AIN SHAMS UNIVERSITY
Faculty of Computer
& Information Sciences
Information Systems Department

Crowd-sourced Map Generation

Thesis submitted to the Department of Information Systems Faculty of Computer and Information Sciences Ain Shams University, Egypt
In fulfillment of the requirements for the Master of Science degree, MSc, of Computer and Information Sciences

By

Mahmoud Ezzat Abdelmobdy

Teaching Assistant, Information Systems Department,
Faculty of Computer and Information Sciences, Ain Shams University

Under the Supervision of

Prof. Dr. Mohammed Essam Khalifa

Professor, Basic Science Department,
Faculty of Computer and Information Sciences, Ain Shams University

Dr. Rania Abdelrahman Elgohary

Associate Professor, Information Systems Department,
Faculty of Computer and Information Sciences, Ain Shams University

Dr. Mahmoud Attia Sakr

Assistant Professor, Information Systems Department,
Faculty of Computer and Information Sciences, Ain Shams University

February 2019

Table of Contents

List of Figures

List of Tables

List of Abbreviations

GPS: Global Positioning System

OSM: OpenStreetMaps

SI: Satellite Imagery

HSR: High Sampling Rate

LSR: Low Sampling Rate

KDE: Kernel Density Estimation

ILM: Intersection Linking Method

ITI: Incremental Track Insertion

DTW: Dynamic Time Warping

DBSCAN: Density-Based Spatial Clustering of Applications with Noise

EPDBSCAN: Extended Progressive Density-Based Spatial Clustering of Applications with Noise

Abstract

The wide spreading of geo-aware mobile applications provides the opportunity for huge amounts of user-contributed GPS trajectories to be available with various levels of accuracy. Constructing road maps is one important benefit of such datasets. Now, many applications rely on road map data to provide their services. However, keeping the road map up-to-date and validating the correctness of generated maps pose interesting research challenges. Therefore, the research community has been recently interested in proposing alternative means of road map production utilizing the crowd-sourced GPS trajectories. There are some challenges related to the inaccuracies incurred on real datasets, such as missing GPS signals, low sampling rate and bad driving behaviour.

In this thesis, we present a clustering-based method to extract the road map from GPS tracks. Additionally, a new preprocessing algorithm is proposed to adapt with the common problems related to GPS data. Firstly, the tracks are simplified to extract road turns, as well as to remove the noise data. In order to handle the problems due to the low sampling rate, we adjust the core points of the simplified tracks by moving them closer to the positions of the real turns. Afterwards, a progressive clustering is applied to extract turns and intersections. Both of them are connected based on the trajectory information. Finally, we

proposed an efficient method to integrate the turn connections to derive the road segments. Our method is different from related work in that it deals with multiple issues related to real datasets, specifically noise, inconsistent and rather low sampling rate, and the difficulty of tuning parameters. We extract both intersections and turns, allowing applications to make better use of such GPS data.

We evaluate the accuracy of our results by comparing the proposed method with two of the best state-of-the-art methods using a small-scale dataset that was collected in Cairo under low sampling rate. Another experiment is conducted by extracting a part of the road segments of Egypt using a large-scale dataset with more than 12 million GPS points that are captured with high sampling rate by thousands of taxis all over Egypt. Experimental results show that our proposed method outperforms the other methods with regard to F-measure, especially with the low sampling rate datasets. Additionally, the results demonstrate that the proposed method can precisely extract the divided roads that have two lanes of traffic, travelling in each direction and is able to ignore the false trajectories that exist due to GPS errors. Finally, this thesis enriches the area of map production by proposing an efficient method that automates the extraction of the road map from low-sampling-rate GPS trajectories. The generated road map is considered a highly-precise routable-map that can be utilized for the purpose of navigation. Furthermore, our method can deal with the complex networks such as the network of Egypt.

List of Publications

1. M. Ezzat, M. Attia, R. Elgohary and M. E. Khalifa, "Extracting road turns and intersections from crowd source GPS tracks," In proceedings of *International Conference on Communication and Signal Processing (ICCSP)*, Chennai, 2017, pp. 0947-0951, IEEE, doi: 10.1109/ICCSP.2017.8286511.
   `https://ieeexplore.ieee.org/document/8286511`
   The paper was presented by Mahmoud Ezzat at the conference that was held from 6 - 8 April 2017 in Adhiparasakthi Engineering College, Chennai, India.

2. M. Ezzat, M. Attia, R. Elgohary and M. E. Khalifa, "Building Road Segments and Detecting Turns From GPS Tracks," In *Journal of Computational Science,* Volume 29, 2018, Pages 81-93, ELSEVIER, doi: 10.1016/j.jocs.2018.09.011, impact factor 1.925
   `https://www.sciencedirect.com/science/article/pii/S1877750318302813`

Acknowledgment

I would first like to thank my thesis advisors Professor Mohammed Essam Khalifa, Associate Professor Rania Abdelrahman Elgohary and Assistant Professor Mahmoud Attia Sakr who really supported and guided me greatly during my study and research. Their comments and suggestions helped me not only in this work, but also in my career. It was my pleasure and honor working with them.

