The Tale of (Fusing) Two Uncertainties

Bing Zhang
Northwestern University
2145 Sheridan Road
Evanston, IL, 60201
bing@u.northwestern.edu *

Goce Trajcevski
Northwestern University
2145 Sheridan Road
Evanston, IL, 60201
goce@eecs.northwestern.edu †

ABSTRACT

This work addresses the problem of combining spatio-temporal uncertainties obtained from heterogeneous location sources. Specifically, we take a first step towards formalizing the process of fusing the uncertainty of moving objects locations obtained from on-board GPS devices and roadside sensors. We develop a model for combining the values from the different sources and analyze the impact of the model on the basic spatio-temporal queries pertaining to object's whereabouts in time. As it turns out, combining the two sources can indeed narrow down the possible locations of a given object. Our experiments demonstrate that the proposed method of fusing the uncertainties may eliminate significant amount of the false positives, when compared to using the traditional bead (equivalently, space-time prism) uncertainty models.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Spatial databases and GIS

General Terms

Theory

Keywords

Uncertain Trajectory, Uncertainty Fusion, Roadside Sensors, Space-Time Prism

1. INTRODUCTION

Miniaturization of computing and sensing devices and advances in networking and communications provided a technological foundation for generating huge volumes of location-in-time data – order of Peta-Bytes per year just from smartphones [13]. Geographic Information Systems (GIS) [27] and many applications relying on Location Based Services (LBS) [25] rely on efficient techniques for storage, retrieval and query processing for such data – topics studied in the fields of spatio-temporal databases [15] and Moving Objects Databases (MOD) [8].

An important feature of the location data in realistic settings is that due to the inherent imprecision of the sensing devices, typically there is a degree of *uncertainty* associated with the measurements/values. The problem of capturing the uncertainty into the data-models [16, 18, 19, 24] as well as queries' syntax and processing algorithms [5, 7, 22, 30, 29] has been recognized and tackled by several earlier works[1].

Complementary to these efforts in spatio-temporal databases and MOD, where the location data is (assumed to be) obtained by an on-board Global Positioning System (GPS) device, in many traffic management applications [1] the location data is obtained from some types of road-side sensors. For example, lane level positioning is an important component in navigation systems widely applied in smart traffic control, automated vehicle location or intelligent transportation systems [6, 28, 26]. Such sensors-data is combined with data from different sensing devices on-board vehicles – e.g., U.S. Xpress gathers 900 to 970 data elements of various engine/component readings [20], used planning loading, routing and servicing regimes of its trucks fleet.

At the heart of the motivation for this work is the observation that combining the uncertain data from two different (heterogeneous) sources – GPS and road-side sensors – may yield more precise answers to certain spatio-temporal queries. For example, if a trucking company is interested in the quality of steering equipment and axles under particular load, it may be interested in queries such as:

Q1: *Retrieve all the vehicles which have crossed the lane in road segment RS1 when driving less than 50km/h and carrying less than 80% of the maximum load.*

Clearly, given the imprecision of the location measurements, **Q1** needs to be re-phrased so that it incorporates uncertainty:

Q1^u: *Retrieve all the vehicles which have had* $> \Theta$ *(*$0 < \Theta \leq 1$*) probability of crossing the lane in road segment RS1*

*Research Supported by the NSF III 1213038 grant
†Research Supported by the NSF grants CNS 0910952 and III 1213038, and ONR grant N00014-14-10215

[1]See [3] for comprehensive list of references

when driving less than 50km/h and carrying less than 80% of the maximum load.

We argue that properly considering the joint impact of – equivalently *fusing* – the uncertainties from the GPS sources and road-side sensors can eliminate some of the moving objects (trajectories) from the answer-set of $\mathbf{Q1}^u$. In other words, what may have been considered an answer under the single (e.g., GPS) source, may become a false-positive after fusing the two location uncertainties. In summary, the main contributions of this work are:

- We present a novel model of spatio-temporal uncertainty for moving objects, which combines the location data obtained by GPS devices on-board moving objects and the location data obtained from road-side sensors.
- We discuss the semantic implications of the model, in terms of the basic *where_at* and *when_at* location-in-time (whereabouts) queries, as well as *lane-crossing* queries (exemplified by $\mathbf{Q1}^u$ above) and basic *range* queries.
- We present experimental observations which quantify the benefits of fusing the two uncertainties for lane-crossing and range queries in terms of the percentage of trajectories which are pruned from the answer-sets when compared to using the traditional bead-model of uncertainty for GPS-based location data.

