Walkway Discovery from Large Scale Crowdsensing

Chu Cao[1], Zhidan Liu[2], Mo Li[1], Wenqiang Wang[3], Zheng Qin[3]
[1]Nanyang Technological University, Singapore
[2]Shenzhen University, Shenzhen, China
[3]Institute of High Performance Computing, Singapore
{ccao001,limo}@ntu.edu.sg,liuzhidan@szu.edu.cn,{wangweq,qinz}@ihpc.a-star.edu.sg

ABSTRACT

Most digital maps are designed for vehicles and miss a great number of walkways that can facilitate people's daily mobility as pedestrians. Despite of such a fact, most existing map updating approaches only focus on the motorways. To fill the gap, this paper presents *VitalAlley*, a walkway discovery and verification approach with mobility data from large scale crowdsensing. *VitalAlley* aims to identify the uncharted walkways from the big but noisy personal mobility data and incorporate these findings into existing incomplete road maps. The implementation of *VitalAlley* faces the major challenges due to the unstructured nature of the walkways themselves and the noise from crowdsensing data. *VitalAlley* leverages different aspects of individual mobility to model and estimate the walkable areas, based on which representative walkways that connect known road segments or points of interest are extracted. To verify the new-found walkways, we further propose image based auto-verification with the help of publicly accessible street image database from Google Street View. *VitalAlley* is implemented and evaluated with real world crowdsensing data from the Singapore National Science Experiment. As a result, 736 walkways (totaling 161 km in distance) are identified from the mobility dataset collected from 108,337 students in Singapore. We manually verify 224 walkways totaling 32.4 km over a 9 km^2 district through on-site inspection. The results suggest over 96% accuracy of *VitalAlley* in discovering the walkways.

CCS CONCEPTS

• **Theory of computation** → **Data modeling**; **Pattern matching**; • **Information systems** → *Data cleaning*; • **Computing methodologies** → Cluster analysis;

KEYWORDS

Walkway discovery, walkable area, road maps, mobility trajectory, crowdsensing, auto-verification

ACM Reference Format:
Chu Cao[1], Zhidan Liu[2], Mo Li[1], Wenqiang Wang[3], Zheng Qin[3]. 2018. Walkway Discovery from Large Scale Crowdsensing. In *Proceedings of The 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN 2018).* ACM, Porto, Portugal, Article 4, 12 pages. https://doi.org/XXX/XXX

1 INTRODUCTION

Digital road maps are of significant importance for route planning and navigation in our daily life. Existing road maps, however, are mostly vehicle oriented and do not contain the information of many walkways that local pedestrians usually travel with. Many areas or walkways (*e.g.,* basement of buildings, interior of shopping malls, open fields, *etc.*) are used as shortcuts by people who are familiar with the local area but are not included in the digital road maps. These walkable areas or paths, although very useful, are uncharted on the maps and thus cannot be made of use by the public.

In this work, we practice the idea of crowdsensing [15, 34] and make use of the mobility data from a large number of local students to discover the uncharted walkways. Our study rides on an ongoing crowdsensing project - Singapore National Science Experiment (NSE) [18] - a city wide initiative that enrolls more than 250,000 local students carrying smart devices that sense their surrounding environment and track their mobility [29] on daily basis. The smart devices periodically log and upload the locations, IMU readings, and other environmental parameters from participating students. Massive mobility data are collected from the students across the entire city, which provides us the opportunity to study and discover the walkways. The rationale of this study is that the local students are active users of the walkable areas or paths in their neighborhood, and by following their footprints we will be able to discover and summarize those walkways.

There have been existing efforts [11, 22, 27, 33] made to completing the digital road maps. Most of such works, however, only mainly focus on discovering vehicle-oriented motorways rather than the walkways for pedestrians. The motorways are structured where vehicles strictly follow the lanes and directions with almost 1-dimensional uncertainty in mobility. On the other hand, the walkable areas or paths are mostly unstructured where people travel with high freedom of 2-dimensional uncertainty (*e.g.,* lawns, shopping malls, open fields, *etc.*). Most previous works for motorway discovery leverage the structured property of motorways to infer missing road segments with the assumption that vehicle trajectories are constrained on the 1-dimensional roads. Techniques like trace clustering [27, 33] and location point clustering [11, 22] are directly applied with the GPS trajectories of vehicles. Due to the unstructured and pedestrian-oriented nature of walkways, neither previous map updating methods nor vehicle trajectory data can be applied to discover the walkways.

In this paper, we present *VitalAlley* that uses the crowdsensing mobility data from NSE for walkway discovery. Unlike previous works for map completion, *VitalAlley* faces special challenges arising from the unstructured nature of walkways as well as the imperfect quality of the crowdsensing data. The location reports from

0-7695-6377-5/18/$31.00 ©2018 IEEE
DOI 10.1109/IPSN.2018.00009

IEEE computer society

participating students may scatter over a wide area of freedom, making it difficult to estimate walkways through the area. The imperfect quality of the crowdsensing data makes the situation more complicated - the smart device takes the locations every 15 seconds (or even longer in certain circumstances) with errors that range from tens to hundreds of meters (with WiFi hotspot based localization).

VitalAlley statistically looks at the big mobility data from many aspects, being able to tolerate errors and noise contained in the mobility data. With both the location reports and step counts derived from IMU readings, *VitalAlley* builds an ellipse model to estimate the probability of how the individual walks between consecutive reported locations. Putting together such micro estimations from all students allows *VitalAlley* to statistically understand how likely different areas are walkable. With such knowledges, *VitalAlley* applies a two-phase clustering method to discover how the potential walkable areas are connected with nearby known road segments or points of interest and then identify walkways that are representative for people who walk through such areas.

VitalAlley further employs an auto-verification method to verify the correctness of the new-found walkways. By invoking the online Google Street View (GSV) APIs [7], *VitalAlley* is able to access an extensive image library containing street images from most road segments. *VitalAlley* retrieves the GSV images from where the discovered walkways join existing roads or points of interest and analyzes the key image descriptors with reference to a library of templates to verify whether the identified walkways are true or not.

To the best of our knowledge, this is the first study for digital map completion with a focus on walkways, which includes the ellipse based walkable area estimation and weighting, representative walkway identification, and GSV image based auto-verification. The systematic study with city scale mobility data from crowdsensing is of the largest scale. We extensively evaluate the performance of *VitalAlley* with the mobility data collected from 108,337 students in Singapore, which lasts 11 weeks. Based on the analysis of more than 400 million mobility data records, we discover 736 walkways (totaling 161 km in distance) and verify 224 walkways totaling 32.4 km through on-site inspection. Those verified walkways can be integrated into the digital map of Singapore from OpenStreetMaps [20]. The results show that *VitalAlley* has a 96% accuracy in discovering the walkways.

The rest of this paper is organized as follows. We present the preliminary and motivation in Section 2. Section 3 elaborates the design details of *VitalAlley* system, and Section 4 presents the evaluations. We review the related works in Section 5. Finally, Section 6 concludes this paper.

2 PRELIMINARY AND MOTIVATION

2.1 Objective

A road map can be represented as a directional graph $G(V, E)$, where E refers to the set of edges that correspond to roads for vehicles and/or pedestrians and V refers to the set of vertices that correspond to intersection points or terminal points of road segments. We define a **road segment** as follows.

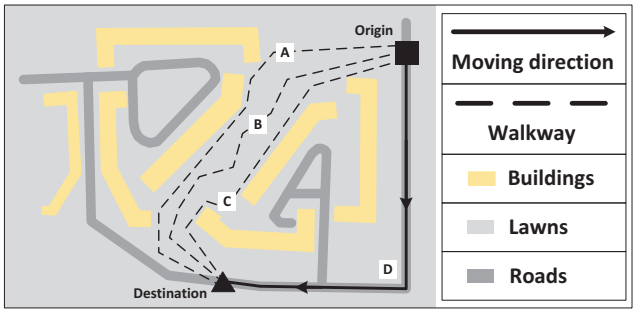


Figure 1: The comparisons of motorway and walkway.

