UC Berkeley
Earlier Faculty Research

Title

Catchline International Journal of Geographical Information System, Vol. X, No. X, Month 2006, xxx–xxx Running heads (verso) J. Zhou and R. Golledge (recto) A three-step map matching method in the GIS environment A Three-step General Map Matching Met...

Permalink

https://escholarship.org/uc/item/5g32k3h3

Authors

Zhou, Jianyu (Jack)
Golledge, R G

Publication Date

2006-10-01

A Three-step General Map Matching Method in the GIS Environment: Travel/Transportation Study Perspective

JIANYU (JACK) ZHOU*† AND REGINALD GOLLEDGE‡

†Department of Geography
University of California Santa Barbara,
Santa Barbara, CA 93106-4060, USA
Telephone: (805) 971-5033
Fax: (805) 456-0362
Email: zhou@geog.ucsb.edu

‡Department of Geography
University of California Santa Barbara,
Santa Barbara, CA 93106-4060, USA
Telephone: (805) 893-2731
Fax: (805) 893-7782
Email: golledge@geog.ucsb.edu

Abstract. Despite its application in many fields, map matching in studies of travel/transport geography is unique in two aspects: 1) The correct road links traversed by the traveler need to be unambiguously identified; 2) All the identified links should form a meaningful travel route. This paper discusses the application of map matching methodologies in the context of deriving people's travel behavior from GPS-traced multi-modal trip data. In recognition of the disadvantages associated with the existing algorithms, this research proposed and implemented a heterogeneous map matching approach suitable for travel/activity research needs which is uniquely characterized by: 1) data preprocessing with point cluster reduction and density leverage; 2) offering the candidate solution within a pool of "the best"; 3) the balancing of matching results from multiple matching factors with rank aggregation; 4) Utilizing the network constraint attributes to increase the matching accuracy; and 5) Use of the Dempster belief test to discern the noise and off-road travel.

1 Introduction

Map matching is defined as the process of correlating two sets of geographical positional information (e.g., GPS records of object positioning versus digital road networks from different vendors). Data types handled by map matching include point-to-line, line-to-line (Xiong, 2000) and polyline to polyline matching. Based on the temporal-response characteristics, map matching algorithms can also be roughly classified into online map-matching and off-line map-matching. Online map-matching methods snap the device-captured geo-spatial feature position to the base reference in real-time. Offline map-matching counterparts post-snap the point data/linear data after the whole set of data is collected (Yin and Ouri, 2004).

This paper discusses the map matching methods in the context of travel/transportation studies. Map matching is used as a means to transfer the road network attributes to the resulting travel route in order to derive certain travel behavior, and, hence, further analysis can be conducted based on the inferred information. Matching GPS recorded points to the correct position on the correct road link is secondary compared to obtaining the topologically correct travel route and the associated attributes/statistics. The paper is organized as follows: Section 2 briefly overviews the currently available map matching methods. Section 3 delves into the unique requirements for map matching algorithms in travel/transportation studies and describes an innovative three-step heterogeneous map matching methodology. Finally, section 4 presents conclusions and discusses future research on the proposed map-matching method and its application in travel/transportation studies.

2 Overview of Map Matching Methods

2.1 Applications and Conceptual Formulation

Map matching has widespread applications, including automatic vehicle navigation (Syed & Cannon, 2004), image processing, network data conflation (Xiong, 2002), and travel/activity surveys (Lexington Travel Survey, 1997). The common procedure for map matching is to establish the correspondence relationship between two sets of spatial features. Spatial information and attributes contained in the different spatial data sets are conflated into one of the spatial data sets via a matching operation, or are processed further to generate a new data set. The process of map matching helps researchers extract the characteristics of matched features from one data set and transfer/update the associated spatial/non-spatial information to/on the other data set. In this paper, we will focus on the research question of how to accurately match GPS-captured positional data onto linear data (road network).

2.2 Evolution of Map Matching Factors

A straightforward solution to the map-matching problem, as mentioned by White and his colleagues (2000), is to snap the GPS recorded location to the nearest road link node or road link. The simple solution, however, only generates perfect matching results in the ideal situation. Typically, due to the dual uncertainty and inaccuracy involved in both the point and network data, matching GPS recorded point data to road link nodes or to a road link itself is extremely prone to errors. The situation is exacerbated when only distance measures are used for guiding the matching process due to the potential map-matching zone overlay (Lakakis et al. 2004). Point to point matching (matching GPS points to road link nodes) could easily fall into the pitfall of matching to the wrong node on the wrong link if the correct and wrong links are close by in parallel and the correct link does not contain as many pseudo-nodes as the wrong one. Point to curve matching, (matching GPS points to road link arcs), on the other hand, suffers from ambiguity issues when the GPS point is close to a road intersection (Bernstein and Kornhauser, 1996, 1998; White et al. 2000).

Notice that a distance measure would only constrain a spatial feature along one dimension. An additional measure - travel direction - naturally became the candidate to add on as the second dimension to quantify the relationship between GPS points and the matching road network. Unsurprisingly it ends up improving the matching accuracy dramatically (White et al. 2000). However, travel direction derived simply from GPS points could be very unreliable, especially when the carrier of the GPS receiver travels at a low speed or makes brief stops along the travel routes. In either case, GPS signal drifts exert unpredictable effects on the GPS-derived travel directions. This has forced many researchers to rely on gyroscope or digital compass as the second data source. These devices generally can provide accurate heading and heading change information at various travel speeds (Quddus et al., 2003; Syed and Cannon, 2004).

