Weekly report: October 31-November 9, 2014

Ivan Rios S.

November 11, 2014

1 Goals for the week

- Plot available data over time.
- Implement an algorithm for Trajectory Simplification of GPS data.
- Determine the optimal parameters for the Trajectory Simplification algorithm.

2 Activities

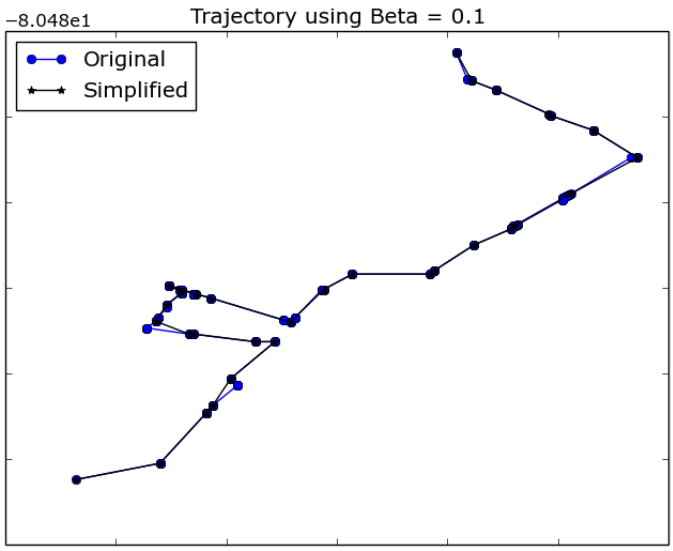
- Set up local copy of the database for development.
- Obtain longest trajectories from database for analysis.
- Modify functions developed by Tommy in order to use them for GPS data cleaning algorithm.
- Implement algorithm using persistence based trajectory simplification [1].
- Analyze the efficiency of the algorithm using various values for the beta parameter. The analysis was done on two trajectories and using 7 different beta values to determine the optimal one that allows to reduce the number of points required while keeping the fidelity of the trajectory image.
- Plot available data based on information stored in the database. The data was analyzed using three criteria. The first one was by checking that there is at least one point of voltage and/or GPS recorded a day; the second one was by analyzing days with at least 5 points available of data; and the last one was by determining if it was possible to identify at least one trip per day using the GPS data.

3 What I learnt

The following table presents a summary of the results obtained from the analysis. The table contains the values used for the parameter Beta, the number of points remaining after applying the algorithm, the percentage of remaining points after applying the algorithm, and the fidelity of the trajectory obtained after applying the algorithm.

Beta	Persistent Points	Percentage	Fidelity	Initial Points
0.05	47	21%	High	225
0.10	38	17%	High	
0.15	32	14%	High	
0.20	32	14%	Medium	
0.25	24	11%	Low	
0.30	22	10%	Low	

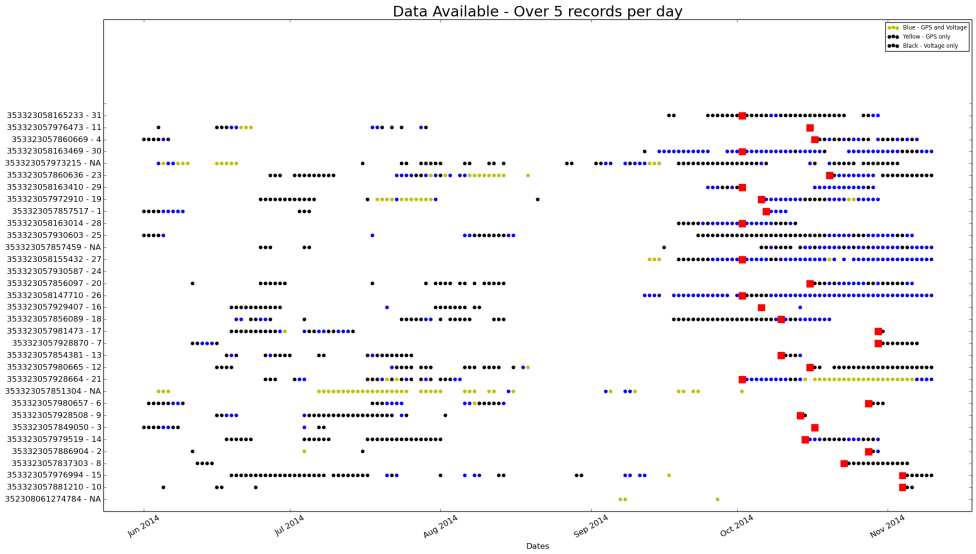
Similar values were obtained for the second trajectory analyzed. It is possible to see that by using a beta value between 0.05 to 0.15 will generate an accurate trajectory very similar to the original one while keeping only 14-21% of the original points. The following graph presents an example of the trajectory obtained after applying the algorithm:



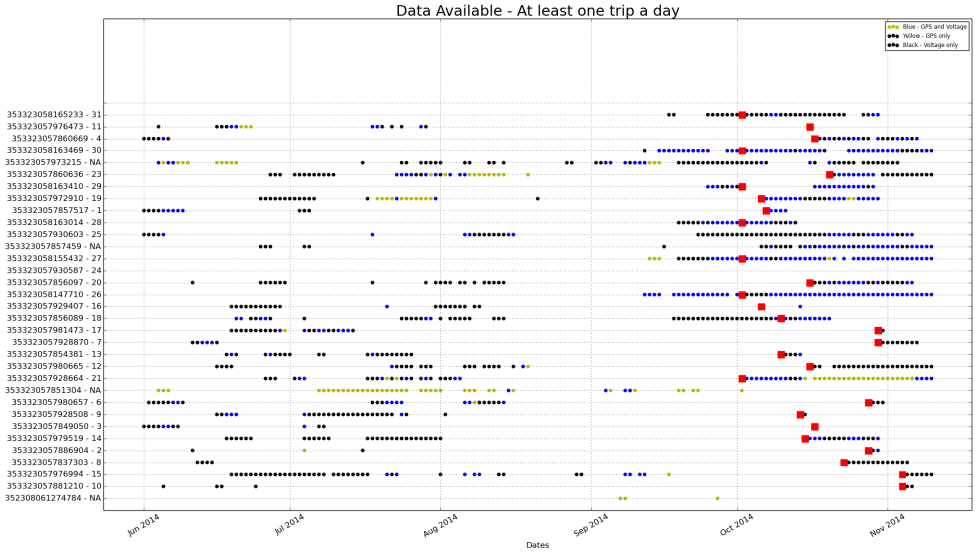
As it is possible to see in the graph, the simplified trajectory is very similar to the original even though there has been a considerable reduction of the number of points. Also, it is important to consider that a high percentage of the points register no changes in the trajectory (i.e. latitude and longitude are the same for subsequent records) which is the reason why there does not seem to be the difference of 84% of the points between the simplified and the original trajectory.

In conclusion, the algorithm effectively reduces the amount of points required to represent a trajectory which is also an advantage to clear the noise generally present in GPS data. This can be later used for trip detection purposes.

Additionally, the following plot represents the available data in the database using the criteria that considers that a day should have at least 5 records of GPS and/or Voltage:



Also, a similar plot was obtained by considering a day to have GPS data only if it is possible to identify at least one trip during that day. The results are presented in the following figure:



Both of these plots present the available data for each one of the IMEI since the project started. The first plot can be used to assure that the system is collecting information effectively, while the second one can be used to keep track of people who is actually using the bikes regularly. Also, each one of the plots contains the dates (red squares) when the caps of the sensor box were changed in order to fix the problem of GPS data collection.

4 Proposed goals for next week

- Use the processed GPS data as the input to do trip detection.
- Find an implement an efficient algorithm for trip detection.

5 References

1. Panagiota Katsikouli, Rik Sarkar, Jie Gao. Persistence Based Online Signal and Trajectory Simplification for Mobile Devices. To appear in the 22nd International Conference on Advances in Geographic Information Systems (SIGSPATIAL), ACM, November 2014, Dallas, USA.