Definition 1. ***Road segment.*** *A road segment is a directed edge in graph G, that is associated with a deterministic traveling direction and two terminal points.*

Existing digital maps do not contain all the road segments and are thus incomplete. There have been some recent efforts, *e.g., CrowdAtlas* [27], *COBWEB* [22], and *GLUE* [33], made to complete such maps. By comparing the GPS trajectories of vehicles with an existing road map, those works mainly focus on discovering those vehicle-oriented road segments. They mainly perform clustering on the vehicle GPS data not matched to existing road segments based on trajectory distance [27], moving direction [11, 33] or distance between location samples [22]. New-found road segments are constructed either by extracting the centric line of clustered trajectories [27] or connecting the centric points of clustered location samples [11, 22, 33]. Those methods perform well with the motorways which are well structured and with vehicles that strictly follow the directions of the road segments. Different from vehicle-oriented road segments, the walkways for pedestrians are often unstructured and different people may travel with very different trajectories even across the same walkable area. We define a **walkable area** as follows.

Definition 2. ***Walkable area.*** *A walkable area is an area bounded by nearby road segments or points of interest (POIs, e.g., residential buildings, schools, etc.). Unconstrained movements of people are allowed within the area.*

Figure 1 presents an illustrative example to compare the motorways and walkways. Given the origin of a trip (shown as the symbol ■) and the destination (shown as the symbol ▲), a user can either take the route on road segments (shown as the black solid line, *i.e.,* route *D*) or choose the walkways (shown as the black dash line, *e.g.,* route *A*) for this travel. For route *D* on road segments, she needs to strictly follow the directions of road segments. By contrast, she owns high freedoms on the walkways. For example, she can take route *A*, *B* or *C* for this travel, and even go with any walking path between the two buildings. Putting all these walking paths together gradually forms a walkable area, where people can freely walk at any direction. Such walkable areas widely exist, *e.g.,* lawns, basement of buildings, floors of shopping malls, open fields, *etc.*, and they may provide significant convenience and save the walking time when compared to using existing road segments on the map. Due to the high travel freedom within such walkable areas, however, pedestrians' movements could be irregular with high

uncertainty. As a result, all existing map completion approaches based on direct clustering of locations or trajectories will fail in obtaining the sketch of a walkable area.

For practical utility, discovering and identifying the entire walkable area is not necessary and thus is not the final goal of this paper. We observe that although different pedestrians may choose different paths when walking through the same walkable area, there are typical walkways that connect certain existing road segments or POIs and are frequently used. We thus define the **representative walkway** as follows.

Definition 3. ***Representative walkway.*** *A representative walkway represents the connectivity a walkable area serves between two known road segments/POIs. If we specify the intersection points between the road segments/POIs and the walkable area, the representative walkway can be denoted as a polyline connecting the two intersection points and integrated into the road graph* G *as an edge. There may be multiple representative walkways connecting different road segments/POIs adjacent to the same walkable area.*

Therefore, instead of precisely discovering the entire walkable area which we may not have sufficient mobility data to support, in this paper we aim at extracting representative walkways that sketch the accessibility of walkable areas in facilitating pedestrians' travel needs. The representative walkways, once identified, are compatible with the current digital road map and can be easily integrated into the map.

2.2 NSE Mobility Data

We make use of a crowdsensing mobility data from the National Science Experiment (NSE) [18] of Singapore to support our goal of walkway discovery. NSE is a nationwide project initiated by the National Research Foundation and supported by the Ministry of Education of Singapore. This project involves more than 250,000 students from primary, secondary, high school and junior colleges. Each student carries a smart device called SENSg [29] with various sensors embedded to record their mobility and sense their surrounding environment everyday. The SENSg device collects sensing data every 15 seconds when it is in active mode. The sensor readings are uploaded to the server through nearby wireless hotspots whenever there are (Wireless@SG [30] provides WiFi coverage with over 4,000 hotspots in Singapore and offers free data bundle to the SENSg devices in NSE. WiFi deployed at schools also can be used to upload sensor readings and are the most frequently used). All data used in the project are anonymized to protect the students' privacy.

Each uploaded record contains 16 attributes, including raw sensor readings and some derived results. Specifically, each record contains the *ID* of the SENSg device, and two timestamps, *i.e.,* ts referring the time the readings are taken at device side and rts referring the time data are received at the server side. Each record has following mobility attributes:

- **Location.** The location is represented as a pair of latitude and longitude, indicating the current position when the sensing data is taken. The location is derived from the MAC addresses and RSSIs of nearby WiFi hotspots and through a third-party localization service called SKYHOOK [23]. In addition to the location itself a localization error estimation is also provided. As shown in Figure 2, the localization service based on WiFi hotspots has varied localization error ranging from tens to hundreds of meters. The 50-percentile and 90-percentile localization errors are 80 m and 143 m, respectively.
- **Step count.** The step count is inferred from the IMU sensor readings in SENSg. The server side keeps tracking the accumulated steps the SENSg user takes.
- **Travel mode.** It indicates the mode of transportation (*i.e.,* walking, bus, MRT train, or private car) the SENSg user takes when the sensing data is taken, which is also inferred from the IMU sensors through an online classification algorithm.

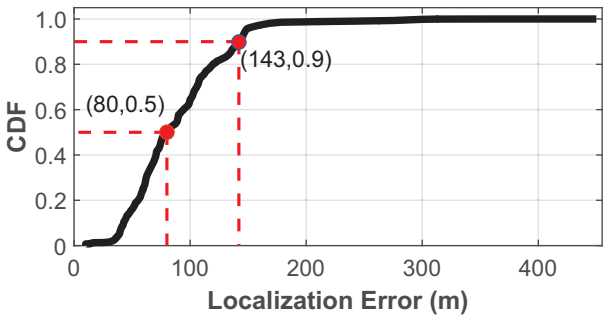


Figure 2: CDF of localization error in NSE mobility data.

Besides the mobility attributes, each record includes other environmental attributes, *e.g.,* temperature, atmospheric pressure, relative humidity, sound pressure level, light intensity, *etc.*. In our study, we rely on mobility attributes from the records of each student which give the student's daily footprint in the city. This study utilizes the mobility data from a total number of 108,337 students for 11 weeks in 2016, which correspond to a distance of trajectories totaling billions of kilometers.

As the natives, students are most familiar with the neighborhood around their home and school. They often make use of the uncharted walkable areas or paths to save their commuting time and their walkway choices contain novel knowledge to existing digital road maps (*e.g.,* open-sourced OpenStreetMaps [20] and commercial maps like Google Maps [6]). In this paper, we primarily exploit the walking trajectories of students to discover the missing walkways.

3 THE SYSTEM

In this section, we present the system overview and then elaborate each component in the following subsections.

3.1 System Overview

The system architecture of *VitalAlley* is illustrated in Figure 3. At a high level, *VitalAlley* consumes the location and step count data from NSE mobility dataset and offers representative walkways to complete existing road maps. Due to localization errors and characteristics of students' mobility (*e.g.,* mainly being active in school and home), not all location data are useful for the walkway discovery. At the very beginning, *VitalAlley* filters out the location data which are not in *walking* status according to the travel mode

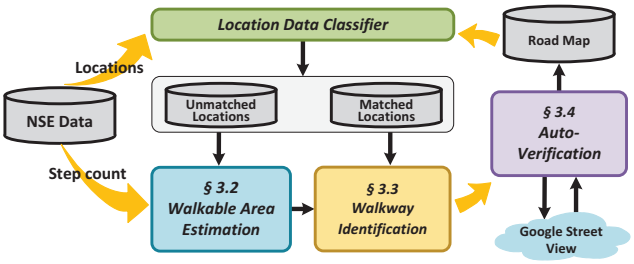


Figure 3: The system architecture of *VitalAlley*.

