Spatio-Temporal Methods to Reduce Data Uncertainty in Restricted Movement on a Road Network

Nirvana Meratnia, Wolfgang Kainz, and Rolf de By

University of Twente, Department of Computer Science, Enschede, The Netherlands, [Meratnia, Kainz]@cs.utwente.nl, deBy@itc.nl

Abstract

Much of the success of location-based applications comes from using spatial analysis in query processing and graphic visualisation. Recently, with the arrival of miniature and inexpensive GPS receivers and cellular phones, new horizons have been opened for these applications. The mobile telephone and Internet technology within the GIS environment has fostered new applications with higher degrees of user satisfaction. The advance this technology affords for data acquisition, facilitates the provision of real-time location-based services.

Recently, considerable work has been undertaken in feasibility assessments and in identifying the type of services and applications that can be offered to mobile persons in an environment equipped with mobile telephones, Internet technology and GIS. No equivalent attention has been given, however, to the data that will be used for running such services. The acquired data have some degree of uncertainty mainly due to system error associated with acquisition devices and limitations in representing continuous data temporally.

By focusing on restricted movement on a road network, this paper proposes a spatial registration technique to reduce the error associated with location data; it also proposes consistency checks and time registration in an effort to compensate for times when no observation is available. The processed data will also reduce the ambiguity in determining the current location of an object within the network.
Keywords: restricted movement, mobile person, spatial/temporal registration, data uncertainty, consistency checks

1 Introduction

The situation in which the whereabouts of objects[1] are constantly monitored and stored are an important class of problem that present-day database/GIS users will find hard to tackle with their current systems. The problem arises as a result of inappropriate data models, data structures and operations, indexing methods and query processing techniques, all of which are necessary to handle continuous changes in moving objects along with the volume of data that is generated. For example, existing query languages are mainly non-temporal and limited to accessing a single (current) database state, not past and future ones (Sistla et al, 1997).

Much of the success of a GIS project depends on the quality of the data that is entered into the system (de By, 2000). Of course, the data uncertainty has various implications on data modelling, querying and indexing (Wolfson et al, 1999). To handle the uncertainty, different methods should be used. As (Márkus, 1997) emphasised, uncertainty in a GIS environment may occur in three phases: data acquisition, data processing and in using information. In general, the uncertainty in data acquisition may occur due to instrument errors, inappropriate frequency of data acquisition, data transmission errors, and poor data acquisition environments. With respect to a poor environment, it is clear that certain data acquisition devices cannot work properly in some situations, such as poor weather conditions. On the other hand, factors such as inappropriate sampling methods, data storage, data representation methods and interpolation techniques also affect uncertainty during the processing phase. Although errors occur as a result of data processing and data use, most GIS errors result from improper or inadequate data input (Demers, 1998).

One of the sources of uncertainty for a moving point object, comes not only from the error associated with the data acquisition devices, but also from the fact that data is acquired in a discrete way. Since computer systems cannot directly represent continuous phenomena of an arbitrary nature, such phenomena must be approximated using a finite structure. The uncertainty of the representation of an object's movement is affected by the frequency with which position samples are taken, i.e., the sample rate. The sample rate may be defined by considering the speed of an object and the desired maximum distance between consecutive samples (Pfoser et al, 1999). It is also necessary to think about how to handle cases when the connection between the object and data acquisition device is lost and data cannot be updated (Wolfson et al, 1999). Uncertainty in the data can accumulate from different sources. Research on uncertainty in geo-spatial information is usually concerned with sources of uncertainty such as the correctness and completeness in the measurement, and the analysis and interpretation of digitally represented, geo-referenced phenomena (Pfoser et al, 1999). As error in input data can be high and may be propagated in further

---

[1] Of all physical objects, i.e., objects with clear boundaries that follow the Newtonian laws in their movement, cars are covered in this study as an example of objects with restricted movement on a road network.

processes, this paper focuses on reducing error in the object locality data while ignoring other types of error.

It is also important to notice that uncertainty has cost, which means higher uncertainty leads to less retrieved information (Wolfson et al, 1998) and consequently results in partially or misleading outcomes.