As more research efforts are devoted to searching better map-matching methods, other measurements have been incorporated to provide better selection criteria for deciding the best match from the neighboring candidate link set. The decision space for matched road link selection, hence, was expanded to a multi-

dimensional level. For example, besides proximity and heading difference, Quddus et al. (2003) have used two additional measures for road link selection -- "GPS position relative to the road link" and "intersection relation between the GPS trace and the road links"; Syed and Cannon (2004) have used the "average distance traveled on current link" and "large distance traveled on current road link". In addition, if the base road network contains detailed road attributes (speed limits, one way lanes, etc.) that potentially restrict a certain routing behavior, they can be utilized to further filter unqualified road links (Najjar and Bonnifait, 2003; Taylor et al. 2001).

Intuitively, taking more factors into consideration helps avoid matching errors that easily result from measurements from a single perspective. Nevertheless, different select criteria could also result in conflicting matching conclusions and therefore cause more confusion. To overcome the difficulty, two approaches have been taken to fuse the multi-dimensional selection criteria and reduce them to facilitate the deterministic matching decision. The one that is commonly used is to simply combine the selection factor with a weighting scheme. The weighting factors are typically derived empirically from data testing (Quddus et al. 2003) or from adaptive-fuzzy-network-based training (Kim and Kim, 2001). The second approach is complicated enough to use Bayesian Belief Theory and Dempster-Shafter's rule for deriving the unique non-ambiguous selection.

2.3 Improvement with Topological Information

With meticulous calibration of selection criteria, map-matching algorithms are sufficient to identify a series of the matched road segments from the pool of candidate links. However, this does not necessarily imply the matching result would be meaningful in terms of truthfully reflecting a traced travel route. A matching result could show up simply as a group of disconnected "paths".

Notice that both point-to-point and point-to-curve matching approaches do not reflect the fact that GPS records indeed represent the travel routes of the carriers but that they only constitute a small sample of the route. Bernstein and Kornhauser (1996) and White et al. (2000) have suggested connecting the GPS points in sequence to form piece-wise linear curves, which are further matched against road network. The method is called curve-to-curve matching. The matched curve, consisting of road links that are selected from the base road network, should have the smallest L2 norm distance to the GPS trace or be viewed as the approximate variation of GPS trace after a small amount of translation and rotation. Techniques have been borrowed from the pattern recognition field that utilizes similar measures to evaluate the shape proximity between two geometrical figures. One of the good examples is Joshi (2001)'s application of a rotational variation metric to measure the shape similarity between vehicle trajectory and the possible travel paths.

On the other hand, there are many other researchers who chose to further improve point-to-point matching or point-to-curve matching. As suggested by Bernstein and Kornhauser (1996, 1998), tested by White et al. (2000), and included in other researchers' work (Quddus et al., 2003; Syed and Cannon, 2004), road network topology information has been incorporated into the matching algorithm to maintain the topological integrity of map matching results and prevent the error of matching points to the wrong road link. The underpinning rational is simple: if, at time t, a road link is selected without ambiguity; at time t+1, the selected road link would most probably remain as the ideal matching candidate; or the end of the current road link has been reached, and, hence, a road link that is connected to it becomes the next candidate. Therefore, the topology

relations (especially, connectivity) among road links restrain the search for the next matching candidate. Unreachable road links from the current match in one GPS epoch thus would be easily eliminated from the candidate link set with confidence. However, the effectiveness of the approach depends greatly on the extent to which we can trust the previous match. A tiny bad match could consequentially lead to a matching blunder as the GPS epoch progresses, as commonly seen in the pure geometry- based matching algorithms (Quddus et al., 2003). It is not an easy task to clearly dichotomize the confidence/trust domain with a deterministic threshold setting.

As a solution to the problem, Pyo et al. (2001) and, later, Marchal et al. (2004) increased the number of testing epochs for determining the best "continuing" road link from the currently matched one. Their algorithms keep multiple possible "continuing" road links and the corresponding accumulated "fitness" values in a hypothesis space over a period of GPS epochs. The fitness measurements that grew smaller than a threshold over time will get pruned. Finally, a road link hypothesis is confirmed as the map matching result either if its fitness score is the highest or the ratio of its score to the next highest one exceeds a certain threshold. Note that there is no single definition for such fitness measures. Marchal et al. (2004) used the aggregated proximity measure between the GPS points and road links, while Pyo et al. (2001) used the recursive conditional probability formulation that was comprised of multiple measurements -- projected position, link direction, link connectivity, road facility, etc. Both approaches seem to perform adequately in empirical tests.

2.4 Statistical-Estimate-Based Map Matching

Distinctive from geometry and topology-based map-matching methods, current trends in map matching development have begun to incorporate more probabilistic and fuzzy elements. These constituents are more tolerant of the uncertainty, partial truth, and approximation involved in map matching processes. Taylor and Blewitt (2000) took a unique but innovative approach by simulating the working mechanism of differential GPS. Their algorithm, called road reduction filtering, defines a pool of virtual GPS "Ref" positions by projecting the "Raw" GPS point to the nearby road links. These "Ref" positions are used as virtual differential corrections for the next "Raw" GPS points and, in turn, to generate another pool of "Ref" positions. Conceivably, the correct matching position on the road link must be among the "Ref" positions. Based on the fact that bearing and distance measures between successive "Raw" GPS points and those between successive "Ref" positions are highly correlated, any false series of "Ref" positions can be filtered out. In essence, the road reduction filter is still mostly geometry/shape based. But, among the few first attempts, statistical rationale (i.e., correlation measurement) for the first time was introduced into the efforts of searching for a better map matching solution. In the mean time, Lakakis (2000) attempted the approach using linear regression analysis to fit GPS points to the road centerlines. Parameter significance tests serve to verify the linear relationship between GPS latitude and longitude coordinates. The fitted lines work as predictors for map matching GPS points to the base road network. This method, however, puts higher demands on GPS data accuracy. For stand-alone GPS data, there is a great possibility that map matching zones for two one-way parallel roads overlay each other. Walter and Fritsch (1999) viewed the map matching problem as equivalent to transmitting information through a communication channel. The ultimate aim of the map matching procedure, therefore, becomes minimizing the amount of information loss during the transmission or maximizing the mutual information shared between data sets. The method does not require any tuning

parameters except a statistical investigation on the evaluation of the conditional probability involved in the mutual information formulation.