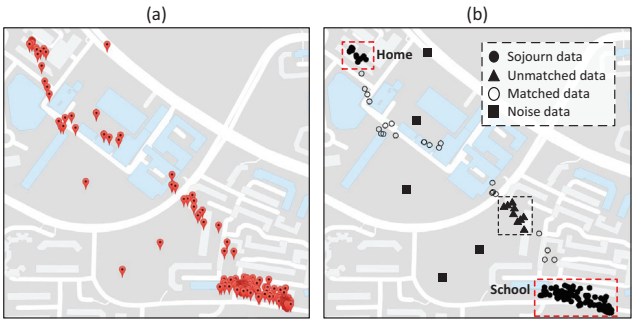


Figure 4: Illustration of location data classification on the mobility data of one student in a typical school day.

in the records. Then it invokes the *Location Data Classifier* module to label the location data as "Noise", "Sojourn", "Matched" and "Unmatched". This module firstly labels the locations with large localization error estimations (*e.g.*, $\geq 143\ m$) as noise data, which are far away form the student's trajectory. The common sojourn places for students are schools, homes, shopping malls, *etc.*, which leads to a much higher data density than usual at a specific place. Thus the module labels such data taken at some sojourn places as sojourn data by exploiting the HDBSCAN algorithm [5]. After filtering out noise and sojourn data, this module classifies the remained data into "Matched" and "Unmatched" through a map matching algorithm [17]. Matched data are those that can well match with existing roads, while the unmatched data cannot well match with any road. Only "Matched" and "Unmatched" locations are helpful for *VitalAlley* as they may contain knowledges about the missing walkways.

Figure 4 illustrates the four types of location data. Figure 4(a) plots the raw mobility data of one student in a typical school day, and Figure 4(b) shows the classification results. The two clusters correspond to home (top left corner) and school (right bottom corner), respectively. The noises are far away from the student's actual trajectory, and the matched locations are distributed near some existing road segments while the unmatched ones locate away from the roads.

Based on the results of *Location Data Classifier*, *VitalAlley* will identify and verify the representative walkways through several modules. Specifically, the *Walkable Area Estimation* module (in Section 3.2) approximates possible walkable areas, within which the *Walkable Path Identification* module (in Section 3.3) extracts

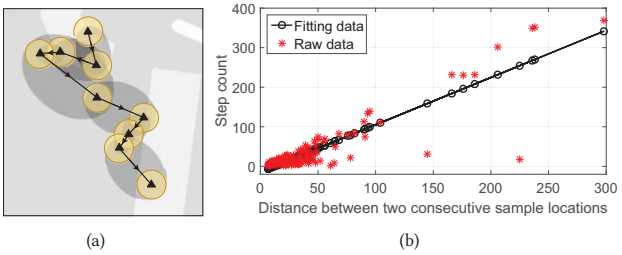


Figure 5: (a) The ellipse model based walkable area estimation *v.s.* individual sample based estimation; (b) Linear relationship between steps taken and sample distance.

some representative walkways. Finally, the *Auto-Verification* module (in Section 3.4) verifies the new-found walkways by analyzing their images retrieved through GSV APIs. All discovered walkaways are integrated into current road map for the public uses.

3.2 Walkable Area Estimation

A straightforward approach to deriving walkable area from coarse locations is to calculate the possible coverage of each mobility sample given a typical WiFi hotspots based localization error (*e.g.*, 80 *m*) in urban city and take the whole covered area as the walkable area estimation. Such an approach simply treats mobility data and fails to exploit the multi-modality sensor data and trajectory information between consecutive locations. Instead, *VitalAlley* proposes an ellipse model to estimate the walkable area between any two consecutive mobility data, jointly leveraging the location and step count data.

Benefits of step counts. When considering the total steps taken from one location to the next, we can estimate the length of walking path between the two locations as the product of total steps and stride length. Such an estimation of walking distance along with the locations themselves together bound the possible walkable area, which form an elliptical region. Based on this observation, We thus propose an ellipse model to estimate the walkable area. In theory, an ellipse is determined by two parameters: 1) positions of two focal points; and 2) sum of distances d from every point on boundary of ellipse to the two focal points. The locations of two consecutive mobility data indicate the focal points and we determine d using the step count data in records. *VitalAlley* calculates d as $\Delta s \times \lambda$, where Δs is the total steps taken between the two locations (*i.e.*, difference of step counts in the two records) and λ is stride length and set as the average stride length of young as $0.78\ m$ according to a recent report [2]. Though the stride length can be accurately estimated for each individual, it requires extra efforts and hardware [9], which is infeasible in the NSE project. With such settings, *VitalAlley* can produce an oval area to capture the walkable area between two locations. Figure 5(a) presents an example of the walkable area estimation from a series of unmatched mobility data. Compared with the disjoint walkable area estimation (*i.e.*, yellow area) derived from coverage of each individual location, the ellipse model can provide more reasonable walkable area estimation (*i.e.*, grey area).

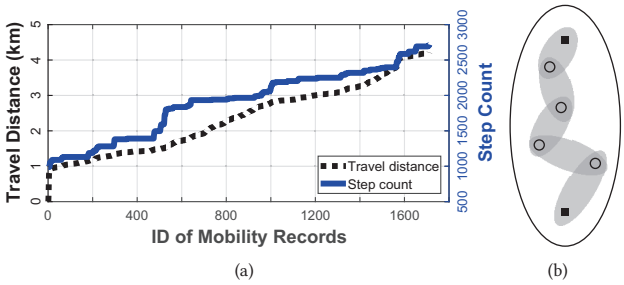


Figure 6: (a) The step counts and accumulated traveling distance of one student over a school day; (b) Fine-grained estimation derived from step count estimator, where ○ denotes locations whose step counts remain and ■ denotes locations whose step counts increase.

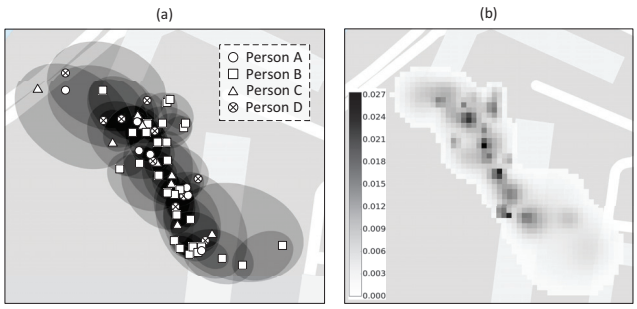


Figure 7: (a) The walkable area estimation from multiple trajectories. (b) The corresponding score map of (a).

Step estimator. Although step counts benefit the walkable area estimation, they are not perfect in the NSE mobility data. To reduce computation overhead of the SENSg device, step counts as the derived results are updated with a longer period than the updating period of locations. Figure 6(a) plots the step count data and accumulated travel distance of a student over one typical school day, where travel distance is calculated from the locations. From this figure, we can see the travel distance keeps increasing while the step count data are discontinuously updated, where flatten lines indicate that the step counts remain the same as previous one. Due to this fact, we can only use the records, whose step counts and locations are simultaneously updated, for walkable area estimation. As a result, the ellipse model can only provide coarser estimations using a few valid records.

To improve the performance of the ellipse model, we propose a step estimator which can estimate the total steps taken between two consecutive locations. We find an interesting relationship between the geographical distance of two consecutive locations and the total steps Δs between them. We extensively study their relationship and plot the statistics in Figure 5(b), which clearly shows that the distance of two consecutive locations and Δs are linearly correlated. Therefore, we can build a step estimator for each student by training a linear regression model based on her records with both step count data and location data available. Since the ellipse model is related to d which is determined by total steps Δs, we thus are able to determine an ellipse only based on the locations of two consecutive records even when their step counts are not updated. The step estimator enables us to derive more fine-grained walkable area estimation, just as demonstrated by gray region in Figure 6(b). We can only generate a large oval area with two records having both updated locations and step counts, while a fine-grained walkable area can be derived by exploiting the step estimator.

Weighting scheme. We further refine the walkable area estimation through a weighting scheme. For an oval area derived from two consecutive records, different parts inside the region may be traveled by pedestrians with varying probabilities. For example, when someone walks from one position to another, it is highly possible for her to walk along the straight line. Based on this intuition, we propose a weighting scheme to assess the probability of being traveled for different parts inside the oval area. To achieve that, we use a bivariate Gaussian model [32], which has been frequently used for home range estimation of animal movements in the biological domain [8]. Researchers in biological domain have limited periodical observations of where the animals are and they need to estimate the moving region of animals. They adopt this model since it can provide the probability of being a part of home range for a specific region. Similarly, we also have limited observations from pedestrians and need to estimate their possible movement area. In our scenario, the bivariate Gaussian model measures the probability density of an area being traveled by the pedestrians. We partition the space of interest into small cells and assess the probability of each cell inside an ellipse. We set the cell size as $2.0\,m \times 2.0\,m$ for a better resolution. The probability density of being traveled by pedestrians for location X is measured by:

$$f(X) = \frac{1}{2\pi\sqrt{|\Sigma|}} exp(-\frac{1}{2}X^T\Sigma^{-1}X), \quad (1)$$

where $X = [x_1, x_2]^T$ and $\Sigma \in S^2_{++}$ is the covariance matrix of X. Σ is calculated as $\Sigma = [a/3, 0; 0, b/3]$, where a and b are major semi-axes and minor semi-axes of the ellipse edges. The probability of each cell being passed by the pedestrians is the integral of Equation (1) on the cell's area. The probabilities of cells outside ellipse are negligible.

