours. The other challenge is the privacy concern. People worry about their privacy, they often don't want to expose the sensitive locations such as the address of home and working place.

We believe a location data collecting system which can provide considerable historical location data should ideally satisfy both the energy efficiency and privacy requirements. For energy requirements, recent technologies try to minimize the GPS sampling rate for energy efficiency [2]. However, the real energy saving is limited, because there is extra energy consumption when the GPS module is initializing and powering off [3]. Other techniques based on wireless fingerprint [4], [5] or video [6] can be energy efficient, but they require to deploy extra infrastructure in the environment, such as Wi-Fi access points, Bluetooth beacons and cameras, which may expose your portrait and other personal information. To deal with the privacy problem, many location privacy protection technologies have been proposed. Such as the Anonymous space technology and False location technology [7], [8]. However, they often need a trusted third party, and need extra communication cost. The extra computation

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Abdur Razzaque.

needed will also affect the energy efficiency. These solutions are mostly used in special application scenarios with high privacy requirements.

In this paper, we propose to design a location data collecting system only using the barometer sensor, which can provide considerable historical location data for the location based services. The system is energy efficient and can protect user privacy to some extent. Before proposing our solution, we need to explain how this kind of location based service works. At first, users provide their history location data to the service provider, such as a set of coordinates obtained by GPS. Then, the service provider will transfer the coordinates into meaningful location results. For example, to analyze the working pressure of the office workers in a city, the provider will transfer user history coordinates into locations such as office, and home, to get users' daily working hours. There is a privacy risk because your detailed address information may be abused for other purposes. How to avoid such problem? Our important observation is that many services only need your non-sensitive location data instead of sensitive physical coordinates. In this example, the service provider only needs to know when you are at the office, but doesn't need to know the detail address of your office. Consequently, if we provide the non-sensitive location directly, the privacy risk can be greatly reduced.

The non-sensitive location can be understood as what's on that physical coordinate. Home, office, restaurant, supermarket, etc. are typical non-sensitive frequent locations of a user. A simple solution is to locally transform the coordinates to non-sensitive locations, and then provide them to the service provider. But that would cause bad energy efficiency, especially for providing considerable historical location data. In this paper, we try to detect the non-sensitive location directly only using barometer sensor readings, and instead of detecting the sensitive physical coordinates. In the rest of the paper, when we say location we mean the non-sensitive location.

In the city area, there often exists different altitude change patterns when the user is arriving and leaving these frequent locations. We call these altitude change patterns as signatures of the locations, and it can be used to detect the location. For example, the altitude of one's home is different from the altitude of one's office and other locations one often go. The altitude change patterns [9], [10] when arriving and leaving the locations can be measured by the barometer sensor on the smartphone, and it is energy efficient. However, it is still not trivial to detect one's location for the following challenges: **a) the barometer sensor reading is affected by the weather, bringing errors to the measurement of altitude change.** If we directly map the barometer reading change to altitude change, the noise can be tens of meters, and in this way, the altitude change patterns of different locations can not be rightly extracted. **b) it is hard to extract signatures of different locations, and what's more it can be different for different users.** For example, people work and live at different places, and the altitude change patterns of home and office are different for every person. Some techniques need to be used to analyze and extract the unique signatures of the locations for every user, which is very challenging.

In this paper, we propose SELoc, a Safe and Energy efficient Locating system to handle these challenges. It is a location data collecting system which provides user non-sensitive location data for location based services. Compared to the traditional location technologies, SELoc has the following strengths. 1) SELoc only makes use of the low power barometer sensor on smartphone to infer the locations, which is very energy efficient, especially when used for providing considerable historical location data. 2) It detects the non-sensitive location directly instead of detecting the physical coordinate, which reduces the privacy risk a lot. In summary, we make the following contributions:

1. To our best knowledge, SELoc is the first system which mainly uses the barometer reading data for location data collection. It provides non-sensitive location data in user's daily life, and is more safe and energy efficient. For this reason, it is very suitable for the location services based on numerous historical frequent location data.
2. We proposed a curve fitting based solution to remove the barometer reading noise caused by weather change.
3. We detected user moving activities using the deep learning algorithm based on the Restricted Boltzmann Machine, and extract signatures of different locations by clustering.
4. We carried out a nation-wide online survey to confirm the desirability for SELoc, and we conducted extensive field studies to analyze the performance of SELoc. The field study shows that SELoc provides user daily locations with an accuracy of 85%. Meanwhile, the average energy consumption is only about 22% compared to GPS based approach.

In the rest of this paper, we give the overview and detailed design in Section II. Later, Section III shows the evaluation results, and Section IV shows the result of an online survey. Section V discusses the related work, and finally, Section VI concludes the paper.

II. SYSTEM DESIGN

The framework of our approach is presented in Fig. 1. The left side shows the scene, people are moving between frequent locations in their daily lives, the purpose of our system is to detect and provide these location data for location based services. The right side shows the sensing data and the model. We continually collect the raw barometer reading data from smartphone. Then, we use a two-step solution to provide the history locations. First, based on the sensing data, it is possible to detect three different movement activities of the user: 1) transportation activity which happens in vehicles, 2) indoor movement activity which happens in buildings and 3) outdoor physical activity which happens in

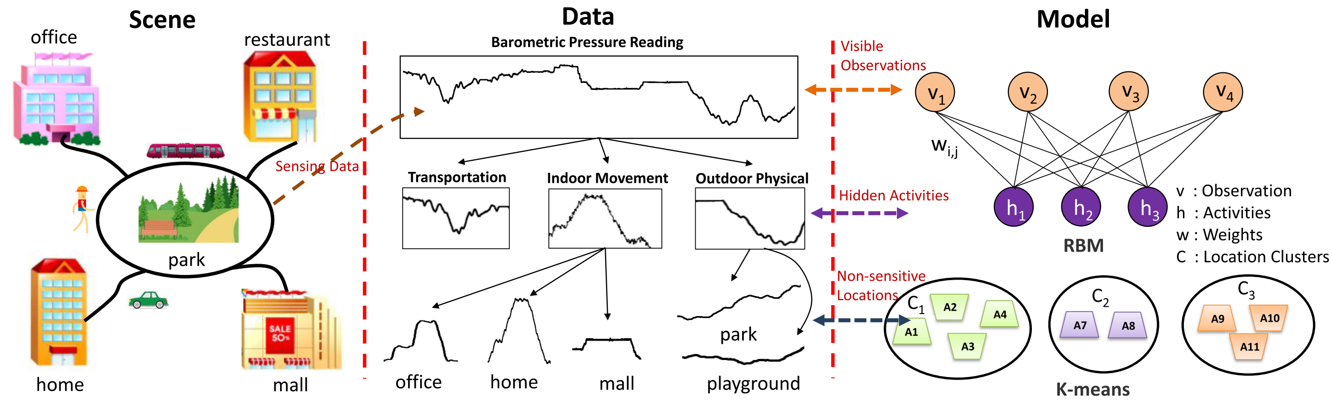


FIGURE 1. Overview of SELoc.

outside locations. We model the classification of three types of user movement activities using a Restricted Boltzmann Machines (RBMs) based activity model, which is one of the most commonly used deep learning algorithms today. In the second step, we use unsupervised learning technique to obtain the locations of every movement activity. In detail, we exploit the CURE based clustering technique to discriminate the altitude change patterns of moving at different locations, and then label the locations such as home, office, and restaurant,.etc. In the following parts of this section, we will show in detail the design of SELoc.