Another critical issue concerns the sufficiency of the number of observations. Too few observations give ambiguous results in interpolation techniques. Although it is possible that an application adds data, it should be done on the basis of rules describing object movement. On the other hand, too many observations consume storage space and higher data acquisition frequency causes more data traffic in wireless environments and therefore may be expensive (Chamberlain, 1999).

Obtaining a "faithful representation" of moving objects with a sufficient number of discrete (possibly erroneous) data points is the goal of this paper. Such representations can open the door for recognition of movement patterns and object classification. These may consequently introduce information about the environment where the objects move and assist the query processing. For example, in locating the shortest path between two points in a road network or in finding the closest feature of interest, not only the static spatial aspects should be considered, but also the characteristics of the road network (such as average speed for each road, delay times at intersections, among other aspects). Such information can be obtained by studying patterns of movement. Thus, unambiguous and optimal representation of moving objects is crucial.

This paper proposes spatial and temporal registration for the reliable representation of moving objects by minimising possible errors in the data. Section 2 contains the description of the methods, addressing the issue of reducing spatial error in the data, followed by explaining the consistency check approach to provide more reliable data for further processing. Time registration is also introduced to handle situations in which no observations are available. Results of experiments and description of data used in these experiments are presented in section 3. Finally, the paper concludes with results of some experiences and recommendations for future work.

2 Methods

Computations about moving objects require the use of positional interpolation techniques, in which one of the fundamental issues is the "medium" via which the objects are travelling. The medium may or may not impose restrictions on where objects go. Among all three possible types of movement, namely unrestricted movement in 2D or 3D, restricted movement in 2D or 3D and restricted movement in a 2D or 3D network, this study focuses on the latter. The observed object can be a car equipped with a GPS receiver or a person travelling in a car carrying a cellular phone that enables position determination.

It is assumed that there is a central base station that receives the location and time data in addition to the queries (if any) from the moving person/car and sends the query results back. The central base station can process two types of query. The first type aims to provide statistical information about a group of objects, i.e. information about the road network status such as average speed for a certain road, or waiting time at a certain intersection. Such information can be obtained based on the historical data from previously monitored objects and can be computed at different times a day, different days a week and so on. The second type of query is initiated by a travelling person and concerns the position of the individual with respect to other persons and phenomena. In this case, the history is collected in addition to the information about the road network (that is provided by the statistical information) and can be used to answer the query. For example, the current and estimated future speed and acceleration of the car, together with information about the estimated waiting time and average speed along the road segments are used to answer the query about the minimum estimated time it takes to reach a destination.

An essential assumption that we have made about data is that there is no error associated with time measurements, while the location data may have error. This can be a fair assumption since at each time (whether correct or incorrect) there is a location, while at a certain location (if it is wrong) there is no time that the person/car passes through. Since there is uncertainty about the location of the various cars at any point in time, we may not be able to answer the queries with absolute certainty; therefore, a degree of probability may be used (Sistla et al, 1999).

Our objects of study include *moving objects,* such as cars or persons, and the *road network*, which consists of nodes with geometry and links with topology. Regardless of the data acquisition method, the main part of the moving object data is in the form of $<O\_id, X, Y, T>$ and we call it *census point*. For a census point, X and Y represent the position of the moving object, with identification $O\_id,$ at time T. While data can be acquired in equal or random time intervals, the size of time intervals should be chosen in a way that does not lead to too few or too many observations. For the sake of simplicity, $<Road\_id, From, To, Speed>$ are assumed to be the main attributes of a road network segment. *From* and *To* indicate start and end nodes, *Speed* is the maximum allowed speed for a road segment with identification *Road_id*. We assume road network segments are one way. Other attributes can be added. The geometry is stored in the nodes.

