Weekly report: Nov 24-Dec 5, 2014

Ivan Rios S.

December 4, 2014

1 Goals for the week

- Research previous work about trajectory identification and determine possible applications to this project.
- Write documentation for scripts, and database to make them public in the ISS4E web page.
- Expand database to store additional data and information about trips.
- Modify the scripts to plot available information in the database to obtain more detailed graphs.
- Add new features to the Webike page and have it fully functional for all users to access it.

2 Activities

2.1 Survey about trip detection.

Paper: Chung, Eh, and A. Shalaby. "A Trip Reconstruction Tool for GPS-Based Personal Travel Surveys." Transportation Planning and Technology 28.5 (2005): 381-401. Print.

Summary: This article describes a software tool developed to process GPS data from a personal GPS tracker (wearable) and identify routes and modes of transportation.

Important Points:

- The system focuses on two main objectives. First, it obtains road links traveled which is the identification of the roads used by the user (based on GIS information). Second, it determines the modes of transportation used based on specific parameters that describe each mode of transportation.
- In the preprocessing stage of GPS data, the authors use the number of satellites available in order to filter valid points. They mention that the GPS antenna should receive signals from at least 3 different satellites simultaneously to obtain an accurate estimation of the position.

- Horizontal Dilution of Precision (HDOP) is used to calculate the quality of GPS data. This is an index that describes how satellites are arranged in the sky at the time of the record. Lower angles between satellites cause lower precision.
- They use Map Matching [1] to determine what route (streets and roads) was followed by the user. This is potentially useful as the walking and biking routes are implemented in mapping systems; in this case, they assumed the users followed regular streets and roads as part of their trajectories.
- Trips are identified using parameters such as distance, and time between subsequent points. If this values are within a certain range, they identify potential trips.
- They identify modes by using some characterization of each (bicycle, bus, car, walk) and additional features such as directionality of roads, transit network, waiting behavior, moving patterns (eg. speed and distance), transit vehicle stop pattern, etc.

Applicable ideas:

- Characterization of electric bicycles as a mode of transportation.
- Available data could be used to identify "biking/waking only" routes in the city of Waterloo.

Paper: Gong, H. M., et al. "A GPS/ GIS Method for Travel Mode Detection in New York City." Computers Environment And Urban Systems 36.2 (2012): 131-9. Print.

Summary: This paper describes an algorithm that is able to reach ˜80% of accuracy to manage GPS data in complex urban environments such as New York City.

Important Points:

- Their algorithm is able to identify any kind of combination of transportation modes. They mention that generally other algorithms used standard combinations for their analysis (eg. bike, bus, bike).
- Since this is a merely GPS study, the software sleeps while the devices is not moving.
- Again, number of satellites available and HDOP are used as a primary tool to filter non-accurate data out.
- Trips start and end when a cluster of points stays inside a radius of 50 meters for over 200 seconds.
- Modes are detected using parameters such as speed, average speed, distance.

- They estimate the warm start time of a GPS between 39s to 106 s. This is the time that the GPS takes to start calculating its location after being in places with no satellites access such as the subway.
- They mention the complexity of identifying modes when there is heavy traffic jams.

Applicable ideas:

- We could reduce the amount of GPS data considerable by not recording location when the phone is not moving. It is possible to create an algorithm that guarantees that the phone is not moving using parameters such as acceleration or GPS data itself.
- We could estimate the warm start time of the phone every time it restarts.
- Analyze the change of moving patterns in rush hour times to see how much it affects bicycles. Also, we could include the types of roads (on road, off road, trail, etc) and other features to see how bike users behave in these situations.

Paper: Kasemsuppakorn, P., and Ha Karimi. "A Pedestrian Network Construction Algorithm Based on Multiple GPS Traces." Transportation Research Part C-Emerging Technologies; Transp.Res.Pt.C-Emerg.Technol. 26 (2013): 285-300. Print.

Summary: This paper describes an algorithm used to process GPS data in order to identify walking routes and implement them in the map. This data is used to merge routes and create a pedestrian network in areas where this data is not available.

Important points:

- The authors make a distinction between 7 different types of path segments: sidewalk, cross- walk, pedestrian walkway or footpath, accessible entrance, pedestrian bridge, pedestrian tunnel, and trail.
- Again, number of satellites available and HDOP is used as a primary tool to filter non-accurate data out.
- The authors use the bearing change in order to identify significant points. This technique calculates the angle variation between subsequent points, and the destination of the trip, and then it filters out the angles under a certain threshold.
- After filtering the significant points of a trajectory, they apply a clustering algorithm called Partitioning Around Medoids (PAM) to isolate outliers and clean the data.

Applicable Ideas:

- Identify turns using bearing change in the different bike trips.
- Combine the current GPS simplification algorithm with a bearing change analysis.