A. BACKGROUND OF BAROMETRIC PRESSURE AND BAROMETER SENSOR

Barometric pressure is the force per unit area exerted on a surface by the weight of air above that surface in the atmosphere of Earth [9]. At low altitudes above the sea level, the pressure decreases by 0.12 *hPa* for going up every 1 meter. Based on this feature, we can calculate the altitude change of a user based on the barometric pressure change measured by the barometer sensor. There also exists a formal formula relates barometric pressure p to altitude h, which is

$$h = 44330 * (1 - (\frac{p}{p_0})^{\frac{1}{5.255}}) \tag{1}$$

However, the pressure around people changes all time, not only because of the altitude change. The change of temperature and humidity will also cause pressure change. We did an experiment to measure the pressure change in an office for half an hour, the result shows that there is a max variation of 1.2 *hPa*, which is about 10 meters if transferred to altitude.

The Barometer sensor here is a digital sensor appears in most smartphones today, which can measure the barometric pressure around the phone. The most commonly used barometer sensors are BMP280, BMP180/182 and LPS331AP. Some information of the sensors are shown in Table 1. These sensors are very sensitive, they even can detect a change of the barometric pressure surround when you move up or down for only 1 meter. Consequently, they can be used to detect the altitude change of the users when they are moving in different buildings. However, the barometric pressure surround not only changes by user's vertical movement but also is changing all the time by the temperature and humidity. In order to measure accurate altitude change of the user, the noise caused by the temperature and humidity change should be handled.

TABLE 1. Barometer sensor parameters.

Property	BMP280	BMP180/182	LPS331AP
Absolute accuracy	± 1 hPa (± 8.5m)	-4.0 , +2.0 hPa (-33,+17m)	- 3.2,+2.6hPa (-27,+22m)
Relative accuracy	± 0.12 hPa (± 1m)	± 0.12 hPa (± 1m)	± 0.2 hPa (± 1.7m)
Noise	0.013 hPa (0.11m)	0.06 hPa (0.5m)	0.06hPa (0.5m)

B. DATA PREPROCESSING

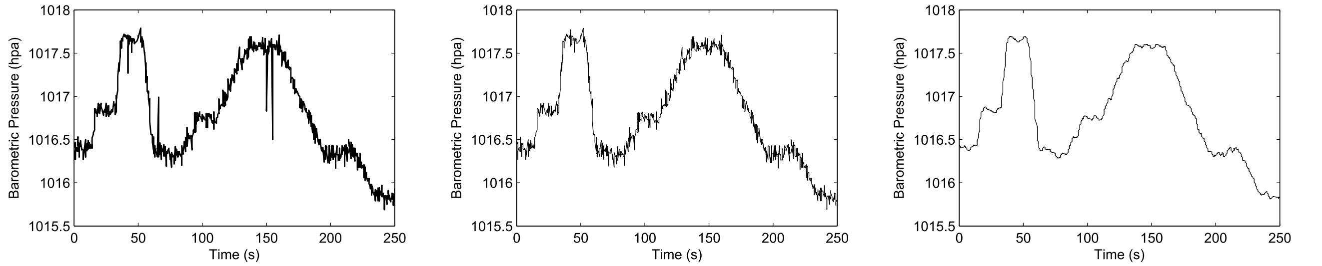
The barometer sensor data are collected continuously in the background with a sampling rate of 2 per second. A data sample is defined by $B = \{t, b\}$, where t is time, and b is the sensor reading at t. Fig. 2(a) shows an example of the raw barometer readings. For noise removing, we first remove the isolated points and high frequency parts of the data, which is done by a typical low-pass filter as shown in equation 2,

$$Y(n) = \beta X(n) + (1 - \beta)Y(n - 1) \tag{2}$$

where $X(n)$ is the nth barometer reading and $Y(n)$ is the output. β is the filter coefficient, and the value is set to 0.5 to make sure the filter result will be neither losing the original data signature nor leaving too much noise. Consequently, the cut-off frequency is 0.16 HZ. Fig. 2(b) shows the result curve. However, the data jitter is still obvious, we further process the data based on discrete wavelet transform using equation 3.

$$WT(\alpha, \tau) = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{+\infty} f(t) * \psi\left(\frac{t - \tau}{\alpha}\right) dt \tag{3}$$

where α is the scale and τ is the shift. The function ψ is the wavelet basis function, and we chose the Daubechies (db N) basis function here. After that, the values are smoothed with



(a) Original barometer readings

(b) After removed high frequency parts by low-pass filter

(c) Further processed based on discrete wavelet transform

FIGURE 2. **Barometer reading data preprocessing.**

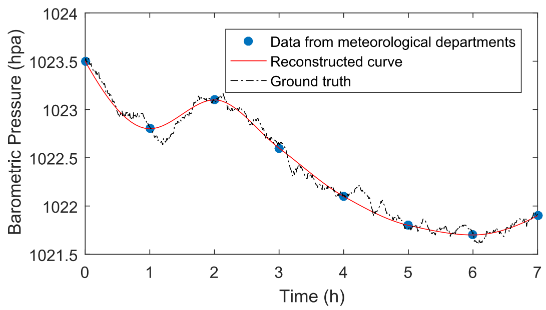


FIGURE 3. **Reconstruct the curve of the barometric pressure.**

a reasonable window size of 2 s,[1] and the final result is shown in Fig. 2(c). After data preprocessing, the signatures are much clearer and can help the moving mode detection algorithms to achieve better accuracy.

C. FILTERING NOISE CAUSED BY WEATHER

The barometer readings can be used to extract different altitude change signatures when user is arriving and leaving different locations. Before extracting these signatures, we need to remove the noise in the readings which is caused by the weather change. The solution is to catch the barometric pressure change that is only caused by weather, and then filter it out from the original barometer reading. However, it is a challenge to get the accurate weather caused barometric pressure change. In our solution, we obtain history meteorological data from open access services provided by meteorological departments. The bad news is that we can only get hourly updated data, which is unacceptable because the accumulative error can be tens of meters in an hour. In order to handle this problem, we try to construct the complete curve of the barometric pressure using curve fitting.

Figure. 3 shows a 7 hour pressure data by the blue points. There are only eight pressure readings, and the important pressure change data within every hour is missing. In order to reconstruct the missing data curve, after trying many solutions, we found out that the cubic curve is very suitable here. An example is shown by the red line in Fig. 3. Compared with the real pressure change data shown by the black dotted line, the two curves are very close to each other.

[1] i.e., the value at time t is the average value from t − 1 to t + 1 s

After that, we remove the weather caused noise by subtracting the weather caused barometric pressure change from the original barometer readings, and only the movement caused reading change is left. Based on this solution, we can reduce the weather noise to a large extent, especially when the real pressure changes gently. Without the weather noise, we can capture more accurate altitude change signatures.

D. MOVEMENT ACTIVITY DETECTION

In order to detect the frequent locations of a user based on barometer readings. Our solution is to first detect different user movement activities which happen in different type of locations. After that, it will be more easy to map the altitude change pattern to the location of the user. The movement activities are divided into three categories, which are: **1) Transportation activity**: That is the activity of a user taking motorized vehicles such as cars, trains and motorcycles in their daily life. When a user is doing a transportation activity, he must be in a physical path. The path is a kind of location. For example, the path of daily taking the subway to work, and the path of driving children to school, which are often fixed paths. **2) Indoor movement activity**: This means the movement activity of a user arriving and leaving a building. The building can be different locations, such as office, home and shopping mall,.etc. When a user is arriving and leaving a building, the altitude change patterns shown in the barometer readings are pretty different from that when taking a vehicle. **3) Outdoor physical activity**: This activity means a user moving outdoor in a non-motorized way. In this activity, the moving speed is much slower compared with the transportation activity, and the way of going up and down is different from the indoor movement activity. These differences are all reflected in the barometer reading data.