2.1 Spatial Registration

As mentioned before, data coming from the data acquisition phase may be erroneous and before any further processing, it is important to correct the data to the best of our knowledge. To do so, the following procedure is proposed, in which restricted movement on the road network plays an important role:

- For each census point, the nearest road segment(s) will be identified. To do so, the minimum distance between each census point and road segment is computed.
- The *Road_id* corresponding to the minimum distance will be assigned to the census point, indicating the most probable road that the object was on. Fig. 1 shows the details. In our figures, road segments are illustrated as *R1, R2, ...* and *R15* and census points are illustrated as i, i+1, i+2, i+3, ….
- Next, the consistency of the assigned *Road_id*s will be checked. We will examine three distinct consecutive *Road_id*s at a time and the key to accept a sequence is the connectivity of the middle road to the next and previous one.
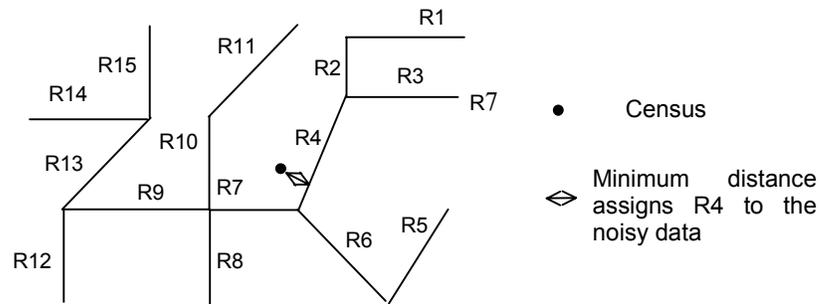

Fig. 1. The selected road segment

Fig. 2 shows acceptable and unacceptable sequences. For the census point sequence shown in A, the logical road sequence will be R1, R2, R3 and R4. Having a data in R'2 obviously indicates an error. For such a census point sequence, case B is not acceptable while case C is an acceptable road order.
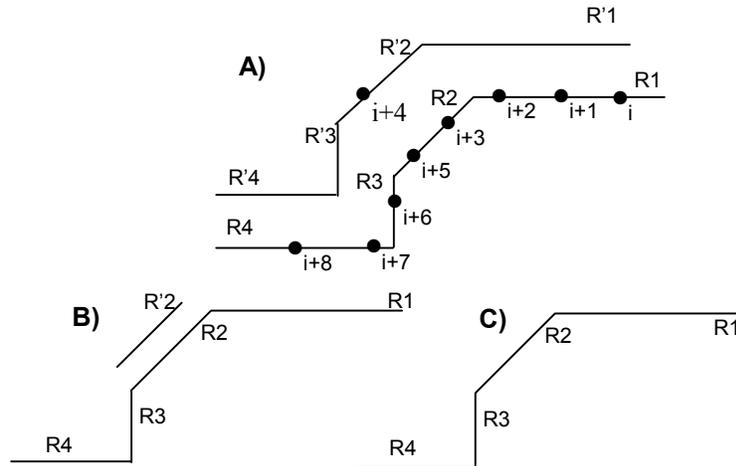
Fig. 2. (Un) Acceptable sequence in consistency check

- Following the consistency check, the census point corresponding to the unacceptable road segment is re-projected to the previous or next road, whichever is closer in order to maintain the consistency.
  The projection of the raw data into the correct or the most probable road leads to the identification of the error associated with the data.
- The length of the projection (perpendicular to the road) describes the error in one dimension; X, Y, or inclined direction depends on the direction of the road. It seems logical to assume an homogeneous error, in which the error is considered to be circular; this implies the error in X is equal to the error in Y or in other words, the error along the road equals the error perpendicular to the road on average.

This method is used to reduce uncertainty in the data. Since the census point is projected spatially to the most probable road segments, this procedure is called "spatial registration".

An important question remains, namely "when should data be modified?" The first option is real-time modification, which leads to storing the modified data. Although it seems to be a time consuming operation, data will be ready for further processes and any posed queries can be answered quickly later on. The second option is storing the noisy data and modifying them when it is necessary, meaning when a query is posed. This option speeds up the data acquisition and storage phases but slows down the query-processing phase. Choosing any of the options definitely depends on the application goals.

2.2 Time Registration

As data is acquired and represented in a discrete way, an appropriate technique is needed to provide the data when no observation is available to determine the historical path and future locations. In such cases, positional interpolation is used. As the method of interpolation is entirely dependent on the characteristic of the moving object, i.e. its maximum speed, maximum acceleration, maximum angular speed, and so on, an important question is which interpolation technique should be chosen (Meratnia, 2000). On the other hand, concerning our example other factors, such as traffic jams, amount of fuel and human decisions have an important influence in predictions concerning location and time.

