2018 17th ACM/IEEE International Conference on Information Processing in Sensor Networks

Demo Abstract: 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

Existing digital maps mainly focus on motorways and miss many walkways facilitating people's daily mobility especially for pedestrians. Based on in-depth analysis of massive human mobility trajectories collected from the National Science Experiment (NSE) in Singapore, we propose a system, discovering walkways from the large scale crowdsensing mobility data. In this demo, we show the new-found walkways discovered by our system in Singapore through a custom visualization platform.

1 INTRODUCTION

Digital maps are vitally important for route planning and navigation. Current digital maps, however, are vehicle oriented, missing many walkways that pedestrians travel with. Such useful walkways are uncharted on maps and cannot be made of use by the public.

In this demo, we show our new-found walkways discovered through a custom visualization platform. The technical details are in [2] and we briefly introduce the design principles here.

We discover uncharted walkways by analyzing the dataset of a crowdsensing project called National Science Experiment in Singapore [5], where students carry smart devices SENSg [6] to sense surrounding environment. SENSg can collect data such as the steps students take each day, and travel patterns. We design algorithms to overcome the unstructured property of walkways and noisy nature of crowdsensing data, finding walkways in a statistical manner. Different from vehicles moving on 1-dimensional direction following the motorways, pedestrians have a high freedom of 2-dimensional uncertainty. To tolerant such uncertainties, we leverage an ellipse to include all possible walking paths between two consecutive locations. The ellipse can be well defined by the locations and step count took in between. The region inside an ellipse is called *walkable area* where pedestrians can freely walk. We refine walkable areas by utilizing a bivariate Gaussian model to weight the region inside ellipses, inspired by home range estimation problem in biological field [3]. Dividing the map of interest into cells, we assess the probability of being walkable for each cell, given an ellipse using the bivariate Gaussian model. With multiple trajectories, each cell is scored by many ellipses and finally we can derive a *score map* measuring the likelihood of being walkable for all cells. Based on such a score map, we propose a two-phase clustering method to discover how the potential walkable areas are connected with nearby main roads or points of interest and then identify representative walkways for pedestrians. Finally, these new-found walkways are integrated into existing digital map for the public uses.

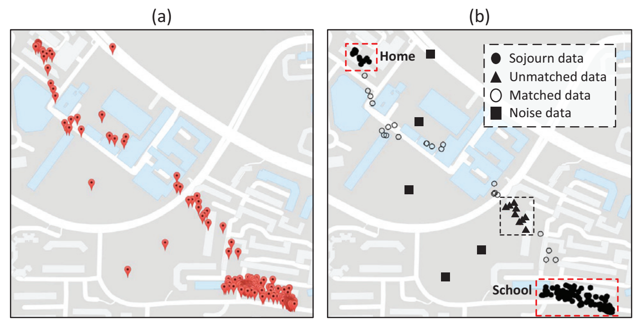

Figure 1: (a) The location data of one student on a specific school day. (b) The results of data classification based on the records shown in (a).

In this demo, we will bring one laptop to show the visualization of all new-found walkways in Singapore from the NSE dataset. During the demo on the conference site, we will show the viewers our results on a screen through our custom visualization platform. Therefore, we need the following facilities: a large electronic screen (around 25 $\sim$ 30 inch) to show the visualization page for viewers and a VGA cable to connect our laptop with the screen.

2 WALKWAY DISCOVERY FROM NSE DATA

From raw data to labeled data. We process NSE mobility data to derive clean inputs for our system. We first remove the data that are not in walking status and of large localization errors. Then we exploit HDBSCAN algorithm [1] to cluster and filter out the location points that correspond to sojourn places (e.g., home or school). Finally we classify the rest data into "Matched" and "Unmatched" through a map matching algorithm [4]. If location data can be well matched with some existing road, then it is considered as "Matched", otherwise "Unmatched". For example, Fig. 1(a) shows the raw mobility data of a student in one day and Fig. 1(b) is the results of labeling. These "Matched" and "Unmatched" data are fed to next module for walkway discovery.

From labeled data to score map. We use an ellipse to include all possible walking paths between two consecutive unmatched

0-7695-6377-5/18/$31.00 ©2018 IEEE
DOI 10.1109/IPSN.2018.00037

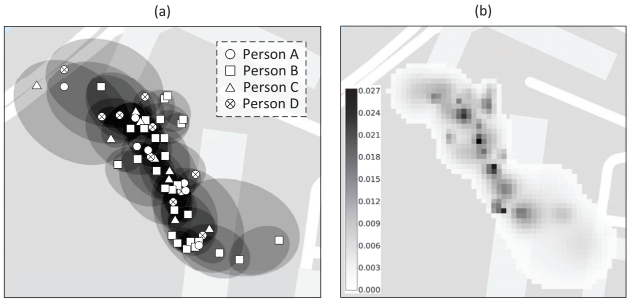

Figure 2: (a) An example of walkable area estimation based on four trajectories of person A, B, C and D. (b) Score map illustration of corresponding walkable areas in (a).

