IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 30, 2016, accepted January 24, 2017, date of publication March 6, 2017, date of current version April 24, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2678471

Emergency-Oriented Spatiotemporal Trajectory Pattern Recognition by Intelligent Sensor Devices

PING ZHANG, QING DENG, XIAODONG LIU, RUI YANG, AND HUI ZHANG
Department of Engineering Physics, Tsinghua University, Beijing 100084, China

Corresponding author: H. Zhang (zhhui@tsinghua.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 91646201, Grant 91546111, and Grant 91224008.

ABSTRACT This paper presents an emergency-oriented procedure to recognize trajectory patterns by analyzing GPS data collected from intelligent sensor devices. An overall description, including design architecture and system modules, is presented. The primary issues are devoted to satisfying the requirements of key group identification and surveillance under normal and emergency circumstance. For the sake of panoramic understanding of human distribution and movement, semantic trajectory information is extracted from dynamic transportation data and static human distribution data. The sequential Monte Carlo method in conjunction with a state-transition model is employed to predict the updating real-time locations. The proposed algorithm selects particles from time-stamped sequential historical data sets. Simultaneously, a resampling strategy is developed to replace low-weight particles. A curve similarity measurement called Fréchet distance is employed to compare trajectories and city roads. Afterward, human daily location and significant locations are identified based on the clustering method. To evaluate the proposed procedure and methods, sequential trajectory data sets come from the GeoLife project along with human distribution logs from smartphone application EMAPP are utilized. Finally, we demonstrate the potential of dealing location information for promoting emergency management.

INDEX TERMS Emergency management, human distribution, moving pattern recognition, spatiotemporal trajectories, trajectory data mining.

I. INTRODUCTION

With the dramatic progress of positioning equipment like the base station, smartphone, RFID, surveillance camera and GPS appliance, individuals' mobility trajectories are continuously captured tagged with time stamp. At the same time, one can move around the city as the traffic road and public transportation become more and more available, which produce a large variety of position logs inevitably. These trajectory data serve as an important foundation for understanding traffic condition and individuals' mobility behavior. Since travel logs accumulate, human periodical behavior patterns are easily recognized based on data mining technology. Alternatively, these findings can be applied to predict the traffic congestion and the places where people like to travel.

Spatiotemporal trajectory analysis has been applied to many aspects such as commuting choice, transportation management, commercial recommendation, urban planning, tourism service, criminal investigation etc. With the development of multi-sensor data fusion, it becomes much easier to integrate varieties of multi-scale spatiotemporal data. Figuring out the characteristics of human movement pattern and group behavior can be extremely useful for normal and emergency circumstances. Taking a large-scale activity for example, hundreds to thousands people will gathering into a small area at particular time. Commonly, camera networks are used to on-site staff surveillance. Meanwhile, the surrounding real-time traffic conditions are captured by on-board GPS in vehicle. By integrating these data, the flow of crowds can be forecasted. It is necessary to take countermeasures to manage and control the stream of people, such as isolating the crowd, and make sure the surrounding traffic are under control. If human movement and distribution cannot be identified with high efficiency, emergency decision-making can be delayed and the rescue resource cannot arrive on time, which cause the stampede and traffic paralysis easily.

Human trajectory and mobility patterns have a high degree of freedom and ambiguity [1]. Considering a countrywide or global scale, human movements are limited with the geographic and social constrains [2]. Especially, it is difficult to analysis spatiotemporal patterns in uncertain hazard circumstance. Meanwhile, it is urgent to unveil the human movement pattern during life-threaten situations. As related study progressed, some researchers suggest that the difficulties were not intractable as formally presumed. Lu *et al*. [3], suggest that both the travel distances and the size of people's movement trajectory are significantly more predictable than previous thought during severe disaster. Brockmann *et al*. [4], conclude that human travel on geographical scale is an ambivalent and effectively super diffusive process. In the field of emergency management, the trajectory recognition problem always contain the distribution of the victims after earthquake, evacuation route choice after hurricane and typhoon, evacuation boundary determination after explosive accident or toxic fumes leakage, human mobility patterns along with the infectious disease, and transportation conditions after terrorist attack. In order to tackle these challenging problems, multi-source intelligent sensor devices are indispensable to collect disaster related spatiotemporal data. Except the human movement patterns and the hotspots identification that have mentioned, inflow and outflow of a pacific area, periodical human transition, and daily transportation conditions are indispensable information. Location based information such as the shelter distribution and the real-time development of disaster can promote the evacuation job. When a person is under a disaster circumstance, it is urgent to push emergency information to him/her based on the development of disaster. Emergency management involves the rescue resource distribution, emergency logistic management, determine the number of victims need to be evacuated, and evacuate victims to safety emergency shelters. In this paper, we aim to unveil the emergency-oriented trajectory pattern based on trajectory data. Recognizing the human distribution change can help to understand the evolution of disaster dynamics. The contribution of this paper can be summarized as follows.

- We give a general trajectory data processing procedure designed for emergency management exclusively.
- We design a smart phone application, which capable of adapting to different disaster scenario. System architecture is depicted and specialized functions are illustrated.
- Along with spatiotemporal information, we integrate the emerging social media function into *EMAPP* to give a comprehensive understanding of disaster development.

The remainder of this article is organized as follows. First, we give an overview of the current state of human location and trajectory analysis. Then, the outline of the framework of our approach is presented. Descriptions on data preprocessing method, mobility models construction and semantic trajectory pattern recognition are illustrated. Experiment design and corresponding results are reported afterwards. Finally, we draw conclusions and offer future work.

II. LITERATURE REVIEW

According to the schemes of data acquisition, technologies for recording human distribution and trajectory can be divided into three categories. ***Direct Method***, which collects position data directly by sensor devices including mobile phone base stations, GPS, and loop detectors on the road etc. These devices are used in different situations depending on the types of data repositories [5]. However, position accuracy and the scope of application are quite different by using this method. The inaccuracy always caused by equipment error or changes in environmental conditions. The data collected from direct method require cleaning. With the development of Internet of Things, all intelligent sensor devices are able to connect with the Internet. Mining location information from the Internet with the support of hardware equipment is called ***Indirect Method***. These devices, such as smart phones, PDAs, and personal computers, always connecting with the World Wide Web. With the proliferation of social media applications, people associate with each other from the cyberspace, which results in the formation of online community. At the same time, ubiquitous location data are mined from these emerging areas. This method compensates for the difficulty of signal sampling and systematic error. For example, geographic location data attached with posting time are able to be collected from microblog API [6], [7]. Geotagged photos uploaded to Flickr can formulate spatial trajectories. Twitter users can add their location to their tweets. In order to get the accurate positioning data, some researchers use social media check-in data. In addition, place recommendations are provided by Foursquare based on detecting a user's location. Compare with Direct and Indirect method, there is a distinct method by reasoning in logical and sensible ways. ***Inference Method***, which means inferring latent location data from existing relevant methods. Such as analyzing the circulation of bank notes to give a quantitative assessment of human travelling statistics [4], multi-agency in situ sensor data integration for providing real-time spatial information [8], and analyzing interconnecting video camera data to count indoor area individuals and rebuild trajectory movement. Inference method made the data, which seem irrelevant, effectively used. Meanwhile, depth data mining from large varieties of potential sources are able to provide real-time information for emergency management and make up the deficiencies of direct method and indirect method.