We can easily extend the scheme to all mobility data collected from many students. For any two consecutive records, *VitalAlley* generates an ellipse and assesses the probability of each cell inside the ellipse. An area covered by more ellipses should be walkable with a higher probability. In other words, these areas are "scored" by the mobility data via ellipses. For any cell, its final score is the sum of probabilities assessed by oval areas covering on it. A cell with larger score implies that it is more likely to be walkable. Finally we can derive a score map of the space of interest. Figure 7(a) presents the ellipses derived from the mobility data of four persons, and Figure 7(b) shows the corresponding score map based on the weighting scheme. We can see that the cells covered by more oval areas have larger scores, *i.e.,* in darker color.

3.3 Representative Walkway Identification

We propose a two-phase clustering to identify some representative walkways from each estimated walkable area. The representative walkways sketch the accessibility of a walkable area and usually

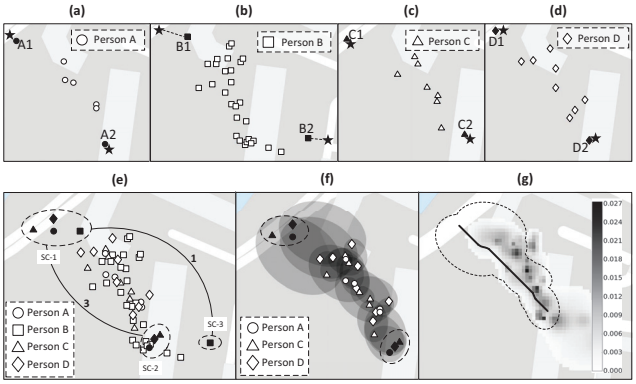


Figure 8: An illustrative example for representative walkway identification. (a-d) plot the mobility data of person A, B, C, and D, respectively. (e) demonstrates the two-phase clustering. (f) shows the covered area by ellipses of trajectory cluster <SC-1, SC-2>, and (g) presents the searching space and derived representative walkaway of trajectory cluster <SC-1, SC-2>.

connect certain road segments or POIs at some intersections for convenient mobility. Such intersections act as the entries/exits of a walkable area. Therefore, we conduct *location clustering* on some matched location data to identify the intersections, and then run *trajectory clustering* on unmatched location data to extract representative walkways that link the derived intersections. In the following, we present the details of each clustering phase.

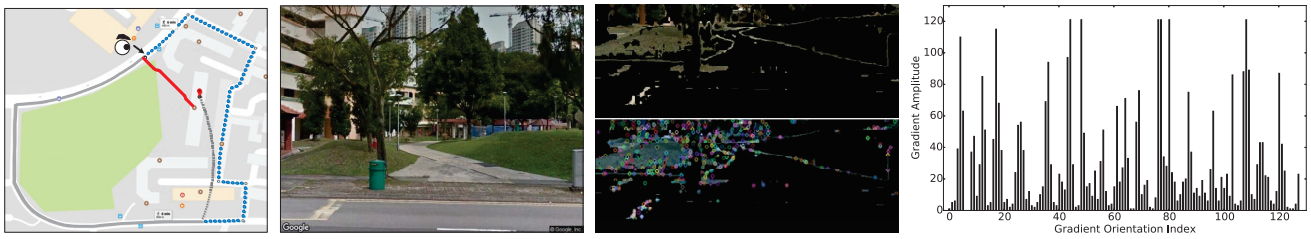
Phase 1: location clustering to identify the intersections. To derive the entries/exits of a walkable area, we make use of the last matched locations collected near the walkable area. A last matched location data is the data that can be well matched with known road segments, after/before which other mobility data cannot be matched. We put the last matched data matching with the same road segment (or POI) together, which may correspond to one or several potential entries/exits. A walkable area may intersect with the same road segment/POI at several points, resulting in multiple entries/exits. Thus location clustering firstly groups the last matched data matching with the same road segment/POIs and then runs the HDBSCAN algorithm [5] on each location group to further classify those locations into one or several clusters, each of which indicates an intersection connecting one road segment/POI and the walkable area. We calculate the position of an intersection through all locations in the same cluster. It is the location on a road segment that has the minimum distance to all last matched locations of the cluster; or just as the average location when the location cluster falls in a POI, *e.g.,* residential building.

Figure 8(a-e) illustrates the location clustering. Figure 8(a-d) plot the mobility data collected from four persons A, B, C, and D, in the same area. Their unmatched locations are denoted by ○, □, △ and ◇, respectively. Meanwhile, their last matched locations are denoted by ●, ■, ▲ and ◆, and the corresponding matched locations are denoted by ★. Figure 8(e) shows that location clustering groups the last matched data into three location clusters, *i.e.,* "SC-1", "SC-2", and "SC-3", which correspond to three entries/exits of the walkable area. "SC-1" and "SC-3" locate on different road segments while "SC-2" locates on one residential building. Through the location clusters, we can calculate the positions of intersections.

Phase 2: trajectory clustering to extract representative walkways. The trajectory clustering takes advantage of the results from the first phase to group trajectories. In practice, each trajectory is associated with two last matched data, and we thus can use their corresponding location clusters to annotate the trajectory. For example in Figure 8(b), location B1 and B2 are paired last matched data of trajectory B, and we annotate this trajectory as <SC-1, SC-3>. We annotate all trajectories following the same rule, and thus can simply group trajectories with the same annotation. Even for the same pair of intersections (*i.e.,* entries/exits), different trajectories exist and may correspond to different representative walkways. Thus we still need to run a typical clustering algorithm [27] to further classify the trajectories in the same group into different clusters according to the metric of Hausdorff distance. Each trajectory cluster corresponds to a representative walkway, and we use its size (named as *support*) to assess whether the potential walkway is frequently used or not. In Figure 8(e), there are two trajectory clusters and their supports are 3 and 1, respectively.

In principle, the potential representative walkway of a trajectory cluster should be a polyline on the score map of a walkable area, traversing the cells with high scores. To find the walkway most likely to be traveled in reality, we transform the score map into a weighted graph and model our representative walkaway identification problem on this graph. For the score map of a walkable area, we generate a graph, where cells are represented as the vertices and edges are formed between any two immediately neighboring cells. In the graph, we set the weight of each vertex as the reciprocal of the score of corresponding cell. For cells with zero score, we set their weights as an maximum value, which implies that those cells are barely traveled by the pedestrians. Each vertex also inherits the position of the corresponding cell. In such a graph, we define a real-valued weight function $f : V \rightarrow \mathbb{R}$ that returns the score of a given vertex v_i, and we will find a representative walkaway $P = (v_s, v_1, v_2, \cdots v_n, v_d)$ between the two intersections v_s and v_d over all possible n vertices such that the total score of its constituent vertices $\sum_{i=1}^{n} f(v_i)$ is minimized. We can run the A^* algorithm to find the optimal solution. Finally, mapping vertices back to the cells on original map obtains the representative walkway.

Optimizations. *VitalAlley* also incorporates some optimizations to speedup the identification of representative walkways. For each trajectory cluster, we extract a subgraph from the original weighted graph by only keeping the vertices (and the associated edges) covered by oval areas of records belonging to this cluster. We thus reduce the searching space for a specific representative walkway. In addition, *VitalAlley* simply discards the trajectory clusters owning small supports, *e.g.,* $< \alpha$, just because they are barely traveled by people. For the example in Figure 8, we plot the covered area by ellipses of trajectory cluster <SC-1, SC-2> in Figure 8(f), which is derived from the unmatched trajectories of person A, C, and D. To identify the representative walkway, *VitalAlley* shrinks the searching space by filtering out irrelevant cells. The final searching space is shown as the dashed region in Figure 8(g). The black line in Figure 8(g) is the representative walkway, connecting a road segment and a residential building.