Since we do not measure the positions in-between two consecutive position samples, the best we can do is to limit the possibilities of where the moving object could have been (Pfoser et al, 1999). Our assumption about restricted movement on a road network necessitates that the location coincides with the road network and helps to narrow the possible area that a car has visited. The procedure to obtain the data in-between two consecutive position samples consists of three main steps:

- Define a general function that describes the car movement at any point in time. It has to be mentioned that the path of the car following the road segments has been determined in the consistency check and is restricted to the road network. The time is known, however, only for those points where there is data. Therefore, the time registration function is chosen as "D(t)" or distance along the network as a function of time. Different interpolation techniques can be used (and compared) to provide better understanding of the car movement. In this study, two interpolation methods are used, namely spline and linear interpolation.
- Constrain the function "D(t)" to the location and time data obtained from the spatial registration step. This means that to estimate the parameters involved in the distance function, in addition to the corresponding time values, the distances along the road network between the census points are used. Fig. 3 shows how the distance between points is restricted to the road network. In this figure, the object is moving from road segment R1 to road segment R3, according to the spatial registration of the points and the consistency check. Distance is measured along the road network. Therefore, the distance D_{i+2} is being forced to pass through the intersection node between R1 and R3, instead of passing through a straight line connecting data points i+1 and i+2. This information would not be available if the spatial registration and the consistency check had not been performed.
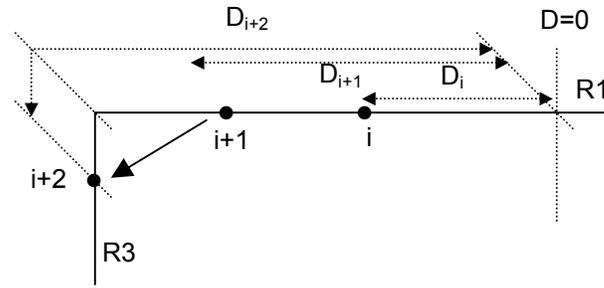
Fig. 3. Constraining distances to the road network

- As a rule of thumb, whenever the number of equations using the available data exceeds the number of parameters to be estimated, a least-square adjustment is used to minimise the possible errors and introduce a confidence measure for the estimated parameters. Examples can be found in (Meratnia et al, 2001).

3 Experiments

The proposed methods were tested using simulation techniques. The evaluation procedure starts by simulating so-called true data for both the object and the road network. In the next two subsections, the data description and the analysis of the results are discussed.

3.1 Data Description

The data has two parts; the road network data and the moving object (car/person) data. The road network consists of road segments (or links) and intersections (or nodes). The structure of the road network has been simulated and includes topology and geometry. The topology is associated with the links as each link has a start and end node. The geometry is associated with nodes as each node has a location. We assume that the road network data is more accurate than the car data. No error is assumed associated with the road network, however, errors can be incorporated in the registration analysis.

Other data that we have simulated are the speed limits and the direction of each road segment. The intersections have been designed to have green and red traffic lights to regulate the incoming and outgoing cars. The green and red traffic light time intervals were simulated independent of the cars and were defined as a global function of time. Therefore, at any point in time, it is determined which car is allowed to pass and which car has to wait and for how long.

The simulation had the cars follow the above road network rules. Each car starts from a randomly selected road segment and continues by maintaining the

connectivity between the consecutive links. The average speed of a car in a road segment was simulated to have a random value, which does not exceed the permitted speed in that road segment. Acceleration (at starting points as well as intersection points) the same as deceleration (when the car approaches an intersection as well as when it is going to stop) were also simulated.

The path determined for cars was temporally discrete in two ways; every second (data was provided for accuracy determination), and every minute (data were introduced to test the distributed errors, for mean, and standard deviation to both X and Y direction of the data).

3.2 Results

Comparing true and noisy data, and computing the RMSE indicates a 2.4m error on average in the data. The RMSE was computed as the average of the spatial errors in the data. Although the directions of the error vectors may differ, this value gives an indication about the average size of the error buffers. The procedure continues by applying the proposed correction method on the noisy data by projecting them on the road network and performing the consistency check. Comparing the noisy and the corrected data, the estimated error results in 1.8m, measured perpendicular to the link directions. It is clear that the average error in data has been reduced by 25%. We will assume that the error is homogeneous and therefore the error component in the direction of the links equals that in the perpendicular directions. The computed RMSE for the corrected data compared to the true data (which we only have in the simulation) is 1.5m, which is very close to our original estimate. Therefore, if we introduce this value as uncertainty along the link we can achieve more accurate results in an adjustment procedure. Fig. 4 shows the difference between the true, noisy (raw), and the corrected data.