Paper: Schuessler, Nadine, and Kay Axhausen. "Processing Raw Data from Global Positioning Systems without Additional Information." Transportation Research Record.2105 (2009): 28-36. Print.

Summary: This paper describes an algorithm that only requires raw GPS data as input and is able to identify trips and modes of transportation. The study was done using data from ˜5000 participants for an average of 6.65 days.

Important points:

- For the data cleaning stage, they use a variance of HDOP which is called PDOP (Position DOP) that considers the position in a 3-d space. Additionally, they consider the number of satellites available as a filtering tool. Finally, they also use altitude to estimate possible miscalculations of the GPS receiver (the paper mentions that errors in GPS data are generally accompanied by strong jumps in altitude).
- Trips are detected by analyzing the data that contains speed close to 0 and stay within a circle with diameter of 30 meters. This value is equivalent to 3 times the standard deviation of the measurement accuracy.
- They mention that most of the trips in literature are detected by detecting periods of inactivity (without considerable movement) between 45 and 300 seconds. Generally, the standard is to use 120 seconds.
- For mode detection purposes, the authors recommend using fuzzy logic because of the overlapping features of the different modes.

Applicable Ideas:

- Include accuracy as input the trip detection algorithm

OTHER PAPERS:

Paper: Usyukov, Vladimir. "Development of a Cyclists' Route-Choice Model : An Ontario Case Study." Waterloo, Ontario, Canada : University of Waterloo, 2013. Print.Waterloo, Ontario, Canada: .

Summary: With 2000 routes obtained from 415 cyclists, the author developed a model to determine Route-Choice based on the level of experience of the biker. He also identifies the main factors that influence in the route-decision with a precision 65%.

Important points:

- The key factors impacting route-choice were found to be trip length, speed, volume, bicycle lane presence and percent of uphill gradient that cyclists face.

Paper: Wolf, J., et al. "Eighty Weeks of Global Positioning System Traces - Approaches to Enriching Trip Information." Data And Information Technology.1870 (2004): 46-54. Print.

Summary: This paper presents a study of the GPS data from 186 personal cars for at least 30 days (approximately 250,000 trips). The authors developed algorithms to correct this data, identify destinations and determine purpose of trips.

Important points:

- For trip detection purposes, the authors considered thresholds such as a minimum trip length of 30 seconds, maximum distance, average speed, and activity time between trips of at least 300 seconds.
- The analysis is divided between full-time workers and retired users.

I also found a paper (I do not think it was published) that characterizes the behavior of a bike user based on demographic data (age, gender, frequency of biking, etc). I think we could try to identify some general patterns. Also, this paper considers other information such as the type of road, traffic, among other, to estimate the speed of the bicycle.

2.2 Write database dictionary.

This document is 70% ready. I will be uploading the last version once we determine the feedback data we will be asking the users.

2.3 Modify the database to save information about trips and their corresponding GPS points. Additionally, write script to upload new GPS data into database periodically.

I wrote the script ebikeblizzardGPS.py which is located in the sensordc home directory. This script is run along with Tommy's ; both scripts are called in the file transfer_files.sh that is executed every hour. This script converts the time stamps from UTC to local time and uploads all the new GPS information to the tables specifically created for this purpose.

2.4 Write manual about how to add new modules to the Webike web page.

This document was emailed on Tuesday and it will be uploaded to the public on the ISSE4 web page.

2.5 Other activities:

- Export part of the database in individual files (available in sensordc/bikeData)
- Add features to script that plots data available in database (emailed on Monday).
- Write new script to show the percentage of voltage data available in the database (emailed on Monday).
- Add changes and modify the Webike web page with new algorithms and features.

3 Questions:

- According to the meeting we had before, these are the options to get feedback from the user:
  - Rate the accuracy of every trip: inaccurate, mostly accurate, accurate
  - Show a naive approach and our approach and ask the user to choose the most accurate one.
  - Show unprocessed data and our approach and ask the user to choose the most accurate one.
- The user can define potential problems of the identified trips such as:
  - Missing data.
  - Too much gap between subsequent points.
  - Trip points are incorrect.
  - Missing sections.
  - Beginning or end of points are incorrect.
- Determine the way to ask for feedback: Should it appear every time we present a trip on Google Maps or should it be presented as a new module in the web page where they can enter optionally?

4 Proposed goals for next week

- Include acceleration to identify trips effectively.
- Expand the survey using other papers.
- Apply trip detection algorithms on new GPS data.

- Modify the scripts that plot the available data so that they can be generated automatically.
- Write a script to automatically process new incoming data and include it in the trip information tables.
- Document GPS identification script.
- Upload final version of documentation to the ISS4E web page.

5 References

- Greenfeld, J. S. (2002) Matching GPS observations to locations on a digital map. Papers presented at the 81th Annual Meeting of the Transportation Research Board, Washington, DC, January.