(a) An example of one newfound representative walkway. (b) The GSV image captured from the position of intersection. (c) One of the clustered images and its key points for the image in (b). (d) One typical key point descriptor illustration.

Figure 9: The illustrations of image processing on the GSV images of one discovered walkway.

3.4 Auto-Verification

We propose an auto-verification method to verify the discovered walkways by invoking the GSV APIs [7]. GSV offers panoramic street views from different positions along the streets in a city, and we are able to retrieve images for a walkway at its entry/exit, which usually locates at some main road. As an independent data source, GSV provides a complementary angle to investigate new-found walkways.

We find that the entries/exits of walkways in a city are relatively regular and pertain a limited number of types. Thus it is possible for us to collect images of typical entries/exits from GSV in advance and treat them as the templates stored in a local library for the auto-verification of walkways.

Retrieving images of walkways from GSV. For each candidate walkway, we project its entry/exit (*i.e.,* the intersection derived from location clustering) to the main road and retrieve multiple images at the projected points by invoking GSV APIs, which need the parameters of location (*i.e.,* projected points), size (*i.e.,* 640×640), field of view, and heading directions. We take the projected point as the origin and generate two rays from this point to include all location points of the last matched data that generate the intersection. The two rays form an angle θ and we retrieve $\lceil \theta/30^\circ \rceil$ images from GSV by varying the heading direction every 30° within the angle. As a concrete example, Figure 9(a) presents a new-found representative walkway that connects a road segment and a residential building, and Figure 9(b) shows the retrieved GSV image from the position where there is an eye in Figure 9(a). From this GSV image, we can clearly see a walkway.

Image matching. For each walkway candidate, we extract the features (such as corner, boundary) from GSV images as the key descriptor using conventional image processing techniques [13]. For one GSV image, we first segment the image into κ clusters based on the pixel values, where we set $\kappa = 5$ for better performances. Second, we remove the upper part of the segmented image because walkways usually locate at the bottom part of an image. Then from the cut image, we extract the key points, which are around the borders of segmentations and describe the image [16]. Then we calculate the image gradient magnitudes and orientations of key points to form an image descriptor, which is denoted as a histogram. We match it with template images by comparing their image descriptors in χ^2 distance, which returns a measure of difference. Once the difference of at least one GSV image is lower enough, *e.g.,* $\leq \rho$, we declare that this candidate walkway is verified and mark it as "*confirmed*"; otherwise as "*suspicious*". We set $\rho = 10$ as the default setting in this work.

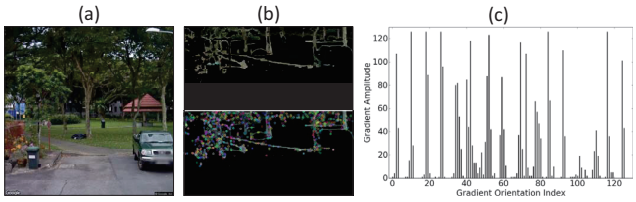


Figure 10: The matched template image in the library.

For the GSV image in Figure 9(b), we derive the image processing results shown in Figure 9(c) and Figure 9(d). In Figure 9(c), the upper figure is the bottom part of one of the 5 clustered images and the bottom figure is the corresponding key points highlighted by circles. Figure 9(d) shows the image descriptor of one key point in Figure 9(c), where bins on x-axis are the gradient orientations and y-axis indicates the amplitude of associated gradient. We match this image with the templates in the library and find one template image shown in Figure10(a). Its clustered bottom part and key points are shown in Figure 10(b). Figure 10(c) is a descriptor of one key point. After the calculations of their descriptors on the χ^2 distance, we find they have low difference with $\chi^2 < 10$, which indicates that the new-found walkway is verified and really exists.

Since GSV does not have images for the whole city, we may retrieve no image at some locations. For such cases, we mark the candidate walkways as "*suspicious*" for prudent use by the publics. We integrate all discovered walkways into existing road map and provide some hints to remind the users whether a walkway is verified or not.

Building the template library. We can build a local template library in advance for the auto-verification. Building a library containing all the templates is out of the problem due to infinite cases. We argue that, however, the types of different walkways in a specific city are limited so we only consider the most frequently used walkways. This library consists of four categories of walkways in total. Each category has many template images and the corresponding image features, which describe the typical characteristics of entries/exits of walkways in a city. Specifically, we review the images of intersections between walkways and main streets from GSV

Figure 11: All new-found walkways discovered from the NSE mobility data.

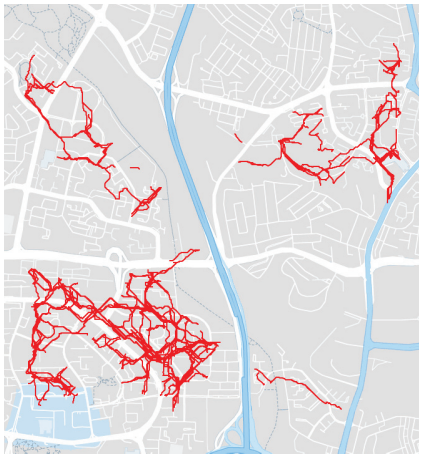


Figure 12: The study area.

and select some regular entries/exits to collect their images as the templates for building the local library. For each typical intersection, we capture 3 images from the street with heading directions at the medial axis of the walkaway and 30 degrees offset of the medial axis. For each image, we extract its image descriptor using the same image processing techniques. We can manually update this library by adding more GSV images of new kinds of entries/exits.

4 EVALUATION

In this section, we implement *VitalAlley* and extensively evaluate its performances using the NSE mobility dataset. We make our source code publicly accessible through GitHub [26].

4.1 Experimental Setup

We implement *VitalAlley* in Python and run the system in a powerful HP Z440 workstation that has 12 3.5GHz Intel Xeon CPU cores and 32GB memory. For data preprocessing, we realize a custom HDBSCAN algorithm [5] to filter out the sojourn mobility data and implement an hidden Markov model based map matching algorithm [17] to accurately match mobility data with a base road map. We leverage some open-sourced Python libraries, *e.g.,* matplotlib, networkx, numpy, *etc.* to compute, visualize, and verify the results on a base road map, as well as capture the screenshots in this paper. We also implement the automatic verification module by invoking the GSV API [7]. We use the NSE mobility dataset and an open-sourced road map for the evaluations of *VitalAlley*.

NSE mobility dataset. The dataset contains mobility data collected from 108,337 students for 11 weeks in 2016 in the NSE project, containing more than 400 million records of totaling billions of kilometers in distance. As mentioned in Section 2.2, each record in the dataset includes 16 attributes about the mobility and environmental parameters during a student's daily activity. *VitalAlley* mainly uses the timestamps, step counts, and locations for walkway discovery.

Base road map. We obtain our base road map from the open-sourced OpenStreetMaps (OSM) [20], which is the largest crowd-sourced mapping project with more than 2 million registered contributors. In this paper, we use the OSM Singapore map of 20/8/2016 for map matching and evaluations, which already includes many passable roads in Singapore. We complete this map by supplementing the missing walkways discovered from daily trajectories of local students. To accelerate the walkway discovery for the whole city, we divide the OSM Singapore map into four regions as shown in Figure 11, *i.e.,* $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$, and run *VitalAlley* for each region in parallel using the multi-threading technique.

After the location data classification, *VitalAlley* divides mobility data into four types: *sojourn* data that are sampled when students are staying at some specific place, *e.g.,* school or home; *noise* data that are with large estimated localization errors and thus far away from the student's trajectory; *matched* data that can well match with the OSM road map; and *unmatched* data that cannot well match with existing roads. Since the schools and homes of students are non-uniformly distributed in Singapore, the records are dispersed in the map. Figure 13 presents the concrete proportions of each data type collected in each region. According to our statistics, the average portion of the four data types (*i.e.,* sojourn, noise, unmatched, and matched) are 80.1%, 0.8%, 8.6%, and 10.5%, respectively. For all regions, the sojourn data occupy the largest portion, which is reasonable as students spend most of their time at school and home. For *VitalAlley*, unmatched data are the most valuable input, which provides implicit information about the missing walkways. The region $\mathcal{D}$ owns the most unmatched mobility data, where we thus may discovery more walkways.