Data collected from smart sensor devices have the potential to provide insightful pedestrian dynamics. Procedures of data processing upon well-recognized human mobility range from tracking, spatial data storage and expression, noise data reduction, trajectory segmentation, travel modes identification, semantic behavior understanding to group activities recognition. In order to simulate fine-granularity of human mobility, heterogeneous data require integration. Many studies have been conducted to establish the microscopic urban structure using graph theory. Crossroads are abstracted as vertices and the traffic road between the

TABLE 1. Representative field studies.

Research	Research Area	Scenario	Main Contributions	Dataset	Time duration
(Brockmann et al, 2006, [4])	United States	normal circumstance	quantitative assessment of human travelling statics	464,670 bank note trajectories	-
(González et al, 2008, [1])	-	normal circumstance	understanding individual human mobility patterns	100 thousand mobile phone users position	6 months
(Eagle et al, 2009, [22])	MIT Campus	normal circumstance	inferring friendship network structure	94 mobile phone logs	9 months
(Zheng et al, 2010, [23])	Beijing, China	outdoor movement	understanding mobility by trajectory data mining	182 users' trajectory data	over 5 years
(Cranshaw et al, 2010, [11])	Pittsburgh, USA	normal circumstance	bridging the gap between physical location and online social network	489 users' location tracking data	several months
(Stange et al, 2011, [24])	Nurburgring, Germany	big event	monitoring the movement behavior of visitors in big event	12,000 Bluetooth recording data	2 days
(Lu et al, 2012, [3])	Haiti	earthquake	analyzing population movement following earthquake disaster	1.9 million mobile phone users movement	384 days
(Song et al, 2014, [15])	Japan	earthquake	predicting human emergency behavior and mobility	1.6 million human mobility GPS records	over one year
(Blanke et al, 2014, [25])	Zurich, Switzerland	large-scale festival	capturing crowd dynamics	28,000 smartphone users' GPS data	3 days
(Phithakkitnukoon et al, 2015, [26])	Japan	tourist behavior	tourist travel behavior analysis	130,861 mobile phone users location data	one year

crossroads is regarded as an edge. Corresponding shortest path problems and maximum flow problems are studied based on the abstract graph theory. Long *et al.*, provide a review of quantitative methods for movement data from geographic information system view, including time geography, path descriptors, similarity indices, pattern and cluster methods, individual-group dynamics, spatial field methods, and spatial range methods [9]. Focusing on urban management, Steenbruggen *at al.* [10], analyze several case studies for understanding the dynamics of cities, which include smart mobility and transportation, smart economy, public safety and land-use sustainably. Lin and Hsu [5], provide a general perspective for studies, which cover the results about inferring significant locations for predicting future moves, detecting modes of transport, mining trajectory patterns, and recognizing location-based activities. Except for these conventional areas, the linkage between cyberspace and physical space becomes an emerging field. The former studies cover the basic spatial problems, but neglect the data from indirect method. Cranshaw *et al.* [11], attempt to bridge the gap between the online social network data and the location traces of its users. Crandall *et al.* [12], employ time and space co-occurrence to infer the likelihood of social ties between people. Alternatively, Cho *et al.*, show that social network ties influence long-distance travel more than short-ranged. Based on social network check-in data sets and cellphone location data sets, they found that social relationships could explain about 10%∼30% of all human movement [2]. In addition, many researchers are devoted to study the daily human distribution by considering the disaster scenarios. Apart from the normal daily circumstances, fusion and mining multi-source heterogeneous data have made the research on emergency circumstance possible. Emergency events have explicit and catastrophic characteristics, which happened with little anticipation. When encountering with the unexpected events, how the human activity pattern changed require understanding [13]. In order to tackle these difficulties, researchers are dedicate to analysis data collected from mobile. phones [3], [14], [15]. Concurrently, social media are able to support urban emergency events analysis, behavior analysis for emergency management, and disaster modeling [16]–[18]. In order to illustrate the characteristic of researches in different scenarios, some representative field studies are summarized in Table 1.

In this article, we developed a smartphone application, which integrate the abstract graph model along with social network, applied to emergency scenario. We highlight the techniques for data preprocessing and model construction. The application is applied into usage, which shows satisfying performance.

III. FRAMEWORK AND MODULES

A. SYSTEM ARCHITECTURE

Emergency management benefits from real-time acquisition of spatial information. Considering the complex situation during disaster response, heterogeneous data, such as human distribution and movement, traffic network, and geographic information require integration. A smart phone application, named *EMAPP*, has been employed to collect users' position information with the permission of its users. The application contains four modules, namely Geo-database Module, Prediction Module, Information Release Module, Scenario and Social Network Module. The GPS logs along with historical datasets are stored on a network server remotely.

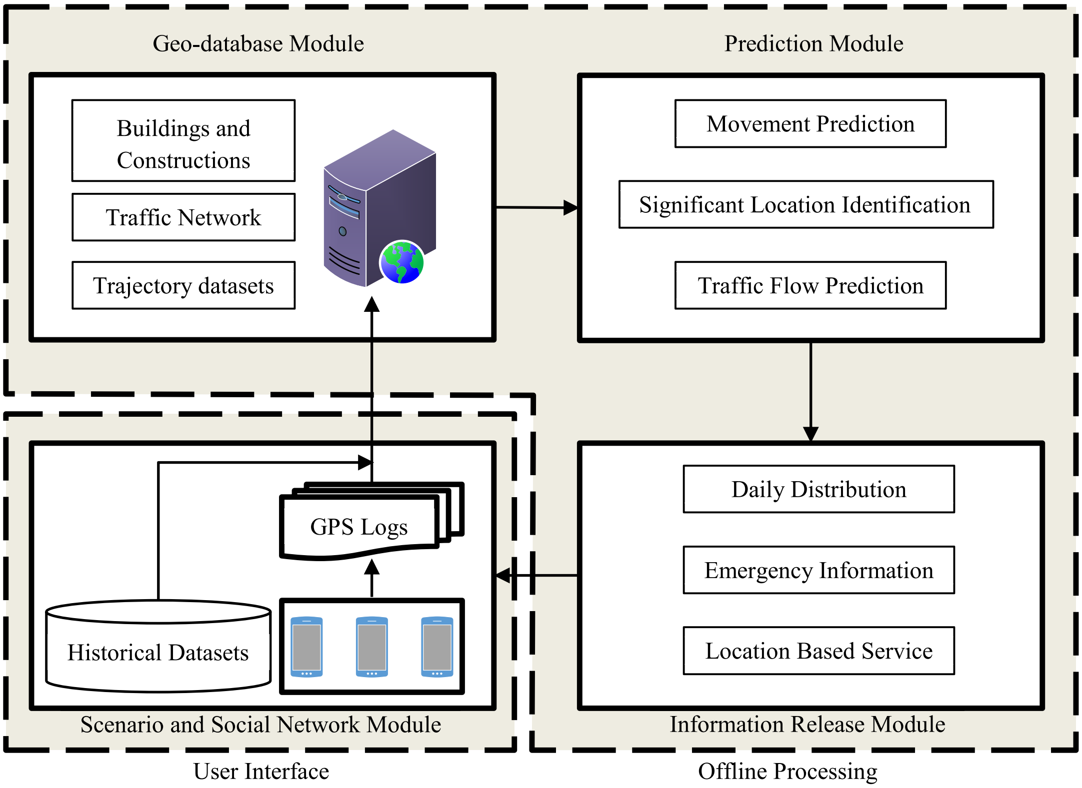

FIGURE 1. System architecture and modules.