I am also grateful to all my colleagues for their continuous support and help. Finally, i must express my very profound gratitude to my family and my friends for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Chapter 1.

INTRODUCTION

1.1. Background

The emergence of applications that base their functionality on road map data brings a great interest in automating the process of road map production. Modern applications such as location-based services, navigation, Geo-enabled social networks, etc, require an accurate and up-to-date road map. Such a map is typically modeled as a graph, where nodes represent the road intersections and edges represent the road segments. Additional attributes can be linked to nodes such as the intersection type, and to edges such as the direction, length and traffic density. Continuous updating of these networks is critical for these applications to ensure the safety of the vehicles and avoid the fatal accidents.

Traditionally, road map production was done using field surveying with specialized equipped cars, or by digitizing a road map from high resolution satellite imagery (SI). These methods are considerably expensive and highly-time consuming, which makes it difficult to keep the road map up-to-date. This is more clearly seen in developing countries and fast growing regions, where the construction rate is relatively high, and the structure and topology of the road map change frequently. By the fast advance of mobile devices and hand-held GPS units, road users, such as pedestrians, cyclists, motorists, vehicle passengers and users of public transportation contribute their GPS tracks to many applications. Now, crowd-sourced applications sit on a huge amount of user-contributed GPS trajectories. One of the major objectives for these applications is to extract an accurate and up-to-date road map from such datasets. Fig. 1.1 shows OpenStreetMaps (OSM) as a background for the black lines that represent thousands of user trajectories that are collected in Egypt. In this figure, the road map is visually clear with some noise data. In addition, GPS trajectories can be utilized to detect the road features such as number of lanes, road type, parking spaces and road surface conditions. For example, Yang et al. [1] presented a system that detects the parking spaces from large volumes of vehicle trajectories and identifies two types of parking, on-street parking and zone parking. Within their preprocessing step, the trajectories are map-matched onto a digital road map that needs to be accurate and up-to-date to ensure the accuracy of the parking locations.

Figure 1.1.: User-contributed GPS Trajectories

1.1.1. GPS Raw Trajectory

An input raw trajectory T is a finite set of GPS points P_0, P_1,....,P_n sorted by their time-stamp t_0, t_1,....,t_n. These samples represent a continuous curve. Usually, the measurements of the samples are only accurate within certain bounds. The measurement error depends greatly on the quality of the GPS receiver, and the accuracy of the movement transition between the consecutive samples varies depending on the sampling rate. We consider every individual trajectory a set of piece-wise linear curves. Fig. 1.2 shows an example of an input raw trajectory with its start and end points.

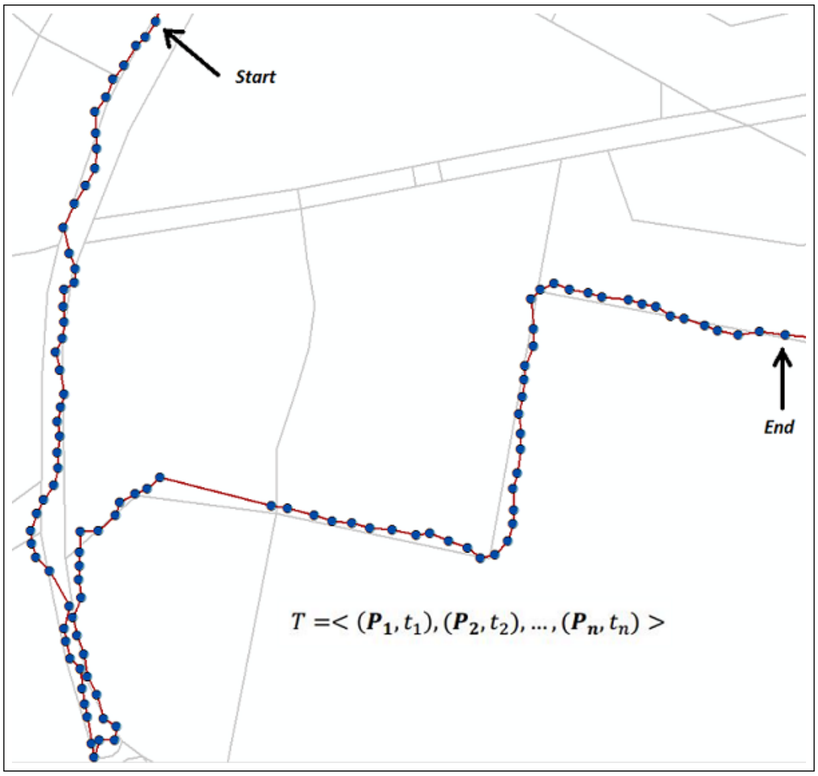

Figure 1.2.: Example for an input raw trajectory