CMSC396H Project Final Report

Abdul A. Hadi
ahadi@terpmail.umd.edu

Isaac Roy
iroy@terpmail.umd.edu

Arnav Whig
awhig@terpmail.umd.edu

Shaashvat Shetty
sshetty@terpmail.umd.edu

1 Abstract

Our research evaluates the performance of various map matching algorithms. We evaluated four map matching algorithms: Euclidean Distance Matching, Buffer Zone Matching, Slope Matching, and Hidden Markov Model-based Matching. Our methodology involved generating routes with data points that had noise added to them synthetically. The selected algorithms were then run and the accuracy of the predicted points was assessed by calculating their deviation from the ground truth points by distance. Our results show that the Hidden Markov Model-based Matching algorithm exhibits much lower errors compared to all the three other algorithms.

2 Description & Summary of the Problem

Map-matching is a concept where a set of trajectories (Essentially vectors containing coordinates and direction) are inputted from GPS is "matched" to some curve in $\mathbb{R}^2$ representing a physical feature such as a road. There are many different algorithms trying to accomplish this feat considering both accuracy and efficiency, ranging from probabilistic to geometric and mathematical approximations. The goal here is to determine which algorithm produces an optimal result. This is important, as many models today are not perfect and can match with incorrect roads and can be affected by noise and sparseness in data. In addition, we are interested in investigating if different kinds of road networks affect an algorithm's performance and if some are more effective in certain types. For example, dense, urban, irregular European road networks are very different from regularly shaped roads that you find in cities like Milton Keynes, UK, or many American cities. Plus, perhaps it is a smart idea to use a normally less effective algorithm in a rural area with fewer roads. Ultimately, benchmarking these algorithms across varied road types will pinpoint the best map-matching approach for each setting, improving GPS-based trajectory accuracy in real-world use.

3 Prior Work

Prior work in the topic of map matching has uncovered many different approaches to solving the problem. The first solutions relied on geometric methods [2] like spatial distance and the shape of trajectories. While these methods worked reasonably well in ideal conditions (consistent sampling, relatively clean data), they struggle when data is sparse and noisy or perhaps when the underlying road network is complex. More modern research into this topic has involved statistical solutions such as the use of Hidden Markov Models (HMMs) [1]. These have shown improvements in accuracy and are commonly used in real-world applications.

4 Methodology

Our methodology's initial step involved selecting two distinct points, A and B. Then, a known, optimal path connecting these two points was identified. To generate the ground truth dataset, we sampled a series of points along this pre-defined path. This collection of sampled points represents the trajectory against which the outputs of the map matching algorithms would be compared to. We then synthetically add noise to each and every point so that they deviate away from the true path so that we can simulate the general inaccuracy of GPS before inputting them into each of our four algorithms. The final phase involved a quantitative assessment of each algorithm. For every predicted point generated by an algorithm, we calculated the distance between it and its corresponding ground truth point. This distance served as a direct measure of the error. This allowed us to calculate some statistics like mean error and standard deviation so that we could perform some analysis.

5 Results

As shared in our presentation, we worked on two sets of ground truth. The first was in a dense network (University of Maryland's campus) while the second was in an area with few paths (Arlington National Cemetery). On both of them, the Hidden Markov Model-based matching algorithm demonstrated superior accuracy with a mean error of 5.42 and 0.69 meters respectively. Likewise, the standard deviation was also lower than that of the other algorithms. Additionally, the median was 0 meters which means that over half of the points predicted by this algorithm were spot on. All the other algorithms, which took on a geometric approach to solving the problem, had mean errors of around 9 and 2 meters respectively on the two data sets. The two tables at the end of this section show the results on our two data sets.

Our results generally align with the theoretical understanding of each algorithm's strengths and weaknesses. Euclidean Distance Matching, while computationally inexpensive, is known to struggle in dense urban areas, and its relatively higher mean and maximum errors support this limitation. One interesting thing to note is that on our second data set, the slope method had a relatively low standard deviation (5.06 meters) despite the higher mean error (2.34 meters) compared to the other two geometric algorithms. Likewise, the same can be seen in the first data set in favor of the Euclidean distance matching method.

Overall, we can conclude that the HMM-based algorithm is best. For users in an enterprise setting, the increased computation required for it likely is overshadowed by its superior accuracy. The other approaches are more suited to road networks that are not complex and dense like in major cities.

Statistic (metres)	**Euclidean**	**Buffer zone**	**Slope/gradient**	**HMM**
Mean error	9.19	8.20	9.29	5.42
Median	6.61	0	0	0
Standard deviation	10.20	11.57	12.12	9.62
Max	32.42	49.77	49.77	34.33

Statistic (metres)	**Euclidean**	**Buffer zone**	**Slope/gradient**	**HMM**
Mean error	2.16	2.14	2.34	0.69
Median	0	0	0	0
Standard deviation	6	5.07	5.06	2.21
Max	28.07	28.07	28.07	8.97

6 Future Work

One of the goals for the future that was not possible in the time slot was to rigorously test our algorithms across a large scale of maps. Right now, only two sample maps per algorithm were used, and although it did leave some interesting insight, stronger conclusions could be created for different kinds of map, especially with varied road density, for example, in wilderness, rural areas, and in cities. Finally, we can use a graph theory approach, where we can treat road (edge) weights as speed limits. When considering trajectories with velocity, such an approach can help refine what specific street a person is driving on.

References

[1] Paul Newsom and John Krumm. Hidden markov map matching through noise and sparseness. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2009. https://doi.org/10.1145/1653771.1653818.

[2] Lianxia Xi, Quan Liu, Minghua Li, and Zhong Liu. Map matching algorithm and its application. 10 2007.