2.5 Data Enhancement Efforts

Apart from continuously refining the map matching procedures and techniques, other efforts have been made to improve map matching accuracy from the perspective of data preparation and compilation, either by increasing the accuracy of the localization estimates or by matching these estimates against a high-accuracy digital map. For an application that involves a larger and more stable carrying platform, data input for map matching typically incorporates simple GPS recorded point data with other data sources, such as from a digital elevation model, digital compass, gyroscope, velocity sensors or Antilock Brake System (ABS), etc. For other applications that allow post-processing in travel or transport-related studies, usually only the GPS point data serves as the input data source. While GPS recorded points comprise the single data source, a Kalman filter typically is executed to estimate the bias associated with a previous map-matching epoch. Then the estimation can be used to compensate the next GPS positioning input. In the situation where a GPS receiver is complemented and integrated with other data sources, a centralized Kalman filter (Extended Kalman Filter-EKF) has been used to incorporate the measurements from all data sources and generate a single stream of complex position estimates (Bétaille and Bonnifait, 2000). To further eliminate the errors and biases generated from the sensors, these EKF methods usually also build the geometrical shape of the platform into the data fusion model. The seminal localization approach has the ability to collect solid measurements even when the GPS signals are cut off by the obstruction of the surrounding environment. The multi-data source combination method is superior in that it handles both of the disadvantages associated with the individual Dead Reckoning (DR) or GPS localization methods by allowing them to complement each other within an integrative measurement framework (Bonnifait et al., 2001).

Compared to the rapid advances of positioning technology, improving map accuracy is more of a long-term, energy-consuming task. Digital maps usually either contain errors or do not possess enough resolution power for some of the map matching applications. Although the current commercial digital maps already have sufficient coverage over the road networks in most metropolitan areas and even most parts of rural regions, mapping accuracy typically varies dramatically, and the map details have not been able to extend to the details of lanes and cover particular types of roads such as bike paths or sidewalks. The map resolution problem could be partially resolved by one of two ways. For simplified two-lane road representation (generalized as single road centerlines), Marchal et al. (2004) have replaced them with two oriented links derived by adding small shifts perpendicular to the link centerline. Therefore, the distance between GPS record and the road segment can be enhanced as the distance from the point to the augmented road links. The method reverses the map generalization process to generate more road link details. But it does require that the network attribute table contain the "number of lanes" details. Rogers (2000) attempted to address the problem by using the repetitive differential GPS (DGPS) measurements from a probe vehicle to augment commercial digital maps down to the drive lane levels. His approach is partially successful, but still suffers higher errors around road intersection and in cases where GPS signals are subject to multi-path effects.

3 A Heterogeneous Map Matching Approach for Travel/Transport Studies

Different from its application in other fields, the map matching in travel/transportation studies aims at: 1) identifying the correct road links traversed by the traveler; and 2) ensuring that the identified links form a meaningful travel route. Its main focus is to correctly transfer the road network attributes data to the GPS recorded travel route, hence further identifying how the spatial features (road, locations visited, function area traversed, etc.) interacted with the traveler during the behavioral process. Ideally, map-matching methodology should be able to help answer queries beyond the direct matching result, i.e., road type distribution along the travel route or delays encountered. Generally speaking, it is impossible to answer these queries if the travelers are off-road, or if the travel roads are not shown on the map. Thus, map matching in travel/transportation studies not only calls for accurate road network maps, but also takes into account the fact that the travel is not necessarily restrained to roads that facilitate a vehicle's travel and that the travel might include pedestrian walks and bicycle journeys.

In this research, three recent map-matching algorithms were empirically evaluated for travel route derivation from GPS point data and their performance is tested against the GPS data that were collected in a travel survey. The three map-matching algorithms are: weight-based map matching by Yin and Wolfson (2004), fuzzy-logic based map matching by Syed and Cannon(2004) and General map matching by Quddus et al (2003). They were implemented in the ARCVIEW network analysis module and the evaluation is conducted from multiple perspectives – data needs, selection factors, matching accuracy, and time complexity.

The empirical testing revealed several facts about these map matching algorithms and has inspired us on how to adapt them or how to design our own in order to solve the matching problem in travel/transportation studies.

- Online map matching algorithms typically ignore or can not use the global information contained in the data. Occasionally a GPS position can be matched to a branching or disconnecting road link (Figure 1). Problems arose when the recorded GPS points are sparse due to fast travel speed by the carrier - a short road link could be ignored and unmatched (Figure 2).
- In traditional GIS digital maps, nodes are not always digitized at street intersections and their density varies across the map. The position of a node plays an important role in the node-to-curve based map matching algorithms. If insufficient attention is paid to maintain the consistent topology of matching results, one minor mismatch could lead to a blunder (broken links or gaps).
- Offline algorithm could use global optimization techniques such as the shortest path algorithm to generate a topologically correct route. However, the assumption that no road links have ever been repetitively visited deprives its ability to differentiate the travel loops.
- Both types of algorithms could utilize the fact that travelers constrain their travel within the attributed road network. Thus, the particular regulations about the road links (speed limits, one-way streets, etc.) could be used to help enhance map-matching accuracy.