In daily life, all user's locations are related to the three movement activities. We collected a typical one hour barometer reading data of a user, and the black curve in Fig. 4 shows the result after data preprocessing and noise removing. The data are divided into different fragments and each belongs to a type of movement activity. For example, the first 8 fragments are transportation activities of taking the subway after work. The two green fragments are indoor movement activities of arriving and leaving the supermarket. The last 7 fragments

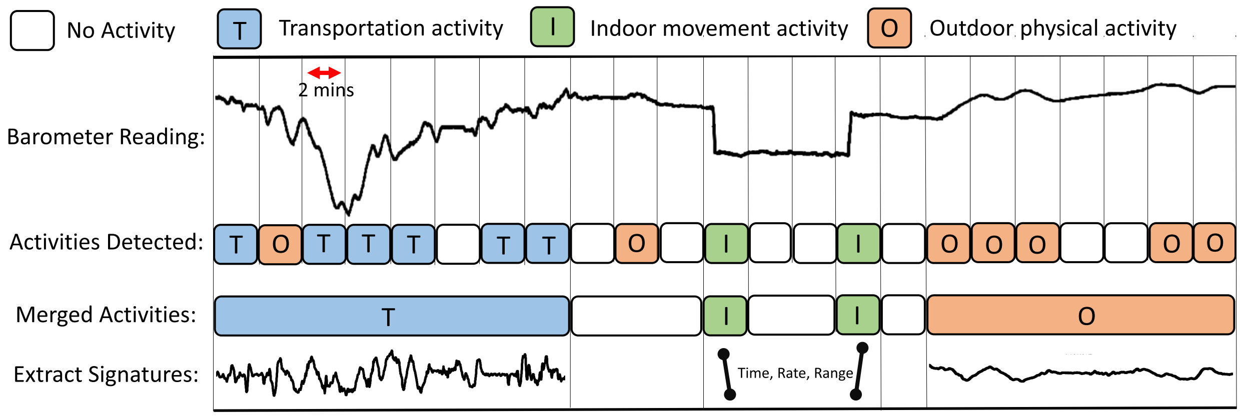


FIGURE 4. **Movement activity detection and signature extraction.**

are outdoor physical activities of having a walk in the park. The barometer readings show different change patterns for different movement activities. In our solution, we try to detect different movement activities based on these patterns.

SELoc is not a real time localization system, it provides historical location data for location based services. In this way, we can choose when and where to do the computation, and the computation cost is not restricted by the limited smartphone resources. Consequently, we are now able to consider much more complex models like deep learning. Motivated by the untapped potential of mobile deep learning combined with increased embedded systems' resources, we do the movement activity detection by using the deep learning algorithm called Restricted Boltzmann Machines (RBMs). Our movement activity detection process contains 3 phases. The first is sample extraction and sample representation. Next phase is about the RBM model, and finally is to solve the model. Here we show details in turn.

1) SAMPLE EXTRACTION AND REPRESENTATION

This is a data preparation phase. We first segment the continuous barometer reading stream and extract windows of sensor readings. The window width $w_d = 2$ minutes. For each window of sensor readings, we extract the representative features as input for the RBM model. Using these input features, deep learning can automatically find a good hierarchical representation of the sensor data. It's pre-training process helps to initialize the model parameter in an unsupervised manner. Therefore, it can decrease the requirement of large amount of labeled data [11]. In detail, we extract time and frequency domain features. The time domain features include standard deviation, range and zero-crossing rate (ZCR). The frequency domain features include the mean and standard deviation of amplitude from power spectral density (PSD).

2) RESTRICTED BOLTZMANN MACHINE

A RBM has binary-valued (Boolean/Bernoulli) hidden and visible units. It is represented by an undirected bipartite graph. The graph is consisted by a set of stochastic visible units $v \in \{0, 1\}^m$ and a set of stochastic hidden units $h \in \{0, 1\}^n$. A matrix of weights $W = (w_{i,j})$ (size $m \times n$) associated with the connection between visible unit v_i and hidden unit h_j, as well as bias weights (offsets) a_i for the visible units and b_j for the hidden units. The example is shown by the model in Fig. 1. The energy function $E : 0, 1^{m+n} \rightarrow R$ associated with a RBM model is given as:

$$E(v, h; \Theta) = -\sum_{i=1}^{m}\sum_{j=1}^{n} v_i W_{ij} h_j - \sum_{i=1}^{m} a_i v_i - \sum_{j=1}^{n} b_j h_j \quad (4)$$

where Θ is the model parameters and is defined by $\Theta = \{W, a, b\}$. The joint probability distributions over hidden and visible units are defined in terms of the energy function, and the function is given as:

$$P(v, h; \Theta) = \frac{1}{Z(\Theta)} e^{-E(v,h;\Theta)} \quad (5)$$

where $Z(\Theta)$ is a partition function defined as the sum of $e^{-E(v,h;\Theta)}$ over all possible configurations. That is, for m visible units and n hidden units, the conditional probability of a configuration of the visible units v, given a configuration of the hidden units h, is:

$$P(v \mid h) = \prod_{i=1}^{m} P(v_i \mid h) \quad (6)$$

and the conditional probability of h given v is

$$P(h \mid v) = \prod_{j=1}^{n} P(h_j \mid v) \quad (7)$$

For Gaussian RBMs, the visible units are the real valued data measured by barometer sensors. The hidden units are binary. The Θ is defined by $\Theta = \{W, a, b, \sigma\}$, where σ is the standard deviation. In this case, the energy function is:

$$E(v, h; \Theta) = -\sum_{i=1}^{m}\sum_{j=1}^{n} \frac{v_i}{\sigma_i} W_{ij} h_j - \sum_{i=1}^{m} \frac{(v_i - a_i)^2}{2\sigma_i^2} v_i - \sum_{j=1}^{n} b_j h_j \quad (8)$$

3) SOLVING THE MODEL

It is not trivial to get an exact solution for Equation 8. In this paper, we make use of efficient Markov Chain Monte Carlo (MCMC) based stochastic approximation techniques to estimate the expected statistics of the model [12]. For initializing the model parameters, we use a greedy layer-wise pre-training before performing back-propagation using the labeled data. It's important to notice that, although the training process is computation-heavy task, it can be done offline, and is once and for all. After training, the parameters of the model can be confirmed and can be used for movement activity recognition by SELoc for every user.

E. DETECT LOCATION BASED ON MOVEMENT ACTIVITY

After detecting different movement activities, now it is easier to detect the location of the user. For example, for an indoor movement activity, the user must be in a building such as home, office or shopping mall,.etc. The search scope is greatly reduced. Similar situation holds for the transportation and outdoor physical activities. In this paper, our solution to detect the location is based on the following observations. 1) The barometer readings show different altitude change patterns when the user appears at different frequent locations in their daily life. This observation helps us to differentiate different locations. 2) Users arrive at their frequent locations with regularity. We can infer the location by the prior knowledge and intuitive logic. Here is an example, after analyzing one week of barometer reading data from a user, we find he arrives at the same location every night until next day, and we can infer that the location is his home with high probability.

Our location detection pipeline spans 3 phases: 1) activity data pre-processing, 2) clustering locations by signatures, and finally 3) label the clusters with locations. We now describe each of them in turn.

1) ACTIVITY DATA PRE-PROCESSING

After activity detection based on RBM, it outputs activity results for every 2 minutes' data window. For indoor activities, one data window time is long enough to contain the activity of the user arriving or leaving the building. However, the transportation and outdoor activities usually contain one or multiple data windows, and we need to merge data windows that belong to the same activity. As shown in Fig. 4, the transportation contains 8 data windows. Among them, there are 2 wrong activity results and 6 correct activity results. Here, we propose a merge algorithm to merge data windows of the same transportation activity or outdoor activity. The basic idea is that these activities are consequent, other activity windows that appear sporadically inside can be treated as wrong detections. The detail is shown in Algorithm 1, it merges the transportation activity windows if there are no more than one continues other kind of activities among them. For easy understanding, Fig. 4 also shows the graphic example of a merged transportation activity and an outdoor activity.