The rest of this paper is structured as follows. In Section 2 we recollect some backgrounds in terms of modeling spatio-temporal uncertainty, and introduce the basic terminology used in the rest of the work. Section 3 presents the details of the new uncertainty model, along with the semantics of the basic whereabouts queries along with lane-crossing and range queries. Section 4 describes our experimental observations. In Section 5 we compare our work with related literature, and we summarize and outline directions for future work in Section 6.

2. PRELIMINARIES

We now overview the techniques for location data relevant for this work, both GPS-based and the ones based on road-side sensors. Subsequently, we proceed with introducing the basic notation used in the rest of the paper.

2.1 Road-side Sensors

Starting in the 1920s, when the traffic signals were still manually controlled, several generations of sensor types have been developed and deployed for traffic management – from pressure-sensitive sensor in 1931 to modern laser sensors [2].

Contrary to the GPS-based data acquisition techniques where each data source is isolated, the roadside sensors are usually connected hierarchically to a server and send their sampled data to traffic control center [14]. Compared with GPS system, the roadside sensors have better measurement accuracy, higher sampling frequency and shorter response time, which enables their use in real time traffic information analysis and control.

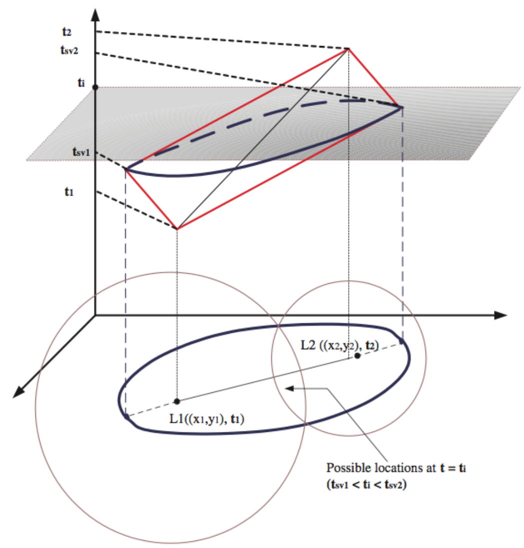


Figure 1: Bead and Ellipse Model

Several types of roadside sensors have been commercialized and deployed on roads. For example, the AMR sensor[11] developed by Honeywell is a type of magnetic sensor with low cost. The WiEye[21] is a passive infrared sensor that can be installed on the motes to sense road condition. The variation of sensing technologies implies different methodologies for modeling of motion in order to capitalize on a particular type of sensors. In this paper, the data model for roadside sensor that we adopt is based on TruSense T-Series, manufactured by Laser Technology Inc.[12] – a kind of active infrared sensor with very high accuracy and repetition/sampling rate.

Table 1 provides a summary of features of several different types of sensors[2]. As shown, all of the popular and commercially available types can detect the presence and speed of vehicles, as well as provide a count value for the number of vehicles that have been detected in their sensing range. However, very few types provide more detailed sensing capabilities, such as classification and multiple lanes detection. In this work, we focus on detection of a presence of a moving object in the sensing range.

2.2 GPS-based Spatio-Temporal Uncertainty

One of the basic approaches for modeling spatio-temporal uncertainty of moving objects is the, so called, sheared cylinder model. The main assumption of the model is that at any time instant t_i, the object's location is inside a given disk with a fixed radius, centered at the expected location at t_i. For time values different from sampling ones, the expected location is obtained via linear interpolation [30]. However, this model is geared towards past/historic trajectories.