[Insert Figure 1 about here]
[Insert Figure 2 about here]

With these rationales in mind, we determined to take a heterogeneous approach for map-matching travel/transportation data. The targeted data input source for the algorithm is limited to GPS records. The algorithm consists of three phases: data preprocessing, multiple hypothesis map matching with rank aggregation and Dempster belief test.

3.1 Data Preprocessing – Cluster Reduction and Density Leverage

The new algorithm adds a data preprocessing step prior to the real map matching work. It consists of two steps: cluster reduction and density leverage.

Cluster reduction is meant to reduce the systematic noise in the data. Usually it is not easy to qualify the moving/still state solely based GPS receiver's input, especially when a tracking device is used to collect travel data across the full spectrum of travel modes. Even when the carrier keeps still at a fixed location, a GPS device would record a cluster of positions indicating random deviations around the true position point, which phantoms the slow moving speed of the carrier and random travel directions. Due to their unpredictability and falsifying characteristics, the GPS point clusters could be extremely misleading to map-matching procedures and are the greatest cause of overshoots and mismatch. With the spatial clustering modeling technique, the GPS collected travel data are filtered first to substitute the point clusters with their centroids. To model and identify these clusters, we selected the DBSCAN (Ester et al., 1996) clustering algorithm for cluster searching since it allows lack of information on the number and shape of the clusters in the input data.

Density leverage is meant to dynamically adjust the data sampling frequency against the model resolution of the base street map. The matching street layer consists of various lengths of street links. Similarly, the sampling interval of the GPS receiver varies with the carrier's moving speed and direction. Whenever the sampling interval is greater than the length of a traversed street link, there might be the chance that the street link is omitted from the matching algorithm, resulting in gaps in the match result. After the cluster reduction handling, the GPS trace data is streamlined in units of two. Every two GPS points are processed to generate a combined buffer area around them. If the sample distance between the two points is greater than half of the minimum-length street link that falls in the buffer, additional false data points are interpolated and inserted into the trace sequence.

3.2 Multiple-Hypothesis Matching Algorithm with Rank Aggregation

Borrowing the concept from the genetic algorithm, the map matching method we propose and implement keeps a pool of the best solutions. The solution pool is updated sequentially with the ordinal encountering of street intersections along the travel route. The GPS recorded travel trace is treated as a translated and rotated version of the match route. During the search for the best candidates, both the accumulated 2-norm distance (A2ND) and the rotational variation metric (RVM) (Joshi, 2001) are used to evaluate the matching result and guide the search directions around street connections. Norms constitute a quantitative measure of the geometric displacement between the GPS trace and the actual travel route. RVM, which accumulates the degree of variance between the orientations of two geometric shapes, measures the geometric distortion between them.

The algorithm starts with creating a pool of (N) seed candidates by buffering around the first valid GPS record. Any street segment that falls within the buffer is selected as one of the potential matches of the travel route start. Continuing with the next temporally adjacent GPS trace point, the norm distance between the GPS points and its projection on the current match link is computed and accumulated into A2ND as the match score of the current match candidate. In the mean time, a direction discrepancy between the current travel direction as indicated by the GPS records and the current match link is computed and accumulated to RVM metric. A2ND and RVM both serve to guide the match search in the street network space. However, we did not try to multiplex them to produce a single matching index, as a fixed or dynamic weighting schema are difficult to specify and unlikely to suit every possible individual tracing case. The partial match results are ranked in A2ND and RVM separately. Only the top N results of both are kept for the next round of match growth.

As the partial match results growth encounters a street intersection, a ground rule is set up to decide the right timing of when to select the next link to further the matching process. Two cases exist to judge when the traveler began to leave the current link and transit to the next one: 1) The projection of current GPS point falls on or out of the end point of the current link, which typically occurs when the travel direction change is less than or equal to 90 degrees; 2) The projection of the current GPS point comes near to the end point of the current link, but the point's position is getting away from the current link, which typically occurs when the travel direction change exceeds 90 degrees. For the second case, we set up two threshold values to switch on the turning signal -20 meters for coming near to the end point of a link and 30 meters for leaving the current link. When determining the next link, all the topologically connected links to the intersection node are considered as the potential next links, including the incurrent link to cover the U-turn situation. However, prohibited maneuver and turn restrictions information has been used to pre-eliminate certain search branches efficiently.

After the matching process is completed, a pool of top N match results is derived with different rankings of A2ND and RVM measures. With the rank aggregation method, we may combine the ranking of the two to obtain an aggregated ordering. Ideally, Kemeny ordering minimizes the sum of the "bubble sort" distances and thus generates the best compromise ranking. However, it is a NP-hard problem (Dwork et al., 2001). Here, we implemented two of the other heuristic/sub-optimal ranking aggregation methods to composite the ultimate matching results (Table 1): 1) the simple Borda's method to generate a combined ranking for the pool of match results: Each candidate in the pool is assigned a score of the number of candidate ranked blow it. Its total score across the different ranking list is finally sorted in a descending order; and 2) a good approximation to Kemeny optimized rank aggregation – footrule optimal aggregation, which finds the median permutation of the rank lists to be combined.