Algorithm 1 Merging Transportation Movement Activities

Input: The time-series activities of all windows S.
Output: The set A which contains all merged movement activity M;

```
Initialization: new empty set A;
Initialization: new empty list M;
Initialization: new int i = 0;
for each activity window s ∈ S do
    if i > 1 and list M is not empty then
        add list M to set A
        new empty list M, i = 0
    else
        if s is a transportation activity then
            add s to M, i = 0
        else
            i = i + 1
            if i = 1 and list M is not empty then
                add activity window s to list M
            end if
        end if
    end if
end for
return Movement activity set A
```

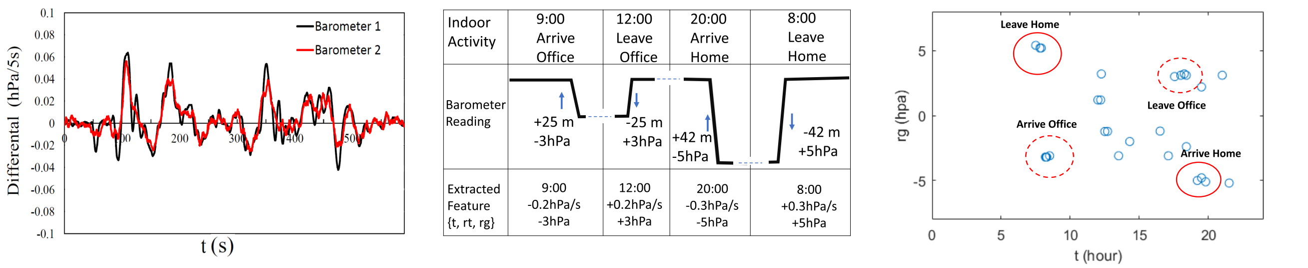
2) CLUSTERING LOCATIONS BY SIGNATURES

The barometer reading show different altitude change patterns when the user moving at different locations, especially these frequent locations. In this step, we find out signatures of different frequent locations using the CURE [13] based clustering algorithm. We extract altitude change pattern features from movement activities and then design distance functions for three types of activities separately.

For the transportation activities, every time when a user moves in the same path, the office-home path for example, the time-series barometer readings will have strong correlations. In another word, user altitude change in the same way. The common approach is to use the absolute barometer reading as the feature and the distance function is to calculate the mean squared error (MSE). This is not applicable here because the absolute readings can be different every time, so does the length of the time-series. Here, the differential of the barometer reading time-series is more suitable to be used as the feature, and the Dynamic Time Warping Distance Measure (DTW) [14] is more suitable as the distance function. Fig. 5(a) shows an example of the extracted differential feature of two transportation activities data in the same path. The correlations are very clear when aligning the two time-series using the DTW distance measure.

To calculate the DTW, the approach is to first align the time-series. For example, as shown in Fig. 5(a). For the barometer differential features of barometer 1 and barometer 2 are D_s and D_r, where

$$D_s = s_1, s_2, s_3, \ldots s_k$$
$$D_r = r_1, r_2, r_3, \ldots r_l$$



(a) Example of two time-series using the DTW distance measure (b) Feature extraction for indoor movement activity (c) Example of indoor movement activity features in four working days

FIGURE 5. **Extract movement activity features.**

the sequences D_s and D_r can be arranged to form a k-by-l grid, where each $point(i, j)$ is an alignment between s_i and r_j. A warping path, W, aligns the elements of D_s and D_r.

$$W = w_1, w_2, w_3, \ldots w_k$$

The DTW distance between D_s and D_r is then:

$$DTW(D_s, D_r) = \partial(F(D_s), F(D_r)) + min \begin{cases} DTW(D_s, R(D_r)) \\ DTW(D_r, R(D_s)) \\ DTW(R(D_s), R(D_r)) \end{cases} \tag{9}$$

where $F(x)$ is the first element of x, and $R(x)$ is the rest except for $F(x)$, and $\partial(i, j) = (s_i - r_j)^2$. The situation for the outdoor physical activities are similar to the transportation activities, and we use the same DTW distance based solution here.

For the indoor activities, every time when a user arriving and leaving the same building, the barometer sensor data show similar altitude change patterns. For example, as shown in Fig. 5(b), the user takes an elevator and goes up for exactly 25 meters to his office in the 6th floor, and that will cause about 3 hpa decrease of the barometer pressure reading. The features of every indoor activity are extracted as a triple $V = \{t, rt, rg\}$, where t is the central time point of the indoor activity, and rt is the barometer change rate, and rg is the range of barometer reading change, which can also be transferred to altitude change. The rg can be positive value or negative value. The negative value means the user is going up and arriving a building, and the positive value means the user is going down and leaving a building. An example of extracting the features is shown in Fig.5(b). The distance function is defined as a Weighted Euclidean Distance as follows:

$$d(V_i, V_j) = \sqrt[3]{\alpha_1(t_i - t_j)^2 + \alpha_2(rt_i - rt_j)^2 + \alpha_3(rg_i - rg_j)^2} \tag{10}$$

where $\alpha_1, \alpha_2, \alpha_3$ are weights of each feature, and $\alpha_1 + \alpha_2 + \alpha_3 = 1$. The values of t, rt, and rg are all normalized, and the absolute value is not greater than 1. Figure. 5(c) shows the indoor activity feature data of 4 working days from the same user. For clarity, we only show the features of time t and barometer reading change rg. We can see from the figure that the points of the user arriving and leaving the same building are gathered together, such as home and office.

The CURE algorithm cannot be directly applied because the resulting number of clusters is unknown. In our solution, the clustering stops when the distance between every cluster above a certain threshold. After that, we find clusters with different signatures, and belongs to different locations. However, the label of the locations are still unknown. In the next step, we will infer the location by labeling the clusters.

3) LABEL THE CLUSTERS WITH LOCATIONS

With the help of the clustering algorithm, we can extract altitude change patterns (signatures) of the user's frequent locations. However, this is an unsupervised learning process, and we need to label the clusters with their locations. In other words, we need to map the clusters to locations. In our solution, for the common locations such as home, office, and home-office path, we label them by defining inferring rules based on the prior knowledge and intuitive logic. For other locations, we provide a summary of space-time characteristics of the cluster and request the users to label them.

Observation 1: the location that a user appears frequently for hours at working time is very likely to be his work place/office.

Rule 1: For indoor activity clusters. If there exists two clusters, the first cluster contains the activities of arriving a building at working time, and the second contains the activities of leaving a building at working time. There are hours between arriving and leaving time. The building is very likely to be his work place/office.

Formally, given that

1) P_1: $\overline{V_1}.t \in$ "working time" and $\overline{V_2}.t \in$ "working time";
2) P_2: $\overline{V_1}.rg < 0 \wedge \overline{V_2}.rg > 0$;
3) P_3: $\left| \frac{|\overline{V_1}.rg| - |\overline{V_2}.rg|}{|\overline{V_1}.rg| + |\overline{V_2}.rg|} \right| <$ "minimum difference ratio";
4) P_4: $\overline{V_2}.t - \overline{V_1}.t >$ "minimum working time" ;
5) A_1: The clusters of $\overline{V_1}$ and $\overline{V_2}$ are activities at work place/office.

$$R_1 : P_1 \wedge P_2 \wedge P_3 \wedge P_4 \rightarrow A_1.$$

where $\overline{V} = \{\overline{t}, \overline{rt}, \overline{rg}\}$, which are the average feature values of an indoor activity cluster. The signatures of the detected work place/office location is extracted as $V_{office} = \{\overline{V_1}.rg, \overline{V_2}.rg, \overline{V_1}.rt, \overline{V_2}.rt, UserID\}$. $UserID$ is the ID of the user, and every location signature is related to a certain user. We use similar inferring rules for other building locations such as home and restaurants. Some special cases cannot be handled here. For example, if the user is not having a common work habit as most other users, the system will provide a interface for the user to custom their own rules.