Combined with geographic information like urban buildings, constructions and traffic network topology, the web server constitutes the Geo-databased Module. Emergency response requires a high demand of real-time information interaction. Therefore, the application is designed to contain a Prediction Module. Human movement prediction, significant location identification, and traffic flow prediction made up the main function in the Prediction Module. The application is devoted to offer real-time surveillance and population monitoring, especially during large-scale activities. Human distributions, emergency information and location-based service are provided to activity committee and the participants. During emergency scenario, users are able to establish temporal social network with each other. Meanwhile, they can upload disaster information depend on the development of emergencies. Different scenario information are stored on another web server. The system architecture and modules are depicted in Figure 1.

B. GEO-DATABASE MODULE

The application is designed to adapt to different emergency scenarios. It is necessary to capture multi-source data as the disaster type is changing. With GIS technology, large variety of location data are stored with latitude and longitude, such as traffic network and buildings. Essentially, *EMAPP* acts as an intelligent sensor. A great volume of real-time spatial data are generated by the users' daily usage. After data collection, spatiotemporal data processing involves noise reduction, spatial data storage and representation. However, storing all kinds of unrelated redundancy data produce a burden to the storage space and reduce transmission efficiency. For trajectory data, outliers are supposed to be detected and noise require removed. Meanwhile, compress trajectory data can save storage space. Yan and Chakraborty [19], describe two representative compression algorithms, the Douglas-Peucker extensions with the application of Synchronized Euclidian Distance, and STTrace. During a large-scale activity, most people heading for a specific place, monitoring and surveillance the location of these people are different from recording their daily locations. To handle the large volume of location updates in the tracking system efficiently, Liu *et al* [20], pose and employ dominant path pattern in the moving object tracking system to reduce overall location updates between moving objects and the server. To tackle the difficulty of data redundancy, three kinds of spatiotemporal data, which include user trajectory, vehicle position, and traffic condition, are integrated in this article. The matching of such critical data are able to facilitate real-time evacuation process. For advanced semantic pattern recognitions, trajectory segmentations are extracted.

C. PREDICTION MODULE

With the cleaned data from the Geo-database module, *EMAPP* can perceive emergency and make useful prediction before hazard evolves into disaster. Concurrently, prediction module contains lots of geographic information processing based on graph theory. By clustering GPS data into meaningful locations, Ashbrook and Starner [21], incorporate location data to predict movement across multiple uses. Like Ashbrook's job, we cluster GPS data to achieve critical place acquisition. Except for human movement and abnormal clustering, traffic congestion also considered in this module.

1) MOVEMENT PREDICTION

Movement prediction provides rich location and context information, which facilitate for adapting to future locations [27]. Similarly, users' movements and mobility patterns are analyzed based on their daily use of *EMAPP*. Therefore, evacuation route and the nearest shelter can be recommended to the victims during emergency circumstance.

2) SIGNIFICANT LOCATION IDENTIFICATION

By clustering trajectories, similar movement patterns can be identified. Meanwhile, heat map displayed in *EMAPP* depicts the population distribution based on the online users' number and their geographic locations. During a large-scale activity, when the number and density of participants exceed a predefined threshold, it becomes necessary to warn the users to leave the high-risk area.

3) TRAFFIC CONGESTION PREDICTION

By calculating trajectory sequence, when the moving speed of a user in accordance with a vehicle. Meanwhile, the user's location is located on the traffic lane. Therefore, it is presumed that the smart phone is inside the vehicle. Therefore, pedestrian crowded problem is changed into a traffic congestion problem. The mechanisms of pedestrian crowded and traffic congestion are similar.

D. INFORMATION RELEASE MODULE

Victims need to know how to take measures under the disaster threat. In addition, real-time situation and rescue information are required. In the prediction module, we use the cleaned data from geo-database module. The transition from data processing to the online message sending is completed in this module.

1) DAILY DISTRIBUTION INFORMATION

Human distribution is always the start and end of their daily travel. Long time movement prediction cannot be achieved without knowing the individual daily distribution. Clearly, people with different careers working and living in different places. The characteristics between their commuting and periodical patterns are diverse. When the volume of distribution data, which collected from intelligent sensor devices, becomes a city-scale, the human daily distribution can be distinguished by combining *Movement Prediction* with *Significant Location Identification*. Determing human daily distribution will foster daily transportation services, and allocation of disaster relief supplies.

2) EMERGENCY INFORMATION COLLECTION

Emergency information flow comes from many resources. *EMAPP* is designed with an incident-reporting interface. When unexpected threats happened, such as fire or terrorist attack, users are become information-sending sources. They can timely report by uploading scenario picture tagged with spatiotemporal information with brief description text. All information require filtering, analyzing, and then sending to other victims by the local emergency agency. These new paradigm of disaster perception can facilitate mutual help and foster emergency rescue which looks like crowdsourcing actually.

3) LOCATION BASED SERVICE

Gathering spatial related disaster response data like the locations of the shelters and emergency medical services are beneficial to the city dwellers. Such information will help the evacuees make correct decisions during emergency, which decrease the evacuation time under the exposure of risk. *EMAPP* extended by coupling the static information with online communication means, such as, social media and SMS text message, is able to broadcast a useful notification to warn the evacuees.

E. SCENARIO AND SOCIAL NETWORK MODULE

People behaves differently from normal circumstance when unexpected emergency happened. Various types of disasters occur dissimilar and the influences are dramatically different, which change the pattern of data analyzing. Social conformity theory indicates that people's movements are influenced by others [28]. Meanwhile, information dissemination via website and social media are quicker than official notification. Communities can share information simultaneously after disaster happened. Community neighbors are capable of contacting with each other and providing mutual support by establishing temporal online social network through *EMAPP*. Human chatting and meeting function are integrated in *EMAPP*. Users can broadcast rescue messages to their friends in case of trapped in a disaster zone. If the disaster type changes, the scenario module varies accordingly.

IV. METHODOLOGY

A. DATA PREPROCESSING

Intelligent sensors collect human trajectory and distribution data at irregular time intervals. For each period, the observation of the GPS readings are inevitably influenced by the user's actual location, movement, surrounding circumstance and equipment defects. When the user steps into a building, the GPS readings of indoor position are no longer accurate. The application, which installed on the mobile phone, will lost location data as soon as the user shut it down. In spite of these inaccuracies, the location data are expressed as three-tuples time sequences with longitude and latitude. For example, a person's raw trajectory can be denoted as:

$$Trajectory = \{(lng_1, lat_1, t_1), (lng_2, lat_2, t_2), \cdots, (lng_n, lat_n, t_n)\} \quad (1)$$

where $t_{1:n}$ denotes the time of each sampling point.