[Insert Table 1 about here]

3.3 Dempster Belief Test for Travel Off-Road/Noise Discernment

As discussed in the previous sections, uncertainties typically exist in both the trace data and the base matching street map. Given the dataset and the matching base, match results are considered as always producible without setting any restraint on the acceptable belief and plausibility level. However, considering the possible travel by walk mode, a matching algorithm could easily map a pedestrian travel onto a highway link nearby. Or under other scenarios, the GPS device could be "blacked out" by the surrounding tall buildings. In either case, the assumption that a candidate match link is identifiable from the base street map becomes void.

This research takes the advantage of Dempter-Shafter theory (Shafer, 1976) to fuse heterogeneous information in order to discern the off-road travel and GPS black-out situations. For each of the matching select criterion (proximity and direction), a frame of discernment {yes, no, perhaps} and its belief assignment functions as similar to Najjar and Bonnifait's (2002) was built to test if a matching link is a "good match". Each discernment type of a select criterion is associated with two quantities: belief and plausibility. The GPS trace point match to a link is considered invalid if the no belief value is greater than the plausibility values of the other two assumptions, the no belief value is 1, or the conflict parameter has a value of 1. A consecutive trace of more than 10 invalid GPS point-to-link match invalidates the corresponding segment of match results, which is then splinted out and replaced with the original GPS travel trace.

3.4 Match Results

Map matching is performed against the Dynamap/transportation data of Santa Barbara from GDT, Inc. It contains complete address information and routing features, including speed limit, cost, turn restrictions, one-way street information, etc. The travel data comes from a travel survey conducted locally. A test run of the algorithm against a single GPS trace generates the following results: Figure 3 and figure 4 show data processing effects of cluster reduction and density leverage, respectively. Figure 5 show the best match result from the match candidate pool as indicated by both the Borda and Footrule ranking aggregation methods. The second best match result as indicated from the Borda method indicates an additional matching link at the end of travel route, while the second best match result as indicated by the Footrule method indicates an alternative matching link at the end of a travel route. They show up with the minimal difference from the best, except the divergence at the end of the identified travel route. Table 2 shows the part of Dempster belief test result for the test route. The final part of the travel is discerned as the part of off-road travel and replaced with the original GPS data. In all, the algorithm generated a perfect match for the GPS trace input, without branch or gap in between, and segmented out the off-road travel portion with high accuracy (Figure 6).

[Insert Figure 3 about here]
[Insert Figure 4 about here]
[Insert Figure 5 about here]
[Insert Table 2 about here]
[Insert Figure 6 about here]

To fully test the effectiveness of the matching method, the algorithm is used to match a series of travel traces collected from a travel behavior survey against the local base map of the Santa Barbara. The matching percentage is derived according to the following equation: 100 * (1- ED/n). Here n is the total number of arcs in the actual route. The difference between the actual travel route and the final match result is evaluated based on the concept of "edit distance" (ED) – the minimum number of insertions, deletions and substitutions needed to transform one route to another. On average the map matching algorithm reaches a matching accuracy of 95.74%. The further break-down of matching results by travel mode is shown in Table 3. It can be seen that travels by bicycle are subject to the most map matching errors, which are mainly due to the lack of bike path information in the base map.

[Insert Table 3 about here]

Considering the performance of the algorithm, suppose the number of the road links in the road network is N, the number of the road link extremities is V, the number of collected GPS points is M, and the size of the candidate Pool is K. The time complexity for the DBSCAN algorithm would be $O(M * \log(M))$ with R-tree implementation. Density leverage involves a spatial buffering operation, and, hence, is the most costly. Its time complexity is up to $O(M * \log(N))$. The map matching step involves $O(M * K)$ for candidate searching, and $O(K * \log(k))$ for pool updating at each step. Its total complexity is $O(M * K^2 * \log(K))$. The Dempster belief test at the end incurs an additional $O(M)$ time cost. When K is small and N/M is large, the time complexity of the algorithm is comparable to most of the algorithms we discussed in section 3.

4 Conclusion

In this paper, we briefly overviewed the currently available map matching methods. Several recent online/offline map matching algorithms were implemented in GIS to provide a case study to evaluate from multiple perspectives. In recognition of the disadvantages associated with the methods examined, this research proposed and implemented an innovative map matching approach suitable for travel/activity research needs which is uniquely characterized by: 1) data preprocessing with point cluster reduction and density leverage, 2) offering the candidate solution within a pool of "the best," 3) balancing of matching results from multiple matching factors with rank aggregation, 4) intelligently utilizing the basic network constraint attributes with "expert rules" to increase the matching accuracy, and 5) Dempster belief test to discern the noise and off-road travel. Our analysis has shown that the performance of the new algorithm is comparable with the others when the candidate pool size is small and network/GPS trace size is large. An application of the matching algorithm to a complete set of travel routes collected from a travel behavior study has shown acceptable matching accuracy across the different travel modes.

Acknowledgement

References

Bernstein D. and Kornhauser A., 1996, An introduction to map matching for personal navigation assistants. New Jersey TIDE Center.

Bernstein D. and Kornhauser A., 1998, Map matching for personal navigation assistants. In *77th annual meeting, The Transport Research Board*, 11-15 Jan 1998, Washington, D.C.

Bétaille D. and Bonnifait P., 2000, Road maintenance vehicles location using DGPS, map-matching and dead-reckoning: experimental results of a smoothed EKF. In IAIN *World Congress in association with the US ION annual meeting*. San Diego, 26-28 June 2000, pp. 409 – 416.

Bonnifait P., Bouron P., Crubille P. and Meizel D., 2001, Data fusion of four ABS sensors and GPS for an enhanced localization of car-like vehicles. In *Proceedings of the 2001 IEEE. International Conference on Robotics & Automation*. 21-26 May 2001, Seoul, Korea, vol. 2, pp. 1597-1621.