The time-geography [9] ideas from the 1970s (and more recently probabilistic time geography [31]) have also permeated MOD research. The implications of the fact that the object's motion was bound by some v_{max} *in-between* two updates was analyzed in [23]. Based on the definition as a geometric set of 2D points, it was demonstrated that the ob-

Table 1: Comparison among different types of sensor

Sensor technology	Count	Presence	Speed	Output Data	Classification	Multiple Lane detection
`Inductive loop`	✓	✓	✓	✓	✓	✗
`Magnetometer (two axis fluxgate)`	✓	✓	✓	✓	✗	✗
`Magnetic induction coil`	✓	✓	✓	✓	✗	✗
`Microwave radar`	✓	✓	✓	✓	✓	✓
`Active infrared`	✓	✓	✓	✓	✓	✓
`Passive infrared`	✓	✓	✓	✓	✗	✗
`Ultrasonic`	✓	✓	✗	✓	✗	✗
`Acoustic array`	✓	✓	✓	✓	✗	✓
`Video image processor`	✓	✓	✓	✓	✓	✓

jects possible whereabouts are bound by an ellipse, with foci at the respective point-locations of the consecutive samples (i.e., samples at consecutive time instants). Subsequently, [10], presented a spatio-temporal version of the model, naming the volume in-between two update points a *bead*[2], and the entire uncertain trajectory, a *necklace*. However, the first works to present a formal analysis of the properties are [17, 18]. An illustration is provided in Figure 1. Letting $d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$ denote the distance between the starting location (at t_1) and ending location (at t_2), the equation of the projected ellipse (cf. [23]) is:

$$\frac{(2x - x_1 - x_2)^2}{v_{max}^2 (t_2 - t_1)^2} + \frac{(2y - y_1 - y_2)^2}{v_{max}^2 (t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2} = 1$$

The corresponding space-time prism is specified with the following constraints:

$$\begin{cases} t_i \le t \le t_{i+1} \\ (x - x_i)^2 + (y - y_i)^2 \le [(t - t_i) v_{max}^i]^2 \\ (x - x_{i+1})^2 + (y - y_{i+1})^2 \le [(t_{i+1} - t) v_{max}^i]^2 \end{cases} \qquad (1)$$

where v_{max} is the maximal speed that the object can take between t_i and t_{i+1}. We note that, what is commonly called *expected* speed in the case of crisp trajectories, now becomes *minimal* expected speed in-between the updates/samples. As shown in Figure 1, at any time instant t between two consecutive samples, the possible locations of the objects are bound by the lens – i.e., intersection of two circles centered at the respective foci and with respective radii $v_{max}(t - t_1)$ and $v_{max}(t_2 - t)$.

If the objects are constrained to move along a road network, then the space-time prisms are restricted in their size. Specifically, if the segments of the road network are assumed to be edges in a graph, then the prisms become restricted to planar figures [7, 16].

2.3 Trajectories and Road Networks

[2]More recently, also called *space-time prism* as used in time-geography.

Throughout this paper, we consider the following definition of a trajectory:

Definition 1. A *trajectory* Tr_i of a moving object with a unique identifier (oID) "i", is a sequence of triplets $Tr = [(L_1, t_1), (L_2, t_2), v_{max1}] \ldots, [(L_{n-1}, t_{n-1}), (L_n, t_n), v_{max\_(n-1)}]$ where each ($L_i = (x_i, y_i)$ is a point in 2D space in a corresponding reference coordinate system, and t_i denotes the time instant at which the object was at location L_i. When it comes to the time-values, $i < j$ implies $t_i < t_j$, and v_{max_i} denotes the maximum speed of the object between samples at t_i and t_{i+1}

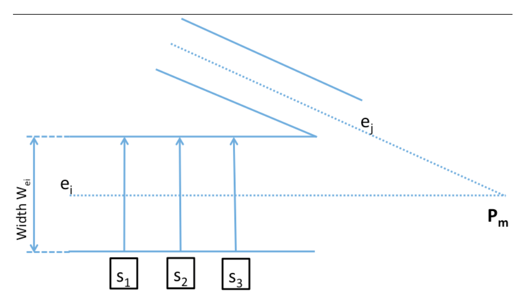


Figure 2: Road Segments and Sensors

We define a road network as an *augmented graph* $G = (P, E_{RS})$ where $P = \{p_1, p_2, \ldots, p_n\}$ denotes a set of points (commonly corresponding to intersections) and $E_{RS} = \{r_{S1}, ..., r_{Sk}\}$ is a collection of triplets of the form $r_{Si} = (e_i, w_{ei}, v_{ei})$ where:

- $e_i = (p_{i1}, p_{i2})$ ($\in P$ X P) is a "regular edge" (i.e., a link between two connected vertices)
- w_{ei} denotes the width of the road segment associated with the edge e_i.
- v_{ei} denotes the maximum speed associated with r_{Si}.