The raw trajectory contains errors, filtering and smoothing are employed. The noisy data are eliminated.

$$Trajectory = \{(lng_1, lat_1, t_1), (lng_2, lat_2, t_2), \cdots, (lng_m, lat_m, t_m)\} \quad (2)$$

where $m \leq n$. Based on different transportation modes, function (2) is segmented.

$$Trajectory(b_1) = \{(lng_1, lat_1, t_1, b_1), (lng_2, lat_2, t_2, b_1), \cdots, (lng_i, lat_i, t_i, b_1)\} \quad (3)$$

$\boldsymbol{b} := \{bike, bus, car, subway, walking\}$. $b_1 \in \boldsymbol{b}$. where $\boldsymbol{b}$ classify a person's transportation mode. Function (3) means that from time t_1 to t_i ($t_i \leq t_m$), the corresponding transportation mode is b_1.

1) PARTICLE FILTERING AND SMOOTHING

The first step after obtaining the trajectory data is noise filtering. Much trajectory data have noise especially when the user change transportation mode. GPS positioning error always caused by the hardware equipment, users' movement, and the interference of surrounding buildings. GPS signals are usually blocked by the indoor or underground circumstance. GPS devices may disrupt near tall buildings and continuously collecting GPS data may reduce device's energy quickly. Meanwhile, low battery will no doubt reduce the positioning accuracy. GPS positioning cannot be used inside tall buildings, the error in the height may cause several floors misunderstanding. In this part, a version of sequential Monte Carlo method, called particle filtering, is applied to integrate position data filtering and estimation.

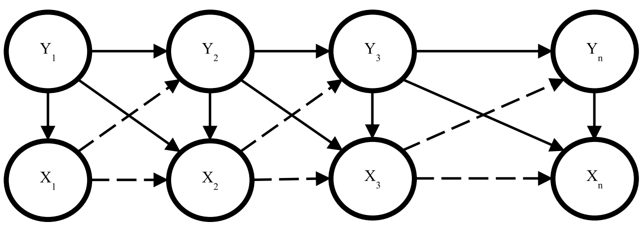

FIGURE 2. Sequential Monte Carlo method in error reduction.

Existing methods in error reduction include mean filtering, median filtering, Lancoz filtering, Wiener filtering, Kalman filtering, and Bayesian filtering etc. The filtering problem consists of not only estimating accurate values upon historical data when partial observations are made, but also the future location prediction in the dynamic system. However, the aforementioned methods excel in processing the existing data instead of making travel position prediction. To address the difficulties, the Sequential Monte Carlo (SMC) method

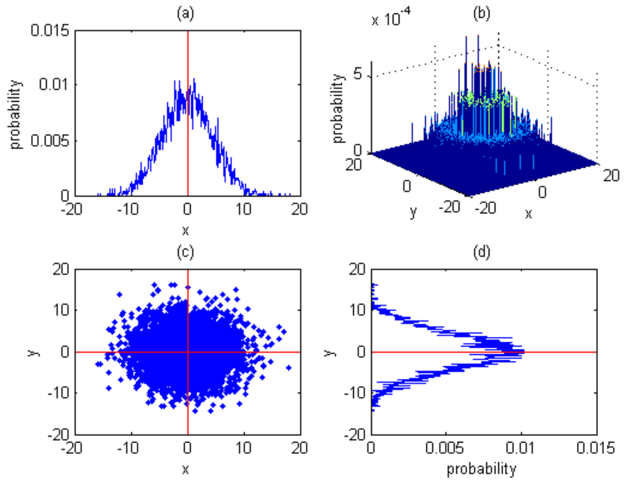

FIGURE 3. Two-dimensional Gaussian distribution. (a) probability density distribution curve above the x-axis. (b) two-dimensional probability density distribution. (c) two-dimensional particles distribution. (d) probability density distribution curve above the y-axis.

in conjunction with a state-transition model is employed to predict and update real-time location. The procedures of the SMC method are illustrated in Figure 2.

Where $Y_{0:n}$ denotes the geographic position of a GPS carrier at time $t_{0:n}$. Correspondingly, $X_{0:n}$ is the prediction position at time $t_{0:n}$. The meanings of each arrow's direction are depicted as follows.

$Y_i \rightarrow Y_{i+1}$: the movement of GPS carrier from time t_i to t_{i+1};

$X_i \rightarrow X_{i+1}$: the state transition of prediction position from time t_i to t_{i+1};

$Y_{i+1} \rightarrow X_{i+1}$: a resampling strategy developed to replace low weight particles;

$X_i \rightarrow Y_{i+1}$: X_i is used to predict the position of Y_{i+1}.

where $i \in \{0, 1, \cdots, n-1\}$.

SMC method is the procedure of sequential particles updating with resampling based on importance factors (sampling-importance-resampling, SIR).

a: SAMPLING (S)

X_i is make up of particles, which produced by a two-dimensional Gaussian distribution $X_i \sim \boldsymbol{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Where (μ_1, μ_2) is the geographic position of Y_i; $(\sigma_1^2, \sigma_2^2, \rho)$ are the covariance and correlation coefficient which can be adjusted according the real scenario. Two-dimensional Gaussian distribution of the particles are shown in Figure 3.

b: IMPORTANCE (I)

The series changing of X_i is the Markov process, which means that a particle only has connection with the former one. The transition of prediction position from t_i to t_{i+1} is denoted by (4).

$$X_{i+1} = X_i + \frac{Y_i - Y_{i-1}}{t_i - t_{i-1}} \cdot (t_{i+1} - t_i) \quad (4)$$

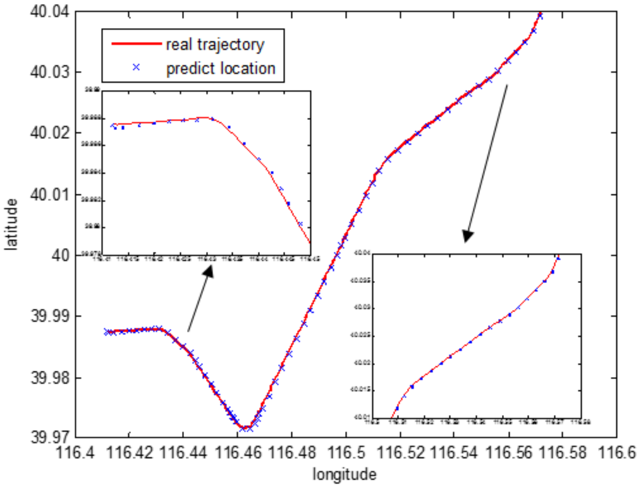

FIGURE 4. **Location estimation of a real vehicle route.**

Let $\boldsymbol{x}$ denotes the particles in position X_{i+1}, the particles are resampled once the distance between X_{i+1} and Y_{i+1} changed. The weights of new particles $\tilde{\boldsymbol{x}}$ in position X_{i+2} are denotes as function (5). $\omega\,(\tilde{\boldsymbol{x}})$ are non-negative weights, called the importance factors which sum to one.

c: RESAMPLING (R)

The weight inversely proportional to the distance between X_{i+1} (predict location) and Y_{i+1} (real location). Through normalization, the low weight particles will be replaced gradually.

$$\omega\,(\tilde{\boldsymbol{x}}) \propto \frac{1}{\sqrt{2\pi}\sigma}\exp(-d\,(X_{i+1}, Y_{i+1})^2/(2\sigma^2)) \tag{5}$$

A representative location estimation of a real vehicle route is shown in Figure 4. The red line depicts the real trajectory and the blue cross demonstrates the predict position. Driving direction is from east to west. The total driving distance is 17 kilometers and time duration is 965 seconds. Five thousand particles, which obey two-dimensional Gaussian distribution, are produced surrounding the existing trajectory points. Based on several former points, consecutive points' locations are updated. In addition, particles are resampled by calculating the weight between the predict locations and the real trajectories. With the prediction process going, only the particles with high weight are reserved.