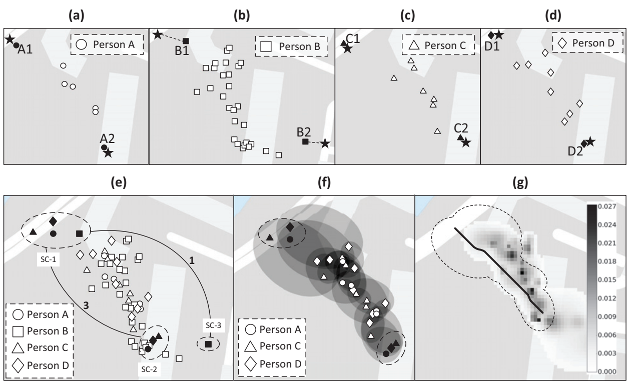

Figure 3: An illustrative example for representative walkway identification. (a-d) show the mobility data of person A, B, C, and D. (e) illustrates the two-phase clustering. (f) Walkable area by ellipses of trajectory cluster <SC-1, SC-2>. (g) Searching space and derived walkaway of trajectory cluster <SC-1, SC-2>.

locations. The shape and size of each ellipse are determined as follows: two locations are the *focal points* and *step-count*×*step-length* determines *sum of distances d* from every point on boundary of ellipse to the two focal points. Fig. 2(a) shows an example of using ellipses to estimate walkable area from multiple trajectories. Then we adopt the bivariate Gaussian model to assess the probability of being walkable for the divided cells given an ellipse and the model parameter can be determined by d. The details can be found in the paper. One cell might be scored by multiple ellipses and finally we derive a score map for the area of interest. Fig. 2(b) shows the final score map given unmatched trajectories in Fig. 2(a).

From score map to walkways. We propose a two-phase clustering method to identify walkways from a score map. The score map is transfered into a weighted graph where nodes with weight in the graph are the cells with score, and edges are generated when nodes are immediately neighboring cells in the score map. In the first phase, we cluster the last matched locations, that are the locations next or before matched locations in time-order, to determine the possible start/end position of a walkway. Fig. 3(e) illustrates three location clusters from four trajectories of four persons A, B, C and D shown in Fig. 3(a-d). In the second phase, we cluster the unmatched trajectories and identify walkways from trajectory clusters. We annotate each trajectory using the clusters of the pair of last matched locations, and group all trajectories with same annotations together. For each trajectory cluster, we formulate the walkway identification problem into a shortest path searching problem in a customized weighted graph that only keeps nodes and edges related with original trajectories. Fig. 3(f) depicts a trajectory cluster and Fig. 3(g) shows the identified walkway.

Figure 4: New-found walkways generated from the dataset in National Science Experiment of Singapore.

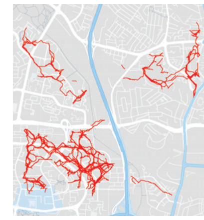

Figure 5: Zooming in at the black box in Figure 4.

Visualization. All new-found walkways are hosted in a database and we visualize them through our own visualization platform. Fig. 4 presents a screen shot of our visualization platform. Viewers can have zooming in and out operations to check the details of a walkway. For example, Fig. 5 is the result when zooming in the region highlighted by a black box in Fig. 4.

3 CONCLUSION

In this demo, we introduce the principle of our system for discovering walkways from NSE mobility data, including the data preprocessing, ellipse based walkable area estimation, bivariate Gaussian model based weighting, and the two-phase clustering based representative walkway identification.

ACKNOWLEDGMENTS

We acknowledge the support from Singapore MOE AcRF Tier 1 grant RG130/16, Tier 2 grant MOE2016-T2-2-023 and the COE grant from NTU, as well as the Natural Science Foundation of SZU under grant No.2018061.

REFERENCES

[1] 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* 10, 1 (2015), 5.
[2] Cao Chu, Liu Zhidan, Li Mo, Wang Wenqiang, and Qin Zheng. 2018. Walkway Discovery from Large Scale Crowdsensing. In *Proceedings of the 17th ACM/IEEE International Conference on Information Processing in Sensor Networks*.
[3] RI Jennrich and FB Turner. 1969. Measurement of non-circular home range. *Journal of Theoretical Biology* 22, 2 (1969), 227–237.
[4] 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.
[5] NSE.SG. 2017. National Science Experiment. (Aug. 2017). Retrieved June 12, 2017 from https://www.nse.sg
[6] NSE.SG. 2017. National Science Experiment. (Aug. 2017). Retrieved June 12, 2017 from http://nse.sg/sensg/about-sensg