Dwork C., Kumar R., Naor M. and Sivakumar D., 2001, Rank aggregation revisited. Available online at: http://citeseer.ist.psu.edu/478775.html (accessed 21 August 2006).

Ester M., Kriegel H., Sander J.and Xu X., 1996, A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining*. Portland, Oregon, August 1996, pp. 226-231.

Joshi R.R., 2001, A new approach to map matching for in-vehicle navigation systems: the rotational variation metric. In *IEEE Intelligent Transportation Systems Conference Proceedings*. 25-29 August 2001, Oakland, CA, USA. pp. 33-38.

Kim S. and Kim J., 2001, Adaptive fuzzy-network-based C-measure map-matching algorithm for car navigation system. IEEE, *Transaction on Industrial Electronics*, Vol. 48, No. 2, pp. 432 – 441.

Lakakis, K., 2000, Land vehicle navigation in an urban area by using GPS and GIS technologies. PH.D. Thesis, Aristotole University of Thessaloniki, Department of Civil Engineering. Thessaloniki, Greece.

Lakakis, K., Savvaidis P., Ifadis M. I., and Doukas D. I., 2004, Quality of map-matching procedures based on DGPS and stand-alone GPS positioning in an urban area. In *Proceeding of FIG Working Week -- the Olympic Spirit in Surveying*. Athens, Greece, 22-27 May 2004.

Marchal, F., Hackney J. and Axhausen K.W., 2004, Efficient map-matching of large GPS data sets - Tests on a speed monitoring experiment in Zurich. *Arbeitsbericht Verkehrs- und Raumplanung*, 244, IVT, ETH Zürich, Zürich

Najjar M.E. EL. and Bonnifait Ph., 2002, Multi-Criteria fusion for the selection of roads of an accurate map. In *15th IFAC World Congress*, Barcelona, 21-27 July 02.

Najjar M.E. EL. and Bonnifait Ph., 2003, A roadmap matching method for precise vehicle localization using belief theory and Kalman filtering. In *Proceedings of ICAR. the 11th International Conference on Advanced Robotics*, June30 – July3 2003, Coimbra, Portugal, pp. 1677-1682.

Pyo J.-S., Shin D.-H., and Sung T.-K., 2001, Development of a map matching method using the multiple hypothesis technique. In *Proceedings of IEEE Intelligent Transportation Systems Conference*, 25-29 August, 2001, Oakland, CA, U.S., pp. 23-27.

Quddus, M. A., OChieng W.Y., Zhao L. and Noland R.B., 2003, A general map matching algorithm for transport telematics applications. *GPS Solutions Journal*, vol. 7. no. 2, pp. 157-167

Rogers, S., 2000, Creating and evaluating highly accurate maps with probes vehicles. In *Proceedings of IEEE Intelligent Transportation Systems*, 1-3 Oct 2000, Dearborn (MI), U.S. pp. 125-130.

Shafer, G., 1976, *A Mathematical Theory of Evidence*. Princeton University Press, Princeton.

Syed S. and Cannon M.E., 2004, Fuzzy logic based-map matching algorithms for vehicle navigation system in urban canyons. *Presented at ION National Technical Meeting*, 26-28 Jan 2004, San Diego, CA.

Taylor, G., and Blewitt, G., 2000, Road reduction filtering using GPS. In *Proceedings of 3rd Agile Conference on Geographic Information Science*, 25-27 May 2000, Helsinki, Finland, pp 114-120.

Taylor, G., Uff J. and Al-Hamadani A., 2001, GPS positioning using map-matching algorithms, drive restriction information and road network connectivity. *Presented at GISRUK 2001, 9th Annual Conference*, 18-20 April 2001, University of Glamorgan.

Walter V. and Fritsch D. (1999). Matching Spatial Data Sets: A Statistical Approach. *INT. J. Geographical Information Science*, vol. 13, no. 5, pp. 445-473.

White, C. E., Bernstein D. and Kornhauser A. L., 2000, Some map matching algorithms for personal navigation assistants. *Transportation Research Part C* 8(1-6): 91-108

Xiong, D., 2000, A three-stage computational approach to network matching. *Transportation Research C*, Vol. 8, pp. 13-36.

Yin H. & Ouri W., 2004, A Weight-based map matching method in Moving Objects Databases. In *Proceedings of the 16th International Conference on Scientific and Statistical Database Management (SSDBM)*, 21-23 June 2004, Santorini Island, Greece.