Figure 5 presents a visualization of the vehicle route filtering and corresponding prediction error. The red line depicts the real trajectory; the green line depicts the trajectory prediction error. The position prediction are divided into four phases. In phase I, the sampling is just begin, reasonable errors cannot avoid. With the sampling going, it becomes more stable to predict the trajectory. Therefore, the errors are going down sharply in phase II. However, the traffic roads are not always straight. When the vehicle change its direction, the errors are accumulated (phase III). In phase IV, with the road become straightness again, the position prediction errors are keep balance.

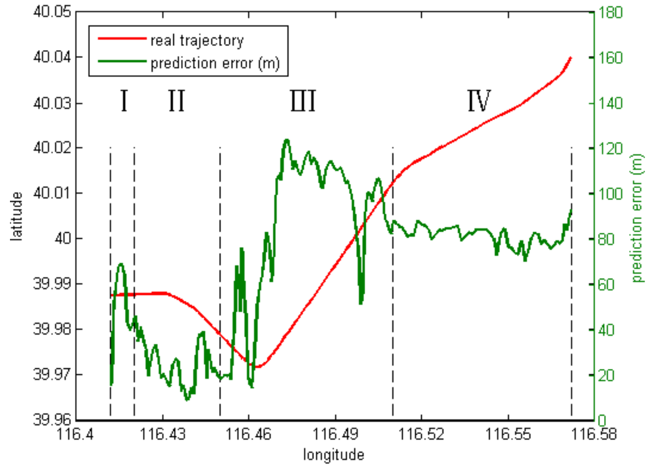

FIGURE 5. **A real vehicle route filtering and corresponding prediction error.**

2) TRAJECTORY MAP-MATCHING

It is quite difficult to satisfy the requirements for exact particle filtering and location prediction unless accurate trajectory and road network data are given. The general purpose of this part is to determine on which road that the GPS carrier is travelling based on his/her historical trajectory data. In order to map matching the correct road, the similarity between the GPS trajectories and traffic road segments are compared. GPS probes are inevitable have positioning errors during actual sampling process. Therefore, it is presumed that an orientation precision ε in advance. If the distance between a road and trajectories exceeds the threshold, it is regarded that the historical trajectory dataset cannot be matched to the road. Considering the actual situation, the Fréchet distance is used as a quantitative measurement to improve the accuracy of the similarity judgement.

Given two curves, $\boldsymbol{f} : \left[a, a'\right] \rightarrow \boldsymbol{R}, \boldsymbol{g} : \left[b, b'\right] \rightarrow \boldsymbol{R}$, where $\boldsymbol{R}$ stands for a Euclidean vector space, then the Fréchet distance $d_F\,(\boldsymbol{f}, \boldsymbol{g})$ between two curves is defined by function (6).

$$d_F\,(\boldsymbol{f}, \boldsymbol{g}) := \inf_{\substack{\alpha[0,1]\rightarrow[a,a'] \\ \beta[0,1]\rightarrow[b,b']}} \max_{t\in[0,1]} \|\boldsymbol{f}\,(\alpha\,(t)) - \boldsymbol{g}(\beta(t))\| \tag{6}$$

Here, $d_F(\boldsymbol{f}, \boldsymbol{g})$ is the discrete Fréchet distance between the two polygonal curves [29], where $\|\cdot\|$ denotes the L_2 norm [30]. Trajectory $\boldsymbol{f}$ consists of sequential points tagged with time stamp. Traffic road is made up of several line segments $\boldsymbol{g}$. The similarity between two curves is defined by function (7) [31].

$$SI(\boldsymbol{f}, \boldsymbol{g}) := 1 - \frac{\inf\limits_{\substack{\alpha[0,1]\rightarrow[a,a'] \\ \beta[0,1]\rightarrow[b,b']}} \max\limits_{t\in[0,1]} \|\boldsymbol{f}\,(\alpha\,(t)) - \boldsymbol{g}\,(\beta\,(t))\|}{\sup\limits_{\substack{\alpha[0,1]\rightarrow[a,a'] \\ \beta[0,1]\rightarrow[b,b']}} \max\limits_{t\in[0,1]} \|\boldsymbol{f}\,(\alpha\,(t)) - \boldsymbol{g}\,(\beta\,(t))\|} \tag{7}$$

Where $0 \leq SI(\boldsymbol{f}, \boldsymbol{g}) \leq 1$. The value of $SI(\boldsymbol{f}, \boldsymbol{g})$ is proportional to the similarity between f and g. The contin-

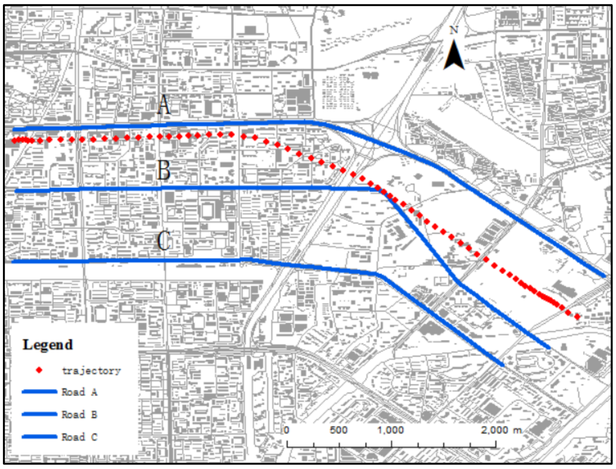

FIGURE 6. Map matching of trajectory and traffic road.