Performance metric. We define $accuracy = \frac{N_{true \cap new}}{N_{new}}$, where $N_{true \cap new}$ denotes the number of new-found walkways that are truly exist and N_{new} means the total number of walkways discovered by our system. The larger *accuracy* is, the better performances *VitalAlley* has.

4.2 Evaluation Results

In this subsection, we first present and analyze the overall results of walkway discovery on the whole OSM Singapore road map, and then conduct a detailed evaluation in a study area to carefully evaluate the accuracy of *VitalAlley* and the effects of system parameters.

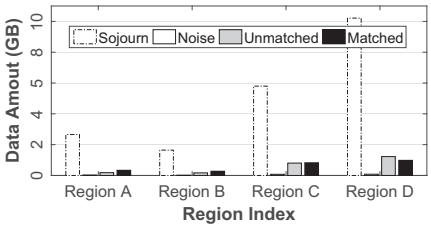


Figure 13: Statistics of data types in each region.

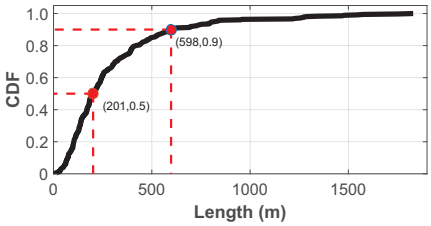


Figure 14: The length distribution of all new-found walkways.

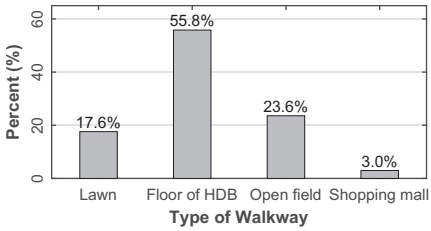


Figure 15: Different types of all discovered walkways.

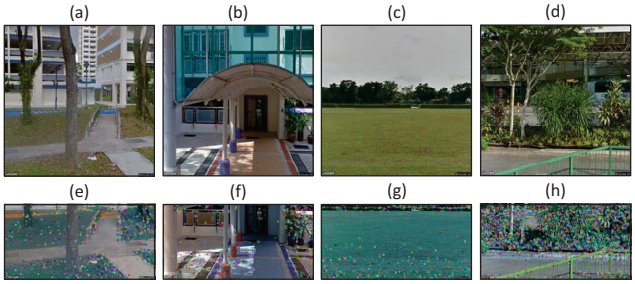


Figure 16: GSV images of the new-found walkways.

4.2.1 Overall performances. Figure 11 depicts all the new-found walkways in Singapore by leveraging the NSE mobility dataset. The discovered walkways scatter over the road map and connect the motorways and POIs. In total, we have discovered 736 walkways, with accumulated distance more than 150km. The number of discovered walkways for region $\mathcal{A}$, $\mathcal{B}$, C, and $\mathcal{D}$ are 110, 109, 147, and 370, respectively. In principle, a region with more mobility data will probably have more discovered walkways. In region $\mathcal{D}$ which is near the downtown area of Singapore, we have the most new-found walkaways as we have the most mobility data there. From Figure 11, we see the walkways are in different lengths, and we present the statistics of lengths of the new-found walkways in Figure 14. Almost all of the new-found walkways are with lengths less than 1 km, and 90% of the walkways are shorter than 598m. This is reasonable as walkways usually act as the small shortcuts between road segments/POIs. According to our statistics, the shortest walkway is only 11 m, while the longest walkway achieves 1829 m.

For each discovered walkway, we invoke the GSV APIs to retrieve its images for auto-verifications, and also check the type each walkway belongs to. Among all the new-found walkways, we mainly have four types of walkways, namely lawn, floor of HDB (the typical residential buildings in Singapore), open field and shopping mall, which account for 17.6%, 55.8%, 23.6% and 3.0%, respectively, as the statistics shown in Figure 15. Walkways derived from the floors of HDB account for the largest proportion mainly because the HDB building widely exist in Singapore and many of such residential buildings have the first floor empty where people can walk through. Besides, the lawns and open fields are also frequently used by the local residents as the shortcuts for accessing to other neighborhood. The interiors of shopping malls are also occasionally used as walkways by some students who are familiar with the inner connectivities of those malls. Figure 16 shows the images of some typical walkways we find in Singapore. Figure 16 (a) plot the walkways in the lawn, where we can see some clear walking tracks used by the pedestrians. Figure 16 (b) shows the walkway around the HDB buildings, which is a small trail connecting a road segment and the HDB residential building. Figure 16 (c) shows a walkway in an open field which is a flat ground, and Figure 16 (d) shows a walkway within a shopping mall, which connects two parallel road segments. Figure 16 (e-h) illustrate the key points of each type of walkways above. The features of walkway in shopping mall are complicated, as shown in Figure 16 (h). We identify walkways passing through shopping malls by exploiting additional information (*e.g.,* positions and names) from OSM. Such walkways widely exist in Singapore and facilitate people's daily mobility.

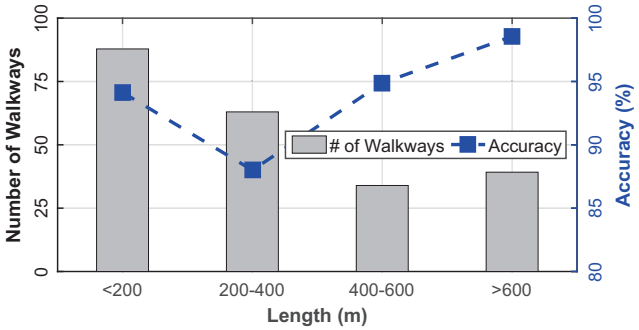


Figure 17: Accuracies for walkways of different lengths.

Accuracy evaluation in a study area. To evaluate the accuracy of our system, we select a small region around the central area of Singapore (*i.e.,* the rectangle region in Figure 11) covering about 9 km^2 area to conduct a detailed study. Within this region, we have discovered 224 walkways in total, as shown in Figure 12. For each new-found walkway in the study area, we manually investigate whether it is true. If we are able to find a walkway according to the positions calculated by our system and the walkway indeed connects two road segments/POIs indicated by our location clustering, then we consider this walkway as true in reality; otherwise we think it is a false walkway and cannot be practically used by the pedestrians. For this specific study area, we finally find 209 true walkways with manual verifications, with the accuracy as high as 93%.

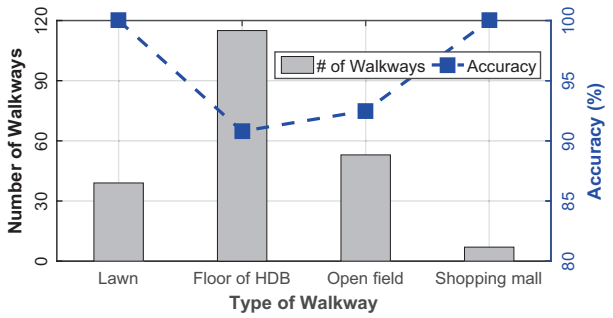


Figure 18: Accuracy of different types of walkways.

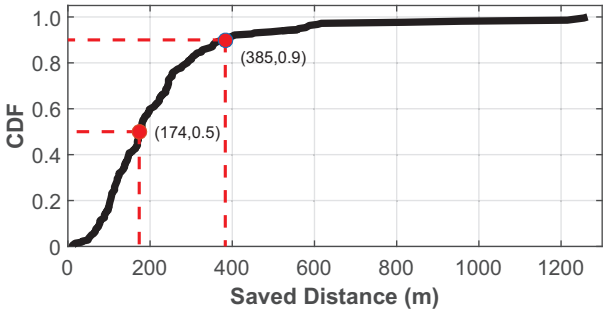


Figure 19: Utility of walkways for route planning.

We explore the accuracies for walkways of different lengths and present the results in Figure 17. Specifically, we classify all walkways into four groups according to their lengths (in m), *i.e.,* $(0, 200]$, $(200, 400]$, $(400, 600]$, and $(600, \infty)$. Among the four groups, we find a large number of walkways are with short length, *i.e.,* $\leq 200\,m$. The distribution of walkways on length is in accordance with the overall distribution in Figure 14. Most walkways are in median lengths and there are few walkways too long, *e.g.,* > 600. With respect to accuracy, group $(200, 400]$ has the lowest accuracy as 88%, and group $(0, 200]$ has the highest accuracy as 94%.