List of Illustrations:
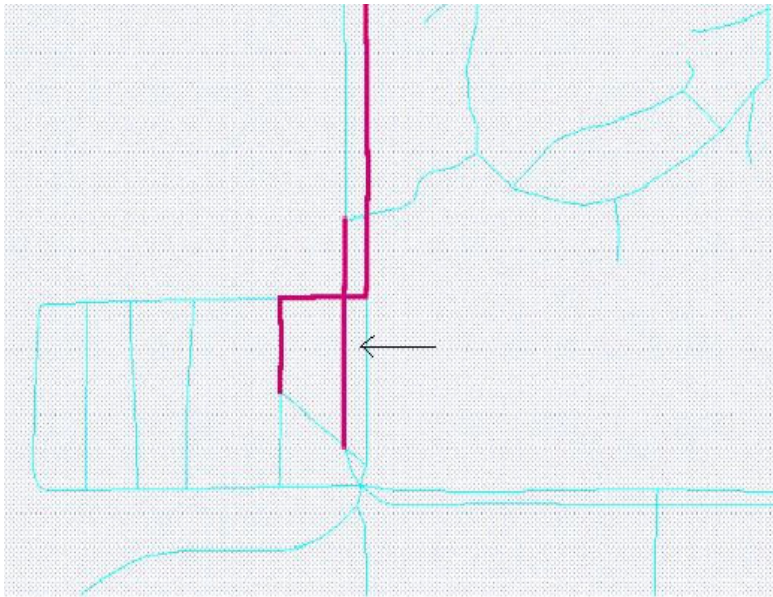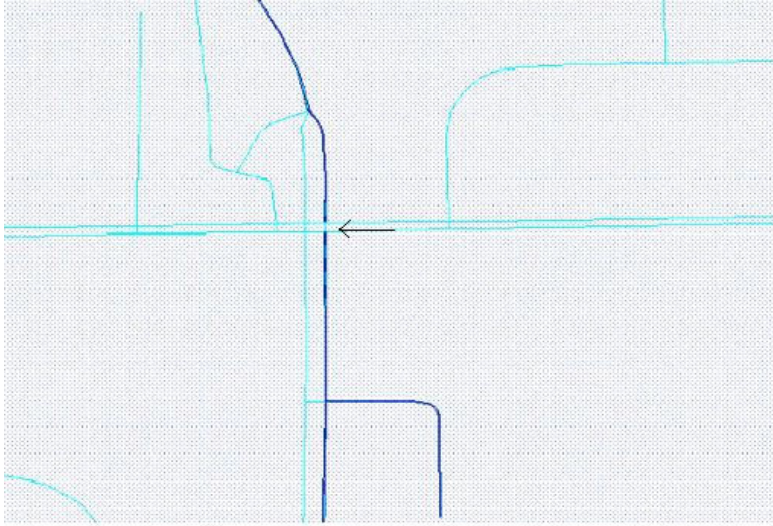
Figure 1 Overshoot scenario

Figure 2 Gap scenario

Figure 3 A cluster of GPS points recorded around street intersection is recovered via DBSCAN algorithm and replaced with its centroid
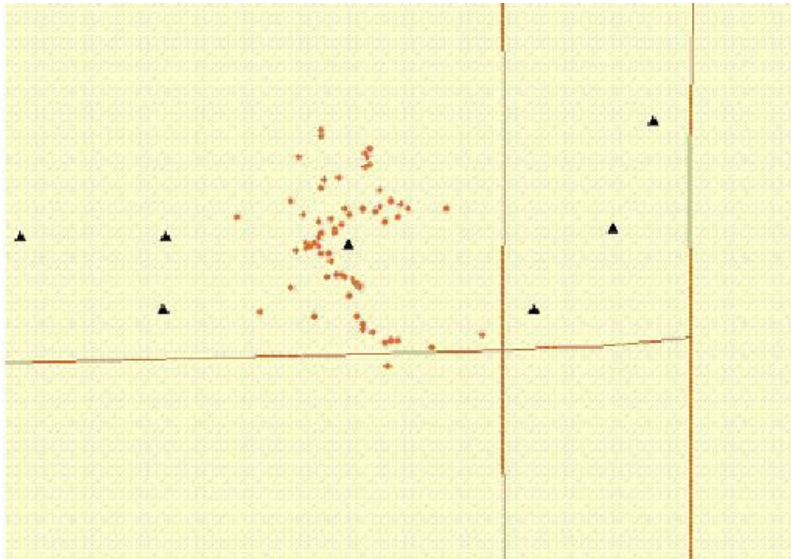
Figure 4 Original GPS trace (as indicated by the triangle symbols) didn't sample the short street intersection. Augmented traces by density leverage added three false GPS data points to cover it.
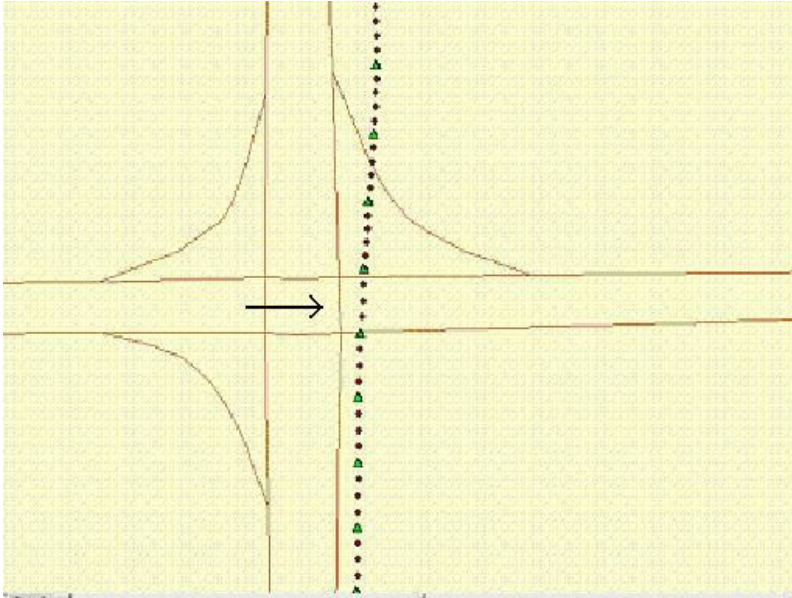
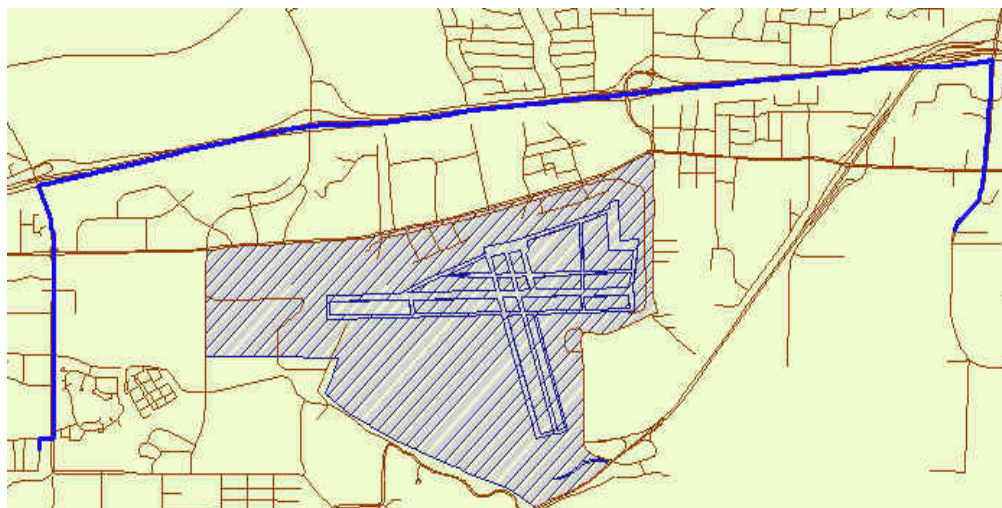Figure 5 The best match result from both Borda and Footrule ranking aggregation.