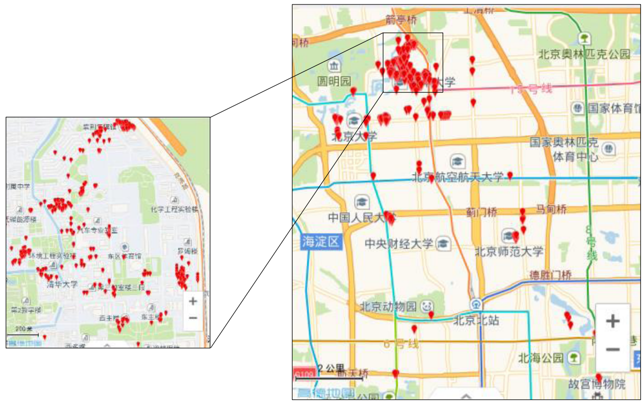

FIGURE 7. Case study area in Beijing, China.

uous line-to-line comparison function (7) is discretized by considering the location and ordering of the trajectory points along the curves. In a real GIS map, traffic road is made up of several polylines instead of a continuous curve. In order to get the road network data, random sampling method is applied to these polylines. Start point, intersections, and end point are chosen as the criticle points. Then, random sampling is applied to the polylines based on these criticle points. Essentially, the map-matching problem is to find the largest SI among the trajectory under the restriction of $d_F(\boldsymbol{f}, \boldsymbol{g}) \leq \varepsilon$. In Figure 6, let T depicts the real trajectory (a chain of red points). Road A, Road B, Road C are three different roads. The map demonstrates the northeast part of downtown Beijing and the trajectory is part of the real trajectoy in Figure 5. By calculating the similarity value SI, we get

$$SI\,(T, C) < SI\,(T, B) < SI\,(T, A)\,.$$

Therefore, Road A makes the best map-matcing result.

B. CONSTRUCTING MOBILITY MODEL

In data preprocessing procedure, the main work is map matching the trajectory data onto the traffic road accurately. Immediately, mobility trajectory in conjunction with critical location information require extracted and reconstructed. We aim to obtain crowd dynamics by recording the location information from city residents. This can be down only through abundant and detailed samples of the crowd.

1) SIGNIFICANT LOCATIONS

Significant locations, where individuals stay for a long time or a crowd of people visit temporary, are easily cause stampede. These places are always vulnerable which require high concerns. Mapping the trajectory data to these places can be accomplished by inferring from GPS traces [5]. However, it is difficult to get all the position data at one time. Therefore, comprehensive understanding of human moving regularity should be made from limited data sets. For example, when a person's location data do not change anymore, we can presume with certain probability that he/she has arrived office or home. When walking into a tall building temporary, the location data tracked by GPS device will stop changing. As the person came out, the data update again. The time difference between missing data can be treated as the duration of staying. Generally, the exact number of crowded area cannot be known in advance. Therefore, the DBSCAN (density-based spatial clustering of applications with noise) method is applied to find regular and significant locations [32]. This method is employed in the Case Study section and therefore human distribution clustering from *EMAPP* are extracted.

2) TRANSPORTATION MODES

The characteristic of individuals' mobility is particularly different for what it contains a variety of different travel modes [33]. During daily routines, people will transfer commuting tools from residential area to business district. Change of transportation tools can be described as a *transfer matrix* mathematically. Different operations on *transfer matrix* can be employed to express the different transformation of transportation modes. When unexpected disaster happened, it is critical to determine the number of victims and emergency vehicles. In addition, deployment of post-disaster relief supplies cannot achieved without urban transportation system. To build a complete mobility map, we combine the sequential dataset from GeoLife project [23], [34], [35] and *EMAPP* smart phone application in this article. The human movement data tagged with the transportation modes are employed.

V. CASE STUDY

A. EXPERIMENT ENVIRONMENT SETUP

Through the development of smart phone application, we put forward a more complete spatiotemporal data acquisition method based on the system framework and its four modules. In this study, we recruit 40 students to conduct the test experiment [36]. These students are mainly come from the Department of Engineering Physics in Tsinghua University, China. They were asked to install *EMAPP* on their smartphones. The experiment was began in June 2016 and lasted for one month. By using the scenario and social network module, each student added at least one student as chatting friend who was using *EMAPP* at same time. Under the permission of these students, their daily trajectories are collected only if *EMAPP* is running. By analyzing the position of their locations, these students appeared in five different provinces nationwide, of which location data in Tsinghua campus up to the rest of other places (see Figure 7). It is because the experiment period was in spring semester and most students were staying in the campus. Therefore, we mainly analyze the students' spatial and temporal trajectories in the campus area, the trajectory data outside are not considered.

B. EXPERIMENT RESULTS AND ANALYSIS

It is difficult to obtain the movement of the whole staff of the university. Therefore, the trajectory of the 40 students are representative samples to reflect the overall situation of the campus. For the reason of lacking abundant trajectory data, density-based spatial clustering method (DBSCAN) is employed [32]. By analyzing the characteristic of the real trajectory data, the parameter of Eps-neighborhood is set 50m, and the minimum number surrounding the core point is set 20. By executing the DBSCAN program, four places that the experiment subjects clustered are discovered (see Figure 8). The geographical coordinate of each cluster center, the number of points in each cluster, and the location of the four places are described in Table 2.

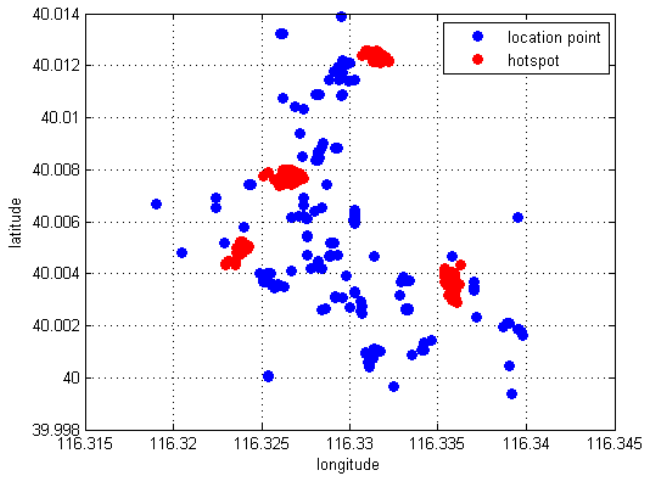

FIGURE 8. Significant location identification based on daily distribution.

TABLE 2. Brief description of the four clusters.

No.	Cluster center	Number of points	Location
1	[116.3358,40.0036]	53	Laboratory
2	[116.3315,40.0123]	73	Dormitory X
3	[116.3238,40.0048]	34	Library
4	[116.3265,40.0077]	47	Dormitory Y

From spatial perspective, the four clusters are all located indoors. It is reasonable that the experiment was conducted in the spring semester and few large-scale activities in the outdoors of the campus. Judging from the habit of the students, they inclined to use the smart phone indoors rather that outdoor area. From the temporal dimension, the users' trajectory patterns present a certain cyclical changes. For students living in dormitory X or dormitory Y, the location information essentially keep unchanged after 23:00 and began to update in the next morning. In addition, the location recordings show that library and laboratory appeared with much greater probability in the office time. Other places, such as teaching area, show lower usage proportion because of the closure of *EMAPP*. In order to demonstrate the historical distribution of the clusters, we made a Gaussian distribution fitting for these four area and visualized the results in Figure 9.