Unless otherwise specified, we will assume that the maximum speed of a given object in-between two consecutive location samples along a particular road segment corresponds to the maximum speed of that segment. We note that, geometrically speaking, the collection all the r_{Si}'s can be obtained as the boundary of the Minkowski sum of each "regular edge" e_i and a disk with diameter w_{ei}.

Lastly, we also assume the existence of a collection of sensors $S = \{s_1, s_2, ..., s_m\}$, where each sensor s_j is located at a point along the outer boundary of some road segment r_{Si}. Each s_j detects when (i.e., the time instant at which) a moving object crosses the line segment going through its location and perpendicular to e_i. The concepts are illustrated in Figure 2.

3. FUSING HETEROGENEOUS LOCATION UNCERTAINTIES

We now introduce the new uncertainty model resulting from combining the GPS-based location data and the location data generated by road-side sensors. We follow with a discussion of the semantics of the basic whereabouts queries, lane-crossing and range queries.

3.1 Combined Model

When combining the location samples, the main observation is that the data obtained from the road-side sensors provides additional constraints on the possible whereabouts in-between two consecutive GPS-based samples (and vice-versa). More specifically, in addition to the system of inequalities (1) specifying the space-time prism (i.e., bead), we now have the constraint that at a particular time instant t_{si}, the locations of the objects are known to also be along a given line-segment determined by: (1) the location of the corresponding road-side sensor; and (2) the direction which is perpendicular to the (boundaries of the) road segment. This can be formalized as:

$$\begin{cases} t_i \leqslant t \leqslant t_{i+1}, \\ (x - x_i)^2 + (y - y_i)^2 \leqslant (t - t_i)^2 v_{max}^2, \\ (x - x_{i+1})^2 + (y - y_{i+1})^2 \leqslant (t_{i+1} - t)^2 v_{max}^2, \\ y = m_i x + b_i, \text{when } t = t_{si} \\ t_i \leqslant t_{si} \leqslant t_{i+1}. \end{cases} \quad (2)$$

The system of constraints (2) is illustrated in Figure 3. Specifically, as shown in Figure 3a, the original GPS-based locations L_1 and L_2 would yield a 2D projection which is an ellipse having them as foci (ligth-grey shade in Figure reffusing1-1) – denote it El_1. However, because of the road-side sensor, we know that the possible locations of the moving object at t_{s1} can only be along the *portion* of line segment originating in $(x_{s1,s1})$, perpendicular to the boundaries of the road segment, and intersecting El_1 – i.e., along the portion of the line segment $\overline{L_1' L_1''}$. Clearly, that intersection has an uncountably many points, and we show 3 such points in Figre 3a – L_{11}, L_{12} and L_{13}. The main observation is that each such point, in turn, can be used as a "generator" for two more space-time prisms: one originating in L_1, and the other terminating at L_2. The corresponding 2D projections (ellipses) are shown in Figure 3a for L_{11}, L_{12} and L_{13}. The most important implication is that when combining (i.e., taking the intersection of) the original ellipse El_1 with the uncountably infinite collection of the ellipses with one of the foci along the line segment due to the road-side sensors, the additional constraint induces a significant amount of a "dead-space" in El_1. A more detailed illustration of the valid range for selecting the points that will generate the infinite collection of (pairs of) new beads is given in Figure 3b. Recall that (cf. Section 2), at any given time instant t_{s1} between the sampling times t_1 and t_2, the object can be located inside of the lens obtained as the intersection of the circles with radii $v_{max}(t_{s1} - t_1)$ and $v_{max}(t_2 - t_{s1})$. Hence, although the ray emanating from the roadside sensor s_1 would intersect the "global boundary" (i.e., the ellipse which is the projection of the bead) at L_1' and L_1'', the only valid points to be considered as possible whereabouts are the ones along (and inside) the lens. As shown in Figure 3b, those are the points along the line segment bounded by L_{11} and L_{13}.