We also study the accuracies for different types of walkway, and present the results in Figure 18. For lawns and shopping malls, *VitalAlley* achieves 100% accuracy, which means that all the new-found walkways on lawns and shopping malls are verified as truly "walkable" after manual investigations. For walkways on lawns, the auto-verification module can accurately identify them as the features of lawns are prominent in images. Besides, our base road map contain the information about shopping malls and once a walkway passing through a shopping mall can be easily identified. As long as those features are extracted, it is highly possible that the walkways endorsed by the GSV-based auto-verification module exist in reality. For walkways belonging to types of "floor of HDB" and "open field", *VitalAlley* still achieves accuracies of as 91% and 93%, respectively.

Utility study of discovered walkways. The motivation for walkway discovery is that those widely existing walkways can provide significant convenience and largely save the walking time when compared to using existing road segments. Thus we conduct a utility study for the discovered walkways in the study area. We initiate 100 travel demands with trip origins and destinations being the students' homes and schools. We implement a shortest path based route planning algorithm, which will return a shortest walking route based on the available road map. For each pair of trip origin and destination, we generate two routes based on the original OSM road map and the new road map supplemented with our new-found walkways. We calculate the distance difference between the two routes and adopt the saved walking distance when compared to the one planned on the original road map as the measure of utility. The larger the saved walking distance is, the more efficiency the walkways can provide. We calculate the distance differences of the 100 trip queries, and plot the statistics in Figure 19. For all trip planing queries, the routes derived from the completed road map are shorter than the ones walking on existing roads. Among all queries, 50% can save walking distance than 174 m, and the 90-percentile saved walking distance is 385 m. The results in Figure 19 demonstrate that it is necessary to discover those missing walkways as they really facilitate people's mobility.

4.2.2 Parameter Setting. We evaluate the impacts of system parameters and design choices in the study area.

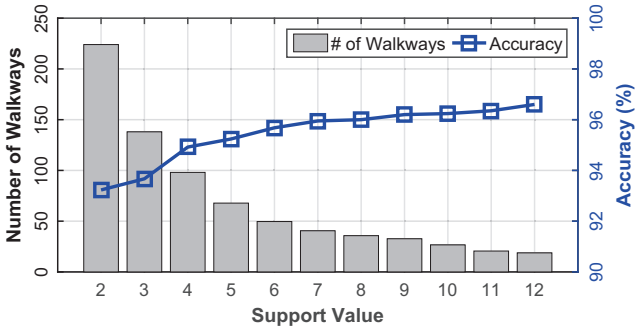


Figure 20: The impacts of support value on number of discovered walkways and accuracy.

Impacts of support value. In the representative walkway identification module, we use a threshold to filter out some trajectory clusters with small support. In fact, the support reflects how likely a trajectory cluster is to generate a representative walkway used by enough pedestrians. Figure 20 shows the effects of choosing different supports as the threshold on the number of discovered walkways and accuracy. The support value is ranged from 2 to 12. When $support = 2$, nearly 224 walkways are correctly found in the specific region with an accuracy of 93%. Larger support requires a walkway traveled by more pedestrians and thus leads to a decreased number of discovered walkways. On the other hand, the accuracy increases along with increase of support values. Only 31 walkways are found with a high accuracy as 97% when we set $support = 12$. According to Figure 20, we find $support = 3$ can balance both number of discovered walkways and the accuracy, and we thus set $support = 3$ as the default setting.

Impacts of data amount used. The data amount used will also affect the effectiveness of walkway discovery. With more data, we can collect the observations about more potential walkways and it

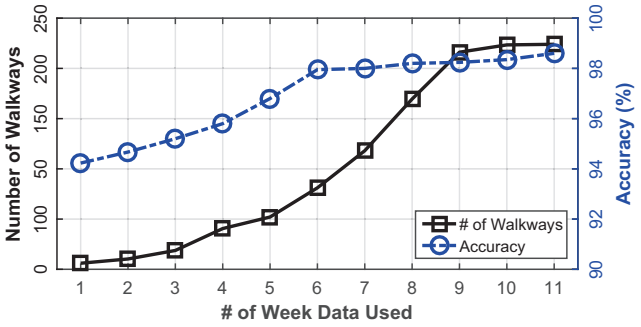


Figure 21: Impacts of data amount used on the number of discovered walkways and accuracy.

is more possible to accurately identify them. We vary the number of weeks of mobility data used, and present the resulting number of discovered walkways and accuracy in Figure 21. When we increase the amount of used mobility data, the number of discovered walkways gradually increase as well. It is reasonable since more mobility data will include the trajectories from more walking paths and we thus can identify more walkways. When we use data more than 9 weeks, the benefit becomes negligible as the frequently used walkways have been covered by existing data and extra data may provide little new knowledges. On the other hand, the impact of data amount on the accuracy is less significant. More data lead to slight increase on the accuracy. When we use more than 6 weeks of data, the accuracy becomes stable around 98.5%. Extra data bring no increase on accuracy. Therefore, 9 weeks of mobility data seem to be a good amount for accurate and comprehensive walkway discovery.

Table 1: Effect of GSV based auto-verification on the accuracy of walkway discovery.

Support	2	4	6	8
w/ GSV	93.2%	94.8%	95.7%	96.0%
w/o GSV	80.9%	88.6%	93.5%	95.8%

Impacts of GSV based auto-verification. We invoke the GSV APIs to retrieve images of the discovered walkways and automatically verify them based on computer vision techniques. To evaluate the performance of this module, we conduct experiments *with* and *without* the GSV-based auto-verification module. We use 9 weeks of mobility data during the comparison and vary support from 2 to 8. The results are shown in Table 1. Note that for both with and without this module, the system can find the same number of walkways. By enabling this module, the system can always achieve higher accuracy than the case without this function, having the largest performance gain of 12.3% on accuracy when $support = 2$. When we increase the support value, the gain benefited from GSV becomes small as the parameter of support value will filter out many invalid walkways.

Impacts of step estimator. In the system design, we propose a step estimator to enhance the walkable area estimation. Thus we also run experiments based on only available step count data

Table 2: Effect of step estimator.

	# of walkways	Accuracy
Basic method	95	93%
Enhanced method	224	94%

(denoted as *basic method)* and enabling the step estimator (denoted as *enhanced method*). Table 2 shows the comparison results. For both methods, they achieve similar accuracy while the enhanced method still has a bit higher accuracy. On the other hand, *VitalAlley* can discover much more walkways with the step estimator when compared using the step count data only, *i.e.,* 224 *vs* 95 with an improvement of 136%. Since the step count data in the NSE project may not be continuously updated, thus some mobility data cannot be directly exploited by the basic method. Once the step estimator is trained for a student, all the mobility data can be used by the enhanced method and as a result more walkways are found.

5 RELATED WORK

Map inference. Map inference aims to automatically generate the whole map from the satellite images or GPS trajectory data. The aerial imagery methods [21, 35] employ image processing techniques to draw only the main roads due to the limit of image resolutions, and thus cannot be used for discovering small trails like the walkways. There are three categories of methods proposed for the GPS trajectory data based map inference, *i.e.,* K-means [1, 31], Kernel Density Estimation [4, 25], trace merging and clustering [10, 14]. Most of these methods, however, build on various impractical assumptions of the GPS data, including low noise and high sampling frequency (*e.g.,* 1 Hz). They perform poorly once the assumptions are not hold [3, 14]. In contrast, our method preserves no assumption on the input data, and is more robust against noisy data like NSE mobility data.

Map updating. Map updating aims at completing a given road map by updating missing roads from GPS data. As discussed in Section 2, several recent works, *i.e., CrowdAtlas* [27], *COBWEB* [22], and *GLUE* [33], have been proposed to find missing roads, which are in particular the well-structured motorways, based on GPS data collected from vehicles. These works extract new-found roads mainly relying on trajectory clustering [27] or location point clustering [22, 33] coupled with some well-tuned thresholds, which thus make them be prone to failure when involving noisy data. In this paper, we consider to complete a given road map by updating the missing walkways from personal mobility data, which are noisy and random in nature. The unstructured and pedestrian-oriented characteristics of walkways implicitly make the problem more difficult and essentially distinguish our work from the existing map updating works. To the best of our knowledge, this is the first work to discover and update walkways from noisy mobility data at large scale.