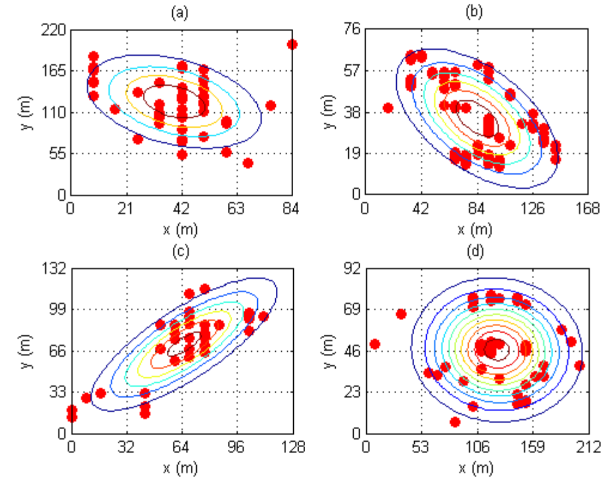

FIGURE 9. Historical participants' distribution in the four identified locations (a) Historical participants' distribution in the laboratory. (b) Historical participants' distribution in the dormitory X. (c) Historical participants' distribution in the library. (d) Historical participants' distribution in the dormitory Y.

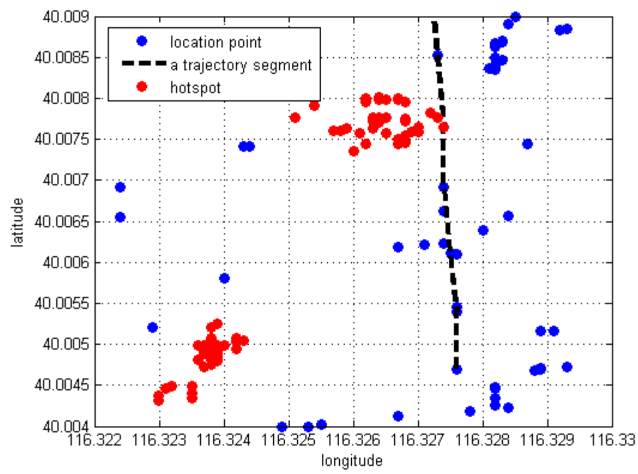

FIGURE 10. A student's trajectory segment in the campus.

Through the experiment, a student's trajectory segment is selected for analyzing in Figure 10. In addition, several regional hot spots are close to the segment. By calculating the moving speed, the user was riding a bicycle between two different places in the morning and afternoon separately. Meanwhile, the trajectory segment is map matching to the Xuetang road in the campus. From the perspective of emergency management, once an unexpected event happened near the gathering crowd, it is easily lead to stampede and personal injury. Therefore, these critical places should be given high concerns.

In addition, many students uploaded photos of the location where they think unsafety through the event report interface. The reserved function is devised that benefit from users participation to improve emergency response capability. This function is not often used in the normal circumstance. However, as a reserved function, it is indispensable in the disaster scenarios.

VI. CONCLUSIONS

In this article, we aim to integrate heterogeneous data to cope with the complex situation during emergency management. Different technologies for recording human distribution and trajectories are compared. To improve the capability of real-time information acquistion, a smartphone application *EMAPP* is developed which acts well as a senor device. We streesed the necessity of the smartphone application and demonstrated its design architecture as well as the four system modules. Especially, feasibility of *EMAPP* used in emergency management is illustrated.

We utilize a Sequential Monte Carlo (SMC) method to reduce the spatial data noise. With the Fréchet distance, trajectory data are managed to map matching to the traffic network appropriately. To build a full mobility map during disasters, human mobility concept is given, which combines the location identification with transportation *transfer matrix*. Finally, we highlight the potential of planning and analyzing large-scale activities by extracting and visualizing location information. With the deployment of *EMAPP*, a complete spatiotemporal data acquisition, data processing, and trajectory pattern recognition are put forward. By doing the experiment, we validate that tracking data analysis are beneficial to understanding human daily distribution and periodical behavior.

There are several challenges need to be tackled. For example, along with the dramatic information increase from the moving objects, high-efficiency of large volume data processing and database management are required. Emergency response and crisis management require timely and accurate information. However, the location-aware data provided by smartphones offer low spatial and temporal resolution. Therefore, fusion for multi-source data, which come from video camera surveillance, social relation network, and geographical information system, can be more convincing.

REFERENCES

[1] M. C. Gonzalez, C. A. Hidalgo, and A. L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, Jun. 2008.

[2] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social network," in *Proc. KDD*, San Diego, CA, USA, 2011, pp. 1082–1090.

[3] X. Lu, L. Bengtsson, and P. Holme, "Predictability of population displacement after the 2010 Haiti earthquake," *Proc. Nat. Acad. Sci USA*, vol. 109, pp. 11576–11581, Jul. 2012.

[4] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," *Nature*, vol. 439, pp. 462–465, Jan. 2006.

[5] M. Lin and W.-J. Hsu, "Mining GPS data for mobility patterns: A survey," *Pervas. Mobile Comput.*, vol. 12, pp. 1–16, Jun. 2014.

[6] Z. Xu, H. Zhang, V. Sugumaran, K. R. Choo, L. Mei, and Y. Zhu, "Participatory sensing-based semantic and spatial analysis of urban emergency events using mobile social media," *EURASIP J. Wireless Commun. Netw.*, vol. 44, pp. 1–9, Dec. 2016.

[7] S. Zhang, Y. Wang, S. Zhang, and G. Zhu, "Building associated semantic representation model for the ultra-short microblog text jumping in big data," *Cluster Comput.*, vol. 19, no. 3, pp. 1399–1410, 2016.

[8] F. Alamdar, M. Kalantari, and A. Rajabifard, "Towards multi-agency sensor information integration for disaster management," *Comput., Environ. Urban Syst.*, vol. 56, pp. 68–85, Mar. 2016.

[9] J. A. Long and T. A. Nelson, "A review of quantitative methods for movement data," *Int. J. Geograph. Inf. Sci.*, vol. 27, no. 2, pp. 292–318, 2013.

[10] J. Steenbruggen, E. Tranos, and P. Nijkamp, "Data from mobile phone operators: A tool for smarter cities?" *Telecommun. Policy*, vol. 39, nos. 3–4, pp. 335–346, May 2015.

[11] J. Cranshaw, E. Toch, J. Hong, A. Kittur, and N. Sadeh, "Bridging the gap between physical location and online social networks," in *Proc. 12th ACM Int. Conf. Ubiquitous Comput.*, Copenhagen, Denmark, 2010, pp. 119–128.

[12] D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg, "Inferring social ties from geographic coincidences," *Proc. Nat. Acad. Sci. USA*, vol. 107, pp. 22436–22441, Dec. 2010.

[13] J. P. Bagrow, D. Wang, and A. L. Barabási, "Collective response of human populations to large-scale emergencies," *PLoS ONE*, vol. 6, no. 3, p. e17680, 2011.

[14] T. Yabe, K. Tsubouchi, A. Sudo, and Y. Sekimoto, "Estimating evacuation hotspots using GPS data: What happened after the large earthquake in Kumamoto, Japan?" in *Proc. 5th Int. Workshop Urban Comput. (URBCOMP)*, 2016.

[15] X. Song, Q. Zhang, Y. Sekimoto, T. Horanont, S. Ueyama, and R. Shibasaki, "Modeling and probabilistic reasoning of population evacuation during large-scale disaster," in *Proc. KDD*, Chicago, IL, USA, 2013, pp. 1231–1239.