We note that there is the "flip-side" context of having a single uncertainty source. Namely, if we only had the roadside sensors available, then, in between two detections by consecutive sensors (say, s_1 and s_2 from Figure 2), the whereabouts of a given object in-between the two sampling time instants t_{s1} and t_{s2} is bounded by the union $\cup(El_{si,sj})$ of uncountably many ellipses for which:

1. The first focus is some point L_{s1} located on the line-segment originating at the location of s_1.
2. The second focus is some point L_{s2} located on the line-segment originating at the location of s_2;
3. The distance between L_{s1} and L_{s2} is smaller than $v_{max}(t_{s2} - t_{s2})$ (i.e., the object could travel the distance within the time-interval $[t_{s1}, t_{s2}]$ for the given speed limit).

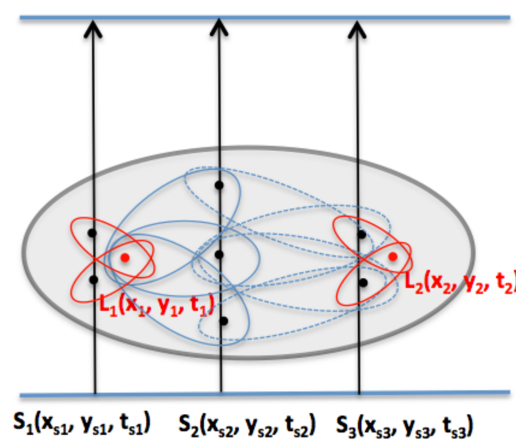
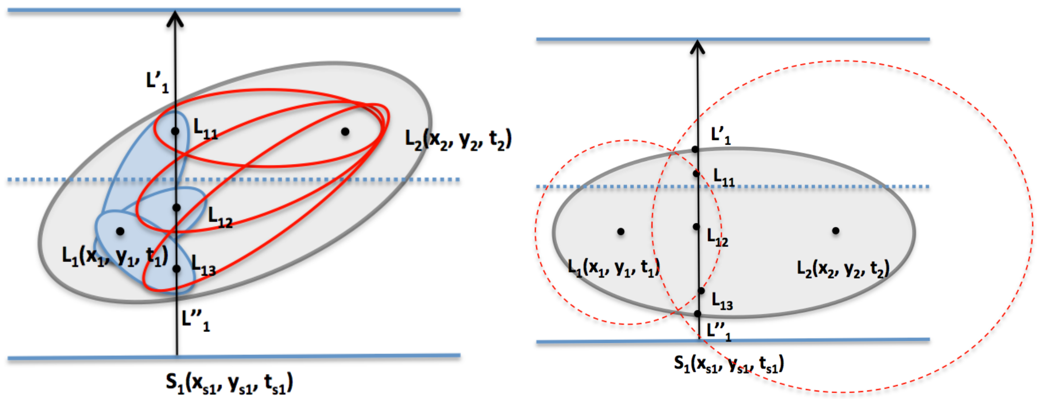


Figure 4: Multiple Roadside Sensors Intersecting a Bead

Incorporating the GPS-based bead in this context would either amount to the case where it intersects one (or more) of the line segments originating at the respective sensors locations, or it has no intersection with any of them. In the latter case, we have a scenario in which GPS sampling frequency is higher than the sampling frequency obtained by the road-side sensors. For such settings, the possible whereabouts will be reduced to the intersection of the $\cup(El_{si,sj})$ and the bead obtained from the GPS-based samples. In the former case, the model is a generalization of the one corresponding to the scenario illustrated in Figure 3 – in the sense that it may be possible to have intersections of the GPS-based