In time registration evaluation, spline and linear interpolation are used separately to provide the complete travelled path with distance as a parametric function of time.
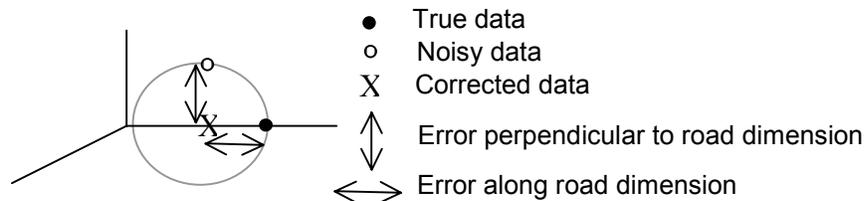

Fig. 4. Homogeneous error

The computed spatial RMSE between the interpolated data points using spline and true data is 33.8m, while spatial RMSE between the interpolated data points

using linear interpolation and true data is 27.6 km. At first, one can consider it a very large error compared to the errors associated with the discrete raw data points. However, they are acceptable since an object is travelling this distance, (with a speed of 50 km/h) in just 2 seconds. It is noticeable that in using both Spline and linear interpolation techniques, errors mainly occur at the intersection points. It is   an expected result, since the behaviour near intersections is not smooth and is highly influenced by human decision and abrupt changes. Fig. 5 and (6) show the distribution of the errors in the direction of the links for the spline and linear time registration, respectively. It should be mentioned that the errors have been exaggerated and plotted normal to the links for better visualisation.
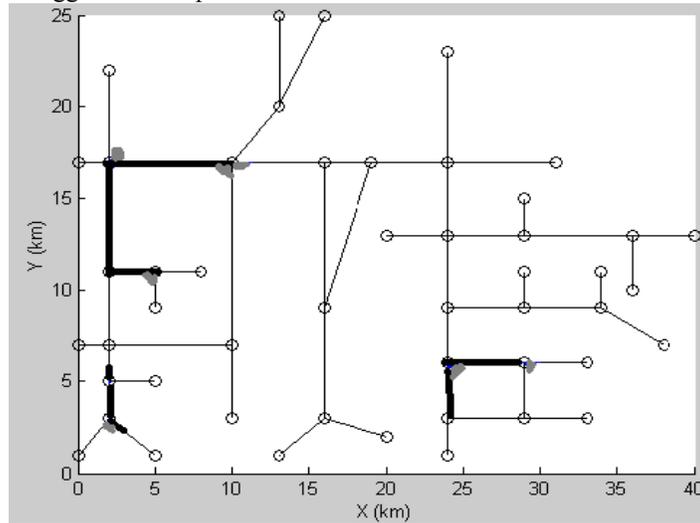
Fig. 5. The spatial error along the road network associated with spline time registration

It is also noticeable that although the linear time registration has less RMSE than the one of the spline time registration, the maximum errors are larger and their distribution is less homogeneous. This can be explained by the fact that more parameters of the Spline (near the intersections) model variations of the movement better, while the same number of parameters may not be required far from the intersections.
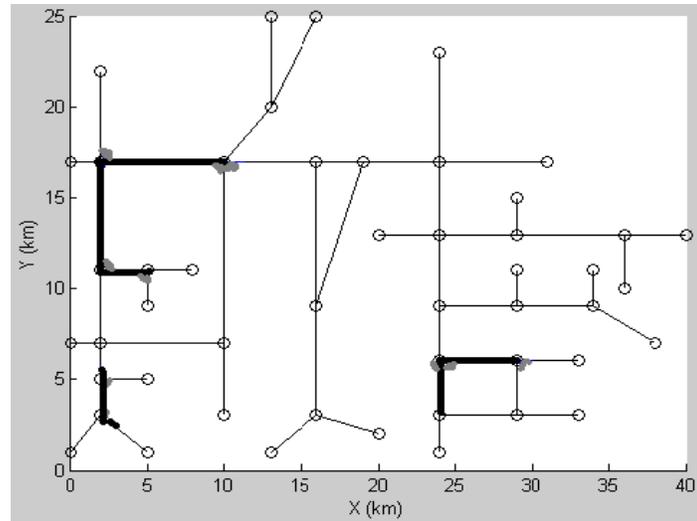
Fig. 6. The spatial error along the road network associated with linear time registration