Observation 2: the path that a user appears frequently before and after working is very likely to be the home - office path.

Rule 2: For transportation or outdoor physical activity clusters. If there exists two clusters, the first cluster happens before working time, and the second happens after working time. One altitude change pattern is matched with the reverse of the other. The moving path is very likely to be the home - office path.

Formally, given that

1) P_5: $\overline{D_1}.t \in$ "before working time" and $\overline{D_2}.t \in$ "after working time";
2) P_6: $|DTW(D_1, Reverse(D_2))| <$ "minimum difference";
3) P_7: $\left|\frac{|\overline{D_1}.dur| - |\overline{D_2}.dur|}{|\overline{D_1}.dur| + |\overline{D_2}.dur|}\right| <$ "minimum difference ratio";
4) A_2: The clusters of $\overline{D_1}$ and $\overline{D_2}$ are activities happen in home - office path.

$$R_2 : P_5 \wedge P_6 \wedge P_7 \rightarrow A_2.$$

where $\overline{D} = \{\overline{t}, \overline{dur}\}$, which are the average happen time and time duration of a transportation or outdoor physical activity cluster. The signatures of the path location is extracted as $D_{home-office} = \{\overline{D_1}, \overline{D_2}, UserID\}$.

Observation 3: some locations that a user appears periodically need the user to label them.

Rule 3: For indoor activity clusters. If there exists two clusters, the first cluster contains the activities of arriving a building, and the second contains the activities of leaving a building. The altitude change of arriving and leaving is the same. The building is very likely to be a frequent location and the system requests the user to label it.

Formally, given that

1) P_8: $\overline{V_1}.rg < 0 \wedge \overline{V_2}.rg > 0$;
2) P_9: $\left|\frac{|\overline{V_1}.rg| - |\overline{V_2}.rg|}{|\overline{V_1}.rg| + |\overline{V_2}.rg|}\right| <$ "minimum difference ratio";
3) P_{10}: $\overline{V_2}.t - \overline{V_1}.t >$ "minimum staying time" ;
4) A_3: The clusters of $\overline{V_1}$ and $\overline{V_2}$ are activities at the same building and need to be labeled by the user.

$$R_3 : P_8 \wedge P_9 \wedge P_{10} \rightarrow A_3.$$

where $\overline{V} = \{\overline{t}, \overline{rt}, \overline{rg}\}$, which are the average feature values of an indoor activity cluster. The signatures of the location are extracted as $V_{label} = \{\overline{V_1}.rg, \overline{V_2}.rg, \overline{V_1}.rt, \overline{V_2}.rt, UserID\}$. When detecting this location, SELoc will request the user to label it.

In this way, we can provide user non-sensitive location data for location based services. After labeling the clusters with locations, we get the signatures of every location. Later, when providing more history location data, we only need to use a kNN based classification algorithm to find the location, and the training process do not need to be done every time. The training process will start again when there are new frequent locations appeared in this user's daily life.

III. EVALUATION

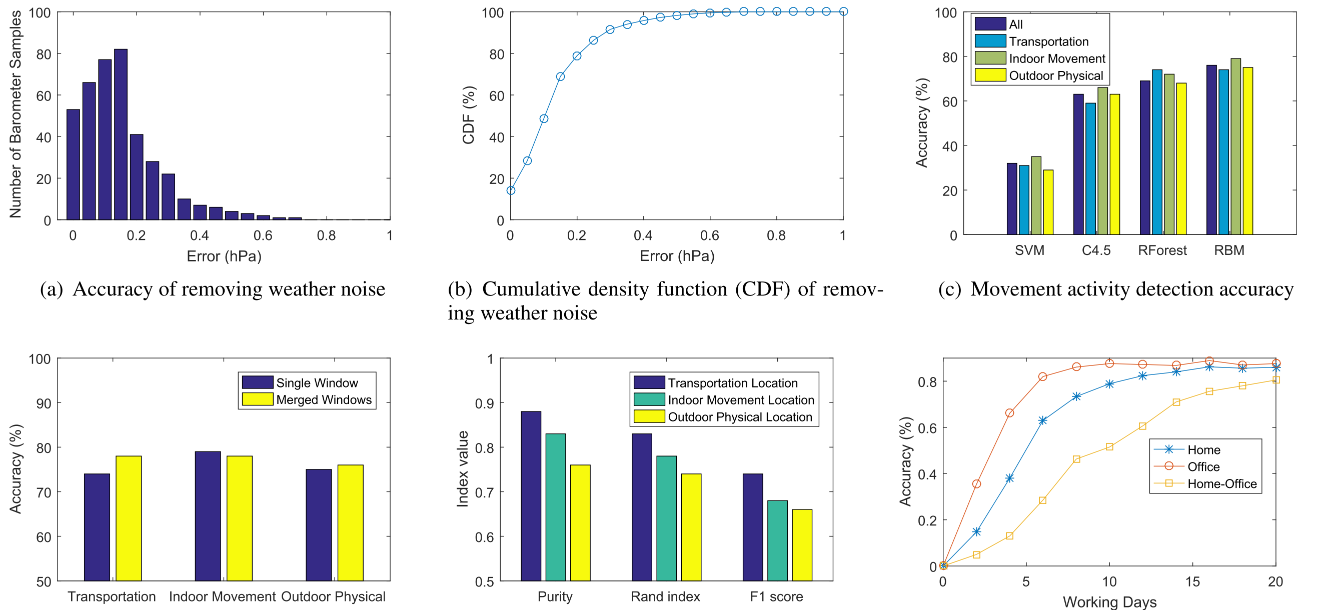
We proposed several field studies to evaluate the performance of SELoc. Mainly divided into two parts. The first part is to evaluate the techniques proposed in SELoc based on a data-set which is continuously collected for two weeks from three users. The next part is to compare SELoc with three existed typical solutions, the GPS-based location system, a wireless fingerprinting based approach [15], and a big data analysis based solution.

A. EVALUATE THE PERFORMANCE OF SELOC

The performance analysis is done off-line based on a barometer reading data-set and its ground-truth. Three participants installed our data collecting tool in their smartphones. The tool keeps on collecting barometer data in the background with a rate of 2 samples per second. To record the ground-truth, the tool collected the GPS sample every 5 minutes, and the participants were also responsible for writing done his movement activity history. The tool can remind the user to record their movement activity and location every hour at day time. We do not give special instructions to control their behaviors during the study, instead, all the users are told to perform their daily routines. The data collection process lasted for two weeks. After the data collection process, with the help of the GPS and user notes, we get the movement activity and location ground-truth of every user during the two weeks. Later, the raw barometer data readings are used as input of SELoc system. The intermediate and final outputs of SELoc are then used to compare with the gourd-truth, for evaluating their performance. We realized SELoc as a Java API, it can be easily integrated into many applications.