Map Matching. Map matching is a technique to match location data to existing roads that minimizes the influences of localization errors. Existing map matching methods are classified into three groups, *i.e.,* based on the geometry of each road [28], based on the topology of the road network [19], and based on matching

probability [17]. In order to compensate for noise and gaps of individual trajectory, leveraging a collection of trajectories to do map matching has been proposed in [12], which increases the density of trajectories, leading to a better matching results, which is similar to [27]. Since the third one globally considers all location observations and achieves the best accuracy [24], in this paper we choose one representative algorithm in [17] to perform map matching on the mobility data. These works are parallel with our work and can benefit from a full road map completed by our system.

6 CONCLUSION

This paper presents *VitalAlley* for discovering and verifying the missing walkways from the NSE mobility data. *VitalAlley* proposes an ellipse model and a novel weighting scheme to estimate walkable areas, and identify representative walkways from each walkable area through a two-phase clustering method. The GSV is also involved to automatically verify the new-found walkways. Experimental results using the large-scale NSE mobility data demonstrate the performances of *VitalAlley* that finds 736 walkways (totaling 161 km in distance) for the OSM Singapore road map. A detailed evaluation in a study area shows that *VitalAlley* can achieve accuracy as high as 96%.

ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers and the shepherd for their constructive feedback and valuable comments for improving the quality of this paper. We also would like to thank National Research Foundation (NRF), Ministry of Education (MOE) of Singapore for the permission of access the National Science Experiment and its data. The access to the data is under a non-disclosure agreement. We acknowledge the support from Singapore MOE AcRF Tier 1 grant RG130/16 and Tier 2 grant MOE2016-T2-2-023, as well as the COE grant from NTU. This research is also supported by Natural Science Foundation of SZU under Grant No.2018061.

REFERENCES

[1] Gabriel Agamennoni, Juan I Nieto, and Eduardo M Nebot. 2011. Robust inference of principal road paths for intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems* 12, 1 (2011), 298–308.

[2] Tiago V Barreira, David A Rowe, and Minsoo Kang. 2010. Parameters of Walking and Jogging in Healthy Young Adults. *International Journal of Exercise Science* 3, 1 (2010).

[3] James Biagioni and Jakob Eriksson. 2012. Inferring Road Maps from GPS Traces: survey and Comparative Evaluation. *Transportation Research Board Annual* (2012).

[4] James Biagioni and Jakob Eriksson. 2012. Map inference in the face of noise and disparity. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*. ACM, 79–88.

[5] Ricardo JGB Campello, Davoud Moulavi, Arthur Zimek, and Jörg Sander. 2015. Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 10, 1 (2015), 5.

[6] Google. 2017. Google Maps. (Aug. 2017). Retrieved June 12, 2017 from http://maps.google.com

[7] Google. 2017. Google Street View. (Aug. 2017). Retrieved June 12, 2017 from https://www.google.com/streetview

[8] RI Jennrich and FB Turner. 1969. Measurement of non-circular home range. *Journal of Theoretical Biology* 22, 2 (1969), 227–237.

[9] Yonghang Jiang, Zhenjiang Li, and Jianping Wang. 2017. PTrack: Enhancing the Applicability of Pedestrian Tracking with Wearables. In *2017 IEEE 37th International Conference on Distributed Computing Systems*. IEEE, 2193–2199.

[10] Sophia Karagiorgou and Dieter Pfoser. 2012. On vehicle tracking data-based road network generation. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*. ACM, 89–98.

[11] Hengfeng Li, Lars Kulik, and Kotagiri Ramamohanarao. 2016. Automatic Generation and Validation of Road Maps from GPS Trajectory Data Sets. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 1523–1532.

[12] Yang Li, Qixing Huang, Michael Kerber, Lin Zhang, and Leonidas Guibas. 2013. Large-scale joint map matching of GPS traces. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 214–223.

[13] Robert LiKamWa, Bodhi Priyantha, Matthai Philipose, Lin Zhong, and Paramvir Bahl. 2013. Energy characterization and optimization of image sensing toward continuous mobile vision. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 69–82.

[14] Xuemei Liu, James Biagioni, Jakob Eriksson, Yin Wang, George Forman, and Yanmin Zhu. 2012. Mining large-scale, sparse GPS traces for map inference: comparison of approaches. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 669–677.

[15] Zhidan Liu, Shiqi Jiang, Pengfei Zhou, and Mo Li. 2017. A participatory urban traffic monitoring system: the power of bus riders. *IEEE Transactions on Intelligent Transportation Systems* 18, 10 (2017), 2851–2864.

[16] David G Lowe. 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60, 2 (2004), 91–110.

[17] Paul Newson and John Krumm. 2009. Hidden Markov map matching through noise and sparseness. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*. ACM, 336–343.

[18] NSE.SG. 2017. National Science Experiment. (Aug. 2017). Retrieved June 12, 2017 from https://www.nse.sg

[19] Washington Y Ochieng, Mohammed Quddus, and Robert B Noland. 2003. Map-matching in complex urban road networks. *Revista Brasileira de Cartografia* 2, 55 (2003).

[20] OpenStreetMap. 2017. Open Street Map. (Aug. 2017). Retrieved June 12, 2017 from https://www.openstreetmap.org

[21] Young-Woo Seo, Chris Urmson, and David Wettergreen. 2012. Exploiting publicly available cartographic resources for aerial image analysis. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*. ACM, 109–118.

[22] Zhangqing Shan, Hao Wu, Weiwei Sun, and Baihua Zheng. 2015. COBWEB: a robust map update system using GPS trajectories. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 927–937.

[23] SKYHOOK. 2017. Skyhook. (Aug. 2017). Retrieved June 12, 2017 from http://www.skyhookwireless.com

[24] Renchu Song, Wei Lu, Weiwei Sun, Yan Huang, and Chunan Chen. 2012. Quick map matching using multi-core cpus. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*. ACM, 605–608.

[25] Albert Steiner and Axel Leonhardt. 2011. Map-generation algorithm using low-frequency vehicle position data. In *Transportation Research Board 90th Annual Meeting*.

[26] GitHub User. 2018. Source code of Walkway Discovery from Large Scale Crowdsensing. (Feb. 2018). Retrieved February 5, 2018 from https://github.com/caochuntu/IPSN2018_guizu

[27] Yin Wang, Xuemei Liu, Hong Wei, George Forman, Chao Chen, and Yanmin Zhu. 2013. Crowdatlas: Self-updating maps for cloud and personal use. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 27–40.

[28] Christopher E White, David Bernstein, and Alain L Kornhauser. 2000. Some map matching algorithms for personal navigation assistants. *Transportation research part c: emerging technologies* 8, 1 (2000), 91–108.

[29] Erik Wilhelm, Yuren Zhou, Nan Zhang, Jacksheng Kee, George Loh, and Nils Tippenhauer. 2016. Sensg: Large-scale deployment of wearable sensors for trip and transport mode logging. In *Transportation Research Board 95th Annual Meeting*.

[30] Wireless@SG. 2017. Wireless@SG. (Aug. 2017). Retrieved June 12, 2017 from https://www.imda.gov.sg/wireless-sg

[31] Stewart Worrall and Eduardo Nebot. 2007. Automated process for generating digitised maps through GPS data compression. In *Australasian Conference on Robotics and Automation*, Vol. 6. Brisbane: ACRA.

[32] BJ Worton. 1987. A review of models of home range for animal movement. *Ecological modelling* 38, 3-4 (1987), 277–298.

[33] Hao Wu, Chuanchuan Tu, Weiwei Sun, Baihua Zheng, Hao Su, and Wei Wang. 2015. GLUE: a parameter-tuning-free map updating system. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 683–692.

[34] Pengfei Zhou, Yuanqing Zheng, and Mo Li. 2012. How long to wait?: predicting bus arrival time with mobile phone based participatory sensing. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*. ACM, 379–392.

[35] MJ Valadan Zoej and M Mokhtarzade. 2004. Road detection from high resolution satellite images using artificial neural networks. In *Proceedings of ISPRS Congress, Istanbul, Turkey*, Vol. 35. 104.