4 Conclusions and Future Work

In this paper, methods for handling error in restricted movement on a road network have been proposed. They play an effective role to reduce error associated with the data. A consistency check has been suggested to reduce ambiguity in the data. Although effectiveness of any interpolation technique depends on the type of application using it, including the estimated error components, can be of help in identifying the appropriate function for each application type. More parametric time registrations are recommended, including automatic detection of the location and time, when abrupt changes occur in the movement. We assumed that error exists only in location of census point, while we recommend studying the case when time also is uncertain. For the time being, error in the GIS itself is ignored, while it needs to be taken into account in future work. Performing these methods on real data is useful in trying to overcome complex problems that have not been fully addressed. More error analysis is recommended including estimation of confidence intervals for the spatio-temporal results.

Issues such as defining the maximum effective time interval (i.e. the minimum data acceptable for economic reasons) to clearly represent the object within tolerance values of space and time, modelling intersections, and defining a cost model (time) are areas to be addressed in future work.

Moving objects can be used to provide valuable information about the environment through which they pass. Obtaining an accurate representation of cars opens the door to studying patterns and classes of cars/persons, which can assist in obtaining spatio-temporal information about the road network in addition to information about traffic conditions. Accessing such information helps to

answer many spatio-temporal queries about the history of persons/cars. Additionally, the information assists in predicting the future location of monitored persons/cars or providing spatio-temporal information for persons looking for closest (in terms of time and/or space) features of interest.

References

Chamberlain S, Wolfson O, Dao S, Jiang L, Medsez G (1998) Cost and Imprecision in Modeling the Position of Moving Objects. In: Proceedings of the 14th International Conference on Data Engineering (ICDE98). IEEE Computer Society, Orlando, Florida, pp 588-596

De By RA (2000) Principles of Geographic Information Systems. ITC. The Netherlands

Demers NM (1998) Fundamentals of Geographic Information System. John Wiley & Sons, INC

Márkus B (1997) Decision Support and Error Handling in GIS Environment. Habilitation presentation. University of Sopron, Sopron

Meratnia N (2000) Implementation of Data Structures and Operations for Moving Objects Modeling. M.Sc. thesis, ITC, The Netherlands

Meratnia N, De By RA (2001) Moving objects data modeling. In: Proceedings of ASPRS 2001 Annual Conference. St. Louis

Pfoser D, Jensen C (1999) Capturing the Uncertainty of Moving-Object Representations. In: Proceedings of the 6th International Symposium (SSD99). Springer-Verlag series LNCS 1651

Sistla P, Wolfson O, Chamberlain S, Yesha Y (1999) Updating and Querying Databases that Track Mobile Units. Invited paper in special issue of the Distributed and Parallel Databases. Journal on Mobile Data Management and Applications, pp 257-287

Sistla P, Wolfson O, Chamberlain S, Dao S (1997) Modeling and Querying Moving Objects. In: Proceedings of the 13th International Conference on Data Engineering (ICDE97). IEEE Computer Society, pp 422-432

Wolfson O, Sistla P, Xu B, Zhou J, Chamberlain S (1999) Tracking Moving Objects Using Database Technology in DOMINO. In: Proceedings of the 4th Workshop on Next Generation Information Technologies and Systems (NGITS). Zikhron-Yaakov, Israel. Lecture Notes in Computer Science, no. 1649, Springer-Verlag, pp 112-119

Wolfson O, Xu B, Chamberlain S, Jiang L (1998) Moving Objects Databases: Issues and Solutions. In: Proceedings of the 10th International Conference on Scientific and Statistical Database Management (SSDBM98). IEEE Computer society, pp 111-122