(a) GPS + Roadside Sensors (b) Determining Boundaries

Figure 3: Fusing GPS and Roadside Sensors Data

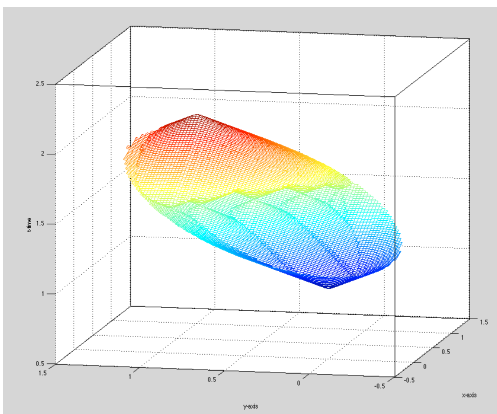


Figure 5: Outer Boundary of the Fused Uncertain Locations

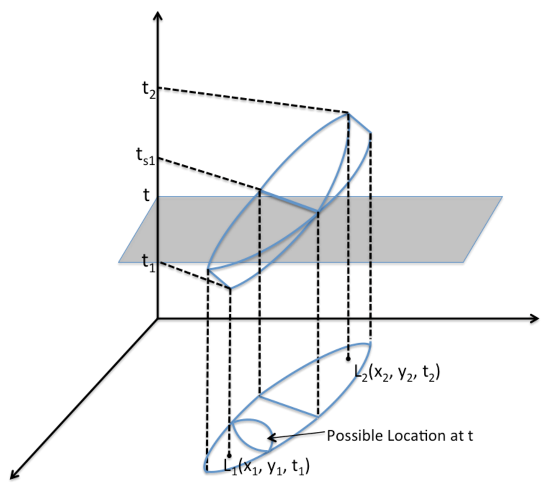


Figure 6: Whereabouts at Time Instant

bead with > 1 sensor lines, as illustrated in Figure 4. In the rest of this paper, we focus on detailed discussion of the scenarios in which a bead is intersected by a line segment from a single roadside sensor.

The spatio-temporal structure induced by combining the two uncertainty sources – GPS and roadside sensors – is called a *Fused Bead* (FB), and it is a sixtuple FB ((x_i, y_i, t_i), $(x_{i+1}, y_{i+1}, t_{i+1})$, $v_{max}, t_s, m, b)$ consisting of:

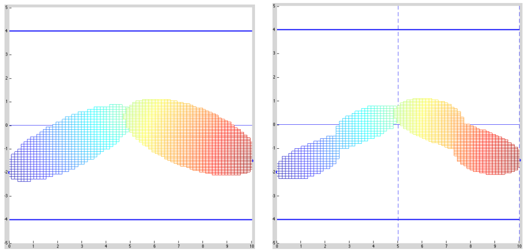
- The 2 GPS-based location-in-time samples and along with the v_{max} speed bound.
- The time instant of detection of the road-side sensor.
- The parameters of the equation of the line specifying the corresponding line-segment of the possible new foci.

When it comes to bounding the possible whereabouts, an additional observation is in order. Namely, some of the points along the intersection of the line segment with the ellipse El_1 may yield possible focal points that would generate ellipses which are not fully contained inside El_1. An example of such extreme-case scenario is when the time (resp. location) of the roadside sensor intersecting the ellipse from the bead is equal to $(t_1 + t_2)/2$, where t_1 and t_2 are the time instants in which the GPS-based locations were taken – equivalently, to foci of the projection of the bead, El_1. However, the set of constraints in (2) will eliminate every portion which is outside the intersection of the original El_1. Hence, in some sense, the original space-time prism obtained from the GPS samples, is an outer-boundary of the volume (2D+Time) of the objects possible (*location, time*) values – as illustrated in Figure 5

3.2 Basic Queries

The first query that we consider pertains to obtaining the possible whereabouts of the object at a given time instant – i.e., *where_at(oID, t)* query. Recall that for the bead obtained by GPS-based samples, one could determine the possible whereabouts of the moving object at time t by intersecting the corresponding bead with a horizontal plane at t (cf. Figure 1) – i.e., an intersection of two circles centered at L_1 and L_2 with the radii corresponding to $v_{max}(t - t_1)$ and $v_{max}(t_2 - t)$.

Similarly to the GPS-based bead, in order to determine the



(a) Bead Model (b) Fused Bead Model

Figure 7: MATLAB Visualization

whereabouts at a given time instant t for a fused bead, we need to obtain the intersection of FB with the horizontal plane $Tme = t$. The corresponding illustration of the volume in 2D space + Time, along with the 2D projection, is shown in Figure 6. We note that the boundary of the 2D projection is obtained as the "envelope" of the union of two collections of uncountably many infinite pairs of arcs. Each pair of arcs represents the boundaries of the intersections of the corresponding pairs of disks – one centered at the focus of the GPS-based bead (e.g., L_1) and the other centered at a point along the intersection chord (exemplified by $\overline{L_{11}L_{13}}$ in Figure 3) resulting from secant due to he roadside sensor and the arc from the lens of the original GPS-based bead. Thus, one of the boundaries is always a circular arc originating at the focal point of the "original" GPS-based bead, centered at focus of the GPS-based bead (say, L_1) and with radius $v_{max}(t - t_1)$. The other part of the boundary is actually the boundary of the union of uncountably many disks with radii $v_{max}(t_{s1} - t)$, and with centers along the intersection-chord.