1) FILTERING WEATHER NOISE

The weather noise is the major factor that affects the movement activity detection accuracy. Our way to measure the filtering performance is to compare it with the real pressure change caused by weather. We deployed an electronic barometer sensor at the root of an office building in the field study area to record the real weather caused pressure change. To compare the estimated pressure change with the real pressure change, we randomly selected 400 barometer reading data points, and the error is measured by the difference of the readings. The result is shown in Fig. 6(a), and the CDF of the accuracy is shown in Fig. 6(b). We can see from the figures that the error is less than 0.25hPa for about 80% of the cases. Compared to the original error caused by weather



(a) Accuracy of removing weather noise

(b) Cumulative density function (CDF) of removing weather noise

(c) Movement activity detection accuracy

(d) Accuracy after activity merging

(e) The clustering performance

(f) Accuracy of labeling locations

FIGURE 6. Performance results of the field study.

such as 1.2hPa in half an hour, our solution of filtering the weather noise is a great improvement.

2) MOVEMENT ACTIVITY DETECTION

We now detail the movement activity detection performance based on RBM. We use the data-set collected in the field study to train the model. The movement activity ground-truth are already known, and the accuracy is calculated using 5-fold cross validation. As a reference, we also compare the accuracy with 3 well-known classifiers using the same input features. These classifiers are based on SVM, C4.5 and Random Forest. The accuracy of these algorithms for the movement activity detection, including that of our RBM solution, are shown in Fig. 6(c). For the accuracy, the SVM based classifier was worst, which is about 32%. C4.5 decision tree based algorithm has an accuracy of about 63%, which is much better. The random forest classifier achieves an accuracy of about 69%. However, the RBM based classifier performed best, and the accuracy is 76%. This order still holds when we compare the accuracy of any type of movement activities. For different movement activities, the indoor activity has a better accuracy than the other two movement activities. The detail accuracy result is shown in Table 2. For example, the 16% transportation activities are wrongly detected as indoor activities, 4% are wrongly detected as outdoor physical activities, and 6% are miss-detected.

3) MERGING ACTIVITIES

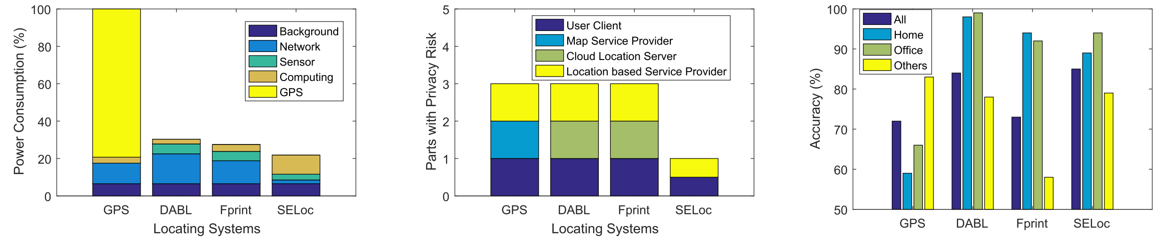
The transportation and outdoor physical activities usually contain one or multiple data windows, and they need to be merged. In the previous section, we evaluated the movement detection accuracy by every data window of 2 minutes. Here, in our merging algorithm, a few wrong detections of data windows sometimes will not affect output. For example, one false indoor activity window appears among transportation windows can be filtered. In this case, the final accuracy of merged windows could be better than the single window accuracy shown in the previous section. In order to verify that, Fig. 6(d) shows the activity accuracy after merging, compared with the accuracy of single window. The accuracy of transportation activity improved a little from 74% to 78%, and outdoor physical activity improved from 75% to 77%.

TABLE 2. Movement activity detection accuracy.

Activity Type	Transportation	Indoor Movement	Outdoor Physical	No Activity
Transportation	74%	16%	4%	6%
Indoor Movement	17%	79%	2%	2%
Outdoor Physical	5%	2%	75%	18%
No Activity	7%	6%	21%	66%

4) LOCATION CLUSTERING AND LABELING

Frequent locations are found by clustering based on different signatures of the same movement activity. To measure the clustering performance, we make use of indexes of purity, Rand index, and F1 score. The purity measures the percent of right clustered samples of the total sample. The Rand index and F1 score are common measures of accuracy in data clustering. Figure. 6(e) shows the accuracy result of clustering the locations of the three types of movement activities. Most samples can be clustered to the right cluster. For indoor buildings,



(a) Total power consumption and estimated ratio of different power consumption sources
(b) Privacy risk of the systems
(c) Locate accuracy of the systems

FIGURE 7. **Evaluation results when comparing with the related works.**

the false positive result often happen among positions with similar altitudes. For example, user A's home and office are in the same floor and same altitude, and this may cause false positive results. However, this doesn't happen often because the indoor locations of a user often have different altitudes. We also evaluated the accuracy of using the inferring rules to label the clusters. The result is shown in Fig. 6(f). It contains inferring locations of home, office and home-office path. The accuracy is related to the days of data. Longer days of data will lead to better accuracy.

B. EVALUATE SELOC WITH EXISTED WORKS

In this section, we compare the performance of SELoc with three existed approaches.

1) *The Traditional GPS Based Localization:* GPS is the common and most used solution for outdoor localization. It does not work well indoors because the signal can't pass through the wall. To find the location of a user, the way is to get the GPS coordinates and then map them to the location using a map service. This can be done easily by developing a small application program in the android platform.
2) *The Wireless Fingerprint Based Approach (FPRINT) [15]:* It's a typical way for location data collection for indoor environments. The way is to locate the user by fingerprint mapping, such as the Wi-Fi and Bluetooth fingerprint. A fingerprint map should be built to map the fingerprint to the location, which is often located in a cloud server. It can locate a user both at indoor and outdoor locations. We implemented a prototype system based on patent US20140012529 [15].
3) *The Big Data Analysis Based Locating (DABL) [16]:* The DABL system covers the methods for locating the user by using big data analyzing, such as [17] and [18]. A user produces large amounts of data in the daily life, and these data can be used to infer the location of the user. For example, the office computer login information can be used to infer that when the user is working at office, and find the user at home when there is a connection between smartphone and equipments at home. The car parking information can also infer the location of the user.

In order to evaluate the performance of these solutions, we developed a prototype application which can provide user history location data based on all the four approaches, and evaluated them under the same scenario. Two participants take four nexus 6 smartphones, and running the four systems respectively. The experiment lasted for two weeks, the participant is responsible for writing done his location history. The requirement for these systems is to provide all the locations and stay time history of the user in the past two weeks. During the experiment, the location output and power file are logged on the disk. After the test, we compare these solutions in different aspects.

1) ENERGY EFFICIENCY

First, we compare the energy consumption in user smartphones. Measurements were performed on Nexus 6 using the Monsoon Power Monitor. Fig. 7(a) shows the comparison of the power consumption of the four systems, and the simulation results when we break-down the power consumption sources of these systems. The power consumption sources of SELoc are 1) collect 2 barometer readings per second, 2) and access the history weather service on internet for barometric pressure data, and 3) run SELoc algorithms to calculate the location of the user. The main power consumption here is computing. The energy consumption of SELoc is much less than the GPS based approach. The GPS sampling has high energy consumption, and it also needs to access the map service through the network to infer user's location. The DABL system needs to collect and upload the different kinds of user data to the cloud server, where running the big data analysis algorithm to calculate user location. So, it costs more energy at network communication. The same situation happens with the Fprint system, which needs to upload the Bluetooth or Wi-Fi fingerprinting to the cloud server for locating. Another fact is that many users are not used to keep Wi-Fi or Bluetooth on all the time, and continuous location data collection cannot be easily done.

2) ACCURACY

Fig. 7(c) shows the location accuracy of the four systems. The average accuracy of GPS is about 73% in the provided location data. Since GPS accuracy is poor indoors, the accuracy of locating the user at home and office is lowest compared with all other systems. The DABL system and Fprint system

TABLE 3. The comparison of the limitation factors of the systems.