[16] Z. Xu *et al.*, "Building knowledge base of urban emergency events based on crowdsourcing of social media," *Concurrency Comput.-Practice Exper.*, vol. 28, no. 15, pp. 4038–4052, 2016.

[17] J. Fry and J. M. Binner, "Elementary modelling and behavioural analysis for emergency evacuations using social media," *Eur. J. Oper. Res.*, vol. 249, pp. 1014–1023, Mar. 2016.

[18] Q. Deng, Y. Liu, H. Zhang, X. Deng, and Y. Ma, "A new crowdsourcing model to assess disaster using microblog data in typhoon Haiyan," *Natural Hazards*, vol. 84, no. 2, pp. 1241–1256, 2016.

[19] Z. Yan and D. Chakraborty, *Semantics in Mobile Sensing*. San Mateo, CA, USA: Morgan & Claypool, 2014, pp. 28–29.

[20] K. Liu, Z. Ding, M. Li, K. Deng, and X. Zhou, "Pattern-based moving object tracking," in *Proc. TDMA*, Beijing, China, 2011, pp. 5–14.

[21] D. Ashbrook and T. Starner, "Using GPS to learn significant locations and predict movement across multiple users," *Pers. Ubiquitous Comput.*, vol. 7, no. 5, pp. 275–286, 2003.

[22] N. Eagle, A. S. Pentland, and D. Lazer, "Inferring friendship network structure by using mobile phone data," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 36, pp. 15274–15278, 2009.

[23] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W. Ma, "Understanding mobility based on GPS data," in *Proc. 10th Int. Conf. Ubiquitous Comput.*, Seoul, South Korea, 2008, pp. 312–321.

[24] H. Stange, T. Liebig, D. Hecker, G. Andrienko, and N. Andrienko, "Analytical workflow of monitoring human mobility in big event settings using Bluetooth," in *Proc. ISA*, Chicago, IL, USA, 2011, pp. 51–58.

[25] U. Blanke, G. Tröster, T. Franke, and P. Lukowicz, "Capturing crowd dynamics at large scale events using participatory GPS-Localization," in *Proc. IEEE 9th Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process. (ISSNIP)*, Singapore, Apr. 2014, pp. 1–7.

[26] S. Phithakkitnukoon, T. Horanont, A. Witayangkurn, R. Siri, Y. Sekimoto, and R. Shibasaki, "Understanding tourist behavior using large-scale mobile sensing approach: A case study of mobile phone users in Japan," *Pervas. Mobile Comput.*, vol. 18, pp. 18–39, Apr. 2015.

[27] X. Liu and H. A. Karimi, "Location awareness through trajectory prediction," *Comput. Environ. Urban Syst.*, vol. 30, pp. 741–756, Nov. 2006.
[28] Y. Wang, N. J. Yuan, D. Lian, and L. Xu, "Regularity and conformity: Location prediction using heterogeneous mobility data," in *Proc. KDD*, Sydney, NSW, Australia, 2015, pp. 1275–1284.
[29] T. Wylie and B. Zhu, "Following a curve with the discrete Fréchet distance," *Theor. Comput. Sci.*, vol. 556, pp. 34–44, Oct. 2014.
[30] H. Alt, C. Knauer, and C. Wenk, "Matching polygonal curves with respect to the Fréchet distance," in *TACS*, A. Ferreira and H. Reichel, Eds. Berlin, Germany: Springer, 2001, pp. 63–74.
[31] J.-F. Wang, C.-D. Xu, S. Tong, H.-Y. Chen, and W.-Z. Yang, "Spatial dynamic patterns of hand-foot-mouth disease in the people's Republic of China," *Geospatial Health*, vol. 7, no. 2, pp. 381–390, 2013.
[32] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. KDD*, Portland, Oregon, USA, 1996, pp. 226–231.
[33] K. Siła-Nowicka, J. Vandrol, T. Oshan, J. A. Long, U. Demšar, and A. S. Fotheringham, "Analysis of human mobility patterns from GPS trajectories and contextual information," *Int. J. Geograph. Inf. Sci.*, vol. 30, no. 5, pp. 881–906, 2016.
[34] Y. Zheng, L. Zhang, X. Xie, and W. Ma, "Mining interesting locations and travel sequences from GPS trajectories," in *Proc. 18th Int. Conf. World Wide Web*, Madrid, Spain, 2009, pp. 791–800.
[35] Y. Zheng, X. Xie, and W.-Y. Ma, "GeoLife: A collaborative social networking service among user, location and trajectory," *Bull. Tech. Committee Data Eng.*, vol. 33, no. 2, pp. 32–39, 2010.
[36] P. Zhang, Y. Liu, R. Yang, H. Zhang, and Z. Gong, "Improving urban traffic evacuation capability in emergency response by using smart phones," in *Proc. APWEB*, Suzhou, China, 2016, pp. 241–252.

PING ZHANG received the bachelor's degree in safety engineering from the Anhui University of Science and Technology, Huainan, China, in 2014, with the excellent undergraduate of Anhui Province. He is currently pursuing the Ph.D. degree with the Institute of Public Safety Research, Department of Engineering Physics, Tsinghua University, Beijing. His research interests include emergency management, object detection and tracking, and spatiotemporal data mining.

QING DENG was born in Xiaogan, China, in 1990. She received the B.S. degree in safety engineering from Central South University, Hunan, in 2013. She is currently pursuing the Ph.D. degree with Tsinghua University, Beijing. Her research interests include public safety, emergency management, and emergency decision-making.

XIAODONG LIU was born in China in 1988. He received the B.S. degree in safety engineering from the Shandong University of Science and Technology in 2011 and the Ph.D. degree in safety science and engineering from the University of Science and Technology, China, in 2016.

He is currently a Research Assistant with the Department of Engineering Physics, Tsinghua University. His research interest includes the emergency management, pedestrian evacuation dynamics, and large-scale crowd evacuation models.

RUI YANG received the B.S. and Ph.D. degrees from the University of Science and Technology of China, China, in 1999 and 2004, respectively. He holds a post-doctoral position with the Department of Engineering Physics, Tsinghua University, China, from 2004 to 2007. He is currently a Research Associate with the Institute of Public Safety Research, Tsinghua University. His research and teaching interests include disaster modeling and simulation, numerical simulation of fire and smoke propagation, prediction of CBNR dispersion, and reverse problem of CBNR dispersion. He has authored or co-authored over 40 technical papers in renowned international journals and conferences. He serves as a Reviewer of the textitJournal of hazardous and materials, the *ASME Journal of Thermal Science and Engineering Applications*, the *Chinese Science Bulletin*, and the *International Journal of Heat and Mass Transfer*.

HUI ZHANG is a Changjiang Scholar Professor and the Vice Director of the Institute of Public Safety Research, Tsinghua University, Beijing, China. He is also the Vice Director of the Tsinghua-Boeing Joint Research Center. He is a member of ISO/TC223 (societal security) and is currently in charge of establishing ISO standard on emergency capability assessment. He received the National Science Foundation CAREER Award (USA) in 1999 when he was at Stony Brook University, USA. He has authored eight book chapters, 120 journal articles, and 140 refereed conference papers.