Figure 6 Map matching result after Dempter belief test with multi-criterion fusion.
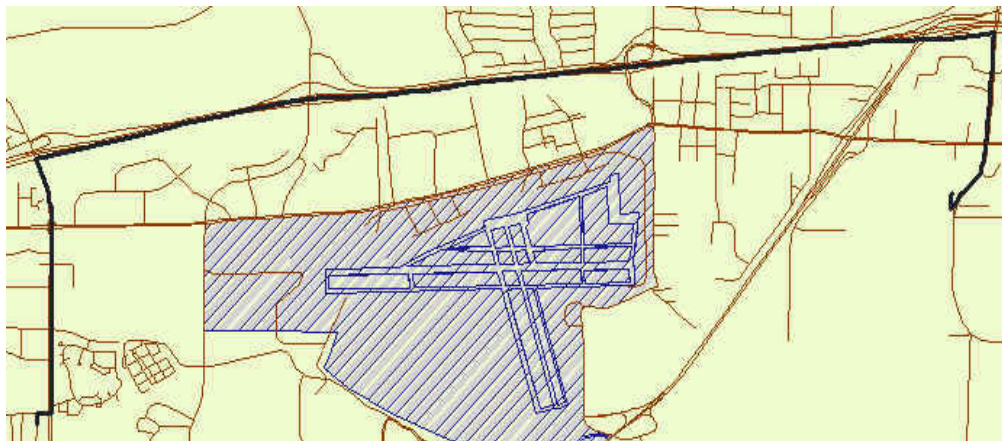
Table 1 Match Result Summary (including L2norm and RVM measures, rankings, ranking results from Borda and Footrule ranking methods, map matching trip distance and final trip distance after Dempster belief test.

L2Norm	RVM	L2NormRank	RvmRank	BordaRank	FootRuleRank	TripDistance	BeliefTripDistance
0.181785	565.459956	1	9	5	2	5.481127	-1
0.181879	570.568439	2	10	8	7	5.618900	-1
0.191356	476.092263	3	3	1	1	5.427727	5.480326
0.191451	481.221254	4	5	4	4	5.565500	-1
0.198498	479.225220	5	4	3	3	5.448950	-1
0.198593	484.334925	6	6	7	6	5.586724	-1
0.209626	441.107106	7	1	2	8	5.429427	-1
0.209627	499.215457	8	7	9	5	5.435327	-1
0.209720	446.224142	9	2	6	10	5.567200	-1
0.209721	504.321745	10	8	10	9	5.573100	-1

Table 2 Portion of the Dempster fusion result for the test travel route (PI = Plausibility)

Attributes of fusionresult

PointID	YesBelief	NoBelief	PerhapsBelief	YesPI	NoPI	PerhapsPI	FusionConflict
1	1	0	0	1	0	0	0.731278
2	0	0	0	0	0	0	1
3	0	0	0	0	0	0	1
4	0	0	0	0	0	0	1
5	0.057267	0	0.942733	0.057267	0	0.942733	0.562581
6	0	0	1	0	0	1	0.737395
7	0.823912	0	0.176088	0.823912	0	0.176088	0.481990
8	0.157464	0	0.842536	0.157464	0	0.842536	0.619850
9	1	0	0	1	0	0	0.960352
10	0.903953	0	0.096047	0.903953	0	0.096047	0.705019
11	0.336617	0	0.663383	0.336617	0	0.663383	0.736978
12	1	0	0	1	0	0	0.140969
13	0.919172	0	0.080828	0.919172	0	0.080828	0.353459
14	0.828980	0	0.171020	0.828980	0	0.171020	0.444416
15	0.900334	0	0.099666	0.900334	0	0.099666	0.611730
16	0.517693	0	0.482307	0.517693	0	0.482307	0.511791
17	0.991111	0	0.008889	0.991111	0	0.008889	0.339207
18	0	0	0	0	0	0	1
19	0.983086	0	0.016914	0.983086	0	0.016914	0.249287
20	0.887415	0	0.112585	0.887415	0	0.112585	0.387752

Table 3 Map Matching Accuracy by Travel Modes

Travel Mode	Total Number of Trips	Average Matching Accuracy (in percentage)
Walk	104	94.75
Car	234	97.04
Bicycle	61	91.19
Carpool	34	97.17
Vanpool	3	100.00
Local bus	45	96.17
Total	481	95.75