Limitation Factor	SELoc	GPS	Fprint	DABL
Training	■	□	■	■
Infrastructure Support	□	□	■	■
Map Support	□	■	■	□
Server Support	□	□	■	■
Outdoor Only	□	■	□	□

performs better when the user is at home and at the office. However, the Fprint accuracy is poor at other locations when there are few Bluetooth or Wi-Fi infrastructure coverage. The DABL performs better but it's still very hard to get such large amount of multidimensional user data, because the available data collection methods are very limited today. The overall accuracy of SEloc is no worse than other three systems, and the average accuracy is 85%.

3) PRIVACY

In Fig. 7(b), we show the privacy risk parts of leaking the sensitive user location data during the whole process of getting a location based service. For the GPS system, the sensitive GPS data can be leaked at the user client, the map service provider and the location based service provider. For the Fprint and DABL system, the sensitive location data can be leaked at the user client, cloud location server and the location based service provider. However, SELoc only provide non-sensitive location data, and the non-sensitive location data can only be leaked at the user client and the location based service provider, which may cause much less privacy problem.

4) LIMITATION FACTORS

Moreover, SELoc is also better for having the least number of limitation factors, which makes SELoc more practical and easy to realize. For detail, we show the comparison in Table 3. For example, SELoc does not need infrastructure support. The Fprint and DABL system need to deploy extra equipments and platforms. They also need a cloud server to support localization, which will cost more energy and resources. The GPS system needs a map support and works poor indoors. SELoc overcomes other approaches except for that it needs a training process to label the locations.

IV. ONLINE SURVEY

The idea of SELoc originated from the authors' experience when trying to design and implement a location based service for university students. We try to obtain students' behavioral habits on campus based on the location data, which can be used to evaluate whether their school life is healthy or not, and accordingly providing related services. With the help of the location based technologies, such as WiFi fingerprinting and GPS, we can get students' location on campus. However, the project was failed to carry out. One reason is that the power consumption is extremely high due to the continuous sampling. However, the main reason is that the students refused it because of the privacy concern, they can't accept that their location is continuously monitored. That's why we try to propose a new technology to avoid using sensitive location data. We believe the user's opinion is very important for us to design such a location data collecting system. Consequently, carried out an online survey using a mobile testing center named BaiduMTC. The survey stopped until totally 600 valid responses were received. In the survey, we first ask the users to choose a characteristic based on their usage frequency of the smartphones. Based on which, the participants are firstly labeled as enthusiast (18%), follower (48%) or ordinary (34%).

A. HOW OFTEN DO USERS USE THE LOCATION BASED SERVICE?

Based on the survey result, totally 56% users agreed that location based services are indispensable and they use them very often today. Other 33% users said they use them sometimes. Still 11% users said they use them rarely. For the ordinary users, 75% use location based services often, their most used location based services are the map applications. They mainly use it for car and pedestrian navigation. About 82% followers are interested in more location based services, especially the emerging new applications. The enthusiasts use it most often compared to others, they are interested and glad to try any new location based services.

B. DO USERS HAVE PRIVACY CONCERN WHEN PROVIDING THE LOCATION INFORMATION?

When asked whether they worry about the privacy problem, 67% ordinary users said they care about the privacy problem. The results were similar in enthusiasts, which is 63%. However, the followers care about the privacy problem most, about 89% followers care about the privacy problem. In all respondents, 82% of them agreed the privacy problem is the biggest obstacle now that prevents them from using the location based services.

C. DO USERS EXPECT FOR NEW TECHNOLOGIES AND WHAT ARE THEIR REQUIREMENTS FOR THESE TECHNOLOGIES?

We found that users care about their privacy when using every application, if the risk of privacy breaches is high, it will cause a great influence on its usability. According to our survey, in public places, 71% participants will accept a new technology instead of providing his sensitive location to get the location based service. In private places such as home and work places, the percent increased to 92%. Besides the privacy requirements, users also care about the energy consumption and convenience of the technology. In public places, 68% participants will refuse the new technologies if it cost more energy or not easy to use. Even in private places, the percent still have 34%. That means users are very sensitive to energy consumption.

Taking survey responses as ordinal values, we computed the correlations between these responses from different kinds of users. Statistically significant positive correlations were found. 1) Users who rarely use location based services are most likely to be an ordinary user. 2) The followers care about privacy most. 3) Enthusiasts are willing to try new technologies instead of providing location information. This means that enthusiasts and followers are precisely those who need SELoc more.

In summary, we obtained three key observations from the online survey. First, the location based service is a basic requirement for users daily life. Second, user's requirement about privacy is not well satisfied by current technologies. Finally, users are very likely to accept a new technology instead of providing his location to get the location based service if it can protect the privacy, and is energy efficient. The above results confirmed users's desirability for the new technology, and further motivated us to propose a new solution to fulfill users' requirements.

V. RELATED WORK

Traditional localization technologies make use of GPS and the Cellular network [19] to locate the user indoors and outdoors. However, the drawbacks are obvious. When we need large number of history user location data for the location based services, the power consumption is pretty high [3]. Although researchers try to save energy by reducing the GPS sampling rate [20], [21], still it affects the accuracy. The accuracy can also be affected by obstructions such as walls and trees in the city area. More important, these systems provide the physical coordinates of the user, the privacy can be leaked when uploaded to the map service provider or the location service provider. In SELoc, we get the non-sensitive location directly instead of the coordinates, and only sensing the barometer data for localization, which is more energy efficient and safe.

For better energy efficiency, researchers try to utilize some new localization techniques. PlaceLab [22] is one of the earliest solutions which can track devices using wireless signal in the wild. Some solutions locate by wireless fingerprints [23], such as Wi-Fi fingerprints and Bluetooth fingerprints [4], [24]. These systems need an off-line process to build the fingerprint map before locating the user, and also need the wireless infrastructure support. This process needs a large amount of human work and cost a lot. SELoc do not depend on these infrastructures and the total cost is very low. In these systems, the location is done in the cloud server where the fingerprint map is located. The user client should upload the wireless fingerprints to query the location. The sensitive location related data may also cause privacy problems. Other techniques such as WheelLoc [25] only makes use of low power sensors and cell tower information, but the accuracy is not good enough compared to SELoc. Another commonly used technology is dead reckoning, it is the process of calculating the next position by a previously determined position. They often make use of the accelerometer to detect the step number of the user [26]. However, the accuracy cannot be always ensured and it needs cooperation with other devices.

Recently some solutions based on big data technology are proposed [16]–[18]. They infer the location of the user based on large amounts of data related to the user. In these solutions, there exists a could server running the big data platform which can analyze the data uploaded to locate the user. However, the user needs to upload more data and the cost to deploy the platform is not cheap. Solutions based on big data technology also have privacy problems [27], [28]. The sensitive user data have privacy risk in the cloud server.

The barometers has been used to assist indoor localization by inferring the altitude [10] and floor level [29], [30] of the user. Our previous work [31] make use of the barometer and accelerometer for outdoor tracking in mountain roads. The difference is that the system is designed for realtime location tracking instead of providing history location data. As far as we know, SELoc is the first system which mainly uses the barometer readings to infer and provide user history location data, and designed for the location services based on numerous historical location data.

VI. CONCLUSION AND DISCUSSION

In our work, we demonstrate the approach for a new location data collecting system for location-based services. First, SELoc only uses the low power barometer and senor on smartphone, which makes the system very energy efficient, only 22% of GPS. Then, the accuracy of 85% is no worse than other technologies, and the way of only providing the non-sensitive location data gains more privacy protection. At last, SELoc has minimal limitation factors compared with other location systems, which is very practical for real usage. Compared to the existing works, we find that SELoc is especially suitable for the location services based on numerous historical location data. However, SELoc still has some limitations. For example, SELoc needs the user to label some locations, and the non-sensitive location data is not suitable for some kinds of location based services. We will try to propose solutions which can provide safe and energy efficient location data collecting for other location-based services in our future work.

REFERENCES

[1] *Location-Based Services (LBS) and Real-Time Location Systems (RTLS) Market by Location Type, Software, Hardware, Service, Vertical, and Region—Global Forecast to 2023*, Res. Markets, Bus. Wire, San Francisco, CA, USA, 2018.

[2] X. Zhu, Q. Li, and G. Chen, "APT: Accurate outdoor pedestrian tracking with smartphones," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 2508–2516.

[3] M. B. Kjaergaard, J. Langdal, and T. Godsk, "Entracked: Energy-efficient robust position tracking for mobile devices," in *Proc. 7th Int. Conf. Mobile Syst., Appl., Services*, Jun. 2009, pp. 221–234.

[4] A. Zhang, Y. Yuan, Q. Wu, S. Zhu, and J. Deng, "Wireless localization based on rssi fingerprint feature vector," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 11, Nov. 2015, Art. no. 528747.

[5] K. Dong, W. Wu, H. Ye, M. Yang, Z. Ling, and W. Yu, "Canoe: An autonomous infrastructure-free indoor navigation system," *Sensors*, vol. 17, no. 5, p. 996, Apr. 2017.

[6] J. Teng, B. Zhang, J. Zhu, X. Li, D. Xuan, and F. Yuan Zheng, ''EV-Loc: Integrating electronic and visual signals for accurate localization,'' *IEEE/ACM Trans. Netw.*, vol. 22, no. 4, pp. 1285–1296, Aug. 2014.

[7] K. Vu, R. Zheng, and J. Gao, ''Efficient algorithms for k-anonymous location privacy in participatory sensing,'' in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2399–2407.

[8] R. Shokri, G. Theodorakopoulos, and C. Troncoso, ''Privacy games along location traces: A game-theoretic framework for optimizing location privacy,'' *ACM Trans. Privacy Secur.*, vol. 19, no. 4, p. 11, Feb. 2017.

[9] G. Liu, M. Iwai, Y. Tobe, D. Matekenya, K. M. A. Hossain, M. Ito, and K. Sezaki, ''Beyond horizontal location context: Measuring elevation using smartphone's barometer,'' in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.: Adjunct Publication*, Sep. 2014, pp. 459–468.

[10] H. Ye, K. Dong, and T. Gu, ''Himeter: Telling you the height rather than the altitude,'' *Sensors*, vol. 18, no. 6, p. 1712, 2018.

[11] S. Bhattacharya, P. Nurmi, N. Hammerla, and T. Plötz, ''Using unlabeled data in a sparse-coding framework for human activity recognition,'' *Pervas. Mobile Comput.*, vol. 15, pp. 242–262, Dec. 2014.

[12] R. Salakhutdinov and G. Hinton, ''Deep boltzmann machines,'' in *Proc. Artif. Intell. Statist.*, pp. 448–455, Apr. 2009.

[13] K. G. S. Shim and R. Rastogi, ''Cure: An efficient clustering algorithm for large databases,'' *Inf. Syst.*, vol. 26, no. 1, pp. 35–58, Mar. 2001.

[14] D. J. Berndt and J. Clifford, ''Using dynamic time warping to find patterns in time series,'' in *Proc. KDD Workshop*. Seattle, WA, USA, Jul. 1994, pp. 359–370.

[15] J. Zhu, K. Zeng, P. Mohapatra, and K. H. Kim, ''Bluetooth beacon based location determination,'' U.S. Patent 8 965 398, Feb. 24, 2015.

[16] M. V. Moreno, F. Terroso-Sáenz, A. González-Vidal, M. Valdés-Vela, A. F. Skarmeta, M. A. Zamora, and V. Chang, ''Applicability of big data techniques to smart cities deployments,'' *IEEE Trans. Ind. Inform.*, vol. 13, no. 2, pp. 800–809, Apr. 2017.

[17] F. Zhu, C. Luo, M. Yuan, Y. Zhu, Z. Zhang, T. Gu, K. Deng, W. Rao, and J. Zeng, ''City-scale localization with telco big data,'' in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2016, pp. 439–448.

[18] K. Lin, J. Luo, L. Hu, M. S. Hossain, and A. Ghoneim, ''Localization based on social big data analysis in the vehicular networks,'' *IEEE Trans. Ind. Inform.*, vol. 13, no. 4, pp. 1932–1940, Aug. 2017.

[19] M. Ibrahim and M. Youssef, ''A hidden Markov model for localization using low-end GSM cell phones,'' in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2011, pp. 1–5.

[20] H. S. Ramos, T. Zhang, J. Liu, B. N. Priyantha, and A. Kansal, ''Leap: A low energy assisted GPS for trajectory-based services,'' in *Proc. Int. Conf. Ubiquitous Comput.*, pp. 335–344, Sep. 2011.

[21] Y. Chon, E. Talipov, H. Shin, and H. Cha, ''SmartDC: Mobility prediction-based adaptive duty cycling for everyday location monitoring,'' *IEEE Trans. Mobile Comput.*, vol. 13, no. 3, pp. 512–525, Mar. 2014.

[22] A. LaMarca, Y. Chawathe, S. Consolvo, J. Hightower, I. Smith, J. Scott, T. Sohn, J. Howard, J. Hughes, F. Potter, J. Tabert, P. Powledge, G. Borriello, and B. Schilit, ''Place Lab: Device positioning using radio beacons in the wild,'' in *Proc. Int. Conf. Pervasive Comput.* Berlin, Germany: Springer, 2005, pp. 116–133.

[23] Q. D. Vo and P. De, ''A survey of fingerprint-based outdoor localization,'' *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 491–506, 1st Quart., 2017.

[24] S. He and S. H. Gary Chan, ''INTRI: Contour-based trilateration for indoor fingerprint-based localization,'' *IEEE Trans. Mobile Comput.*, vol. 16, no. 6, pp. 1676–1690, Jun. 2017.

[25] H. Wang, Z. Wang, G. Shen, F. Li, S. Han, and F. Zhao, ''WheelLoc: Enabling continuous location service on mobile phone for outdoor scenarios,'' in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 2733–2741.

[26] M. Susi, V. Renaudin, and G. Lachapelle, ''Motion mode recognition and step detection algorithms for mobile phone users,'' *Sensors*, vol. 13, no. 2, pp. 1539–1562, Jan. 2013.

[27] L. Wang and X.-F. Meng, ''Location privacy preservation in big data era: A survey,'' *J. Softw.*, vol. 25, no. 4, pp. 693–712, Apr. 2014.

[28] L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, ''Information security in big data: Privacy and data mining,'' *Ieee Access*, vol. 2, pp. 1149–1176, 2014.

[29] S. Vanini and S. Giordano, ''Adaptive context-agnostic floor transition detection on smart mobile devices,'' in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PERCOM)*, Mar. 2013, pp. 2–7.

[30] X. Shen, Y. Chen, J. Zhang, L. Wang, G. Dai, and T. He, ''BarFi: Barometer-aided wi-fi floor localization using crowdsourcing,'' in *Proc. IEEE 12th Int. Conf. Mobile Ad Hoc Sensor Syst.*, Oct. 2015, pp. 416–424.

[31] H. Ye, W. Yang, Y. Yao, T. Gu, and Z. Huang, ''BTrack: Using barometer for energy efficient location tracking on mountain roads,'' *IEEE Access*, vol. 6, pp. 66998–67009, 2018.