The complementary query, *when_at(oID, L)* returns the times during which it is possible for the object *oID* to be at the location $L(x_L, y_L)$, i.e., a time-interval $[t_L^1, t_L^2]$. The time-interval can be defined as the two intersections between the boundary of the fused bead FB and the vertical line (i.e., ray) emanating from L. To calculate the values, we have the following observations:

1. t_L^2 is the latest time that a circle located at the GPS-based focus from the sample at t_1 will "reach" L – hence, it can be obtained as a solution to the equation: $\overline{L_1L} = v_{max}(t_L^2 - t_1)$
2. t_L^1, on the other hand, is the earliest time that any circle with the center on the intersection chord and radius $v_{max}(t_s - t_L^1)$ would pass through L.

Assuming uniform *pdf*s of the possible objects locations within the uncertainty zone defined by the FB model for a given time instant, we now discuss the lane-crossing and range query. Without loss of generality, we will consider an input consisting of a a single fused bead $FB((x_i, y_i, t_i), (x_{i+1}, y_{i+1}, t_{i+1}), v_{max}, t_s, m, b)$ and a region R_q.

The lane-crossing query is a minor variation of $\mathbf{Q1}^u$ (cf. Section 1):

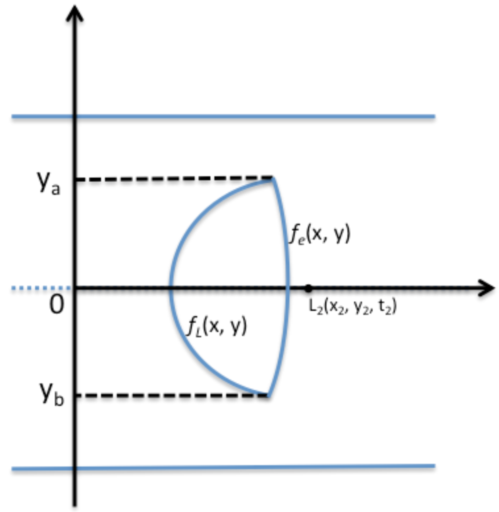


Figure 8: Evaluating lane-crossing query at t_i

$\mathbf{Q}_{lc}^u$: *Retrieve all the vehicles which have* $> \Theta$ $(0 < \Theta \leq 1)$ *probability of crossing the lane in road segment RS1.*

Let C_t denote the planar region corresponding to the answer of the *where_at(oID, t)* and let f_L and f_e denote the two curves defining the boundary of C_t. Also, let (x_a, y_a) and (x_b, y_b) denote the intersection points between f_L and f_e (i.e., the cusps of the boundary of C_t). To calculate the probability that the object oID is crossing the lane at time instant t, one needs to calculate the area of C_t $(A(C_t))$ and the area of the portion of C_t on "the other side" of the lane. While in some special cases – e.g., when the GPS-based location samples are along the line parallel to the lane-separator line (cf. Figure 8) and both are axis-parallel – it may be possible to have closed-form formula, we note that, in general one would need to rely on numerical integration.

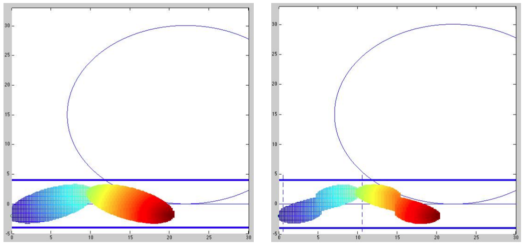
When it comes to the range query, the methodologies applied for the lane-crossing query would require a minor modification in order to cater to the boundary of the region of interest for a given query (e.g., polygon, circle, etc...).

4. EXPERIMENTAL OBSERVATIONS

In order to get quantitative evaluation of the proposed model, we examined how many answers obtained when using the GPS-based bead model actually become false positive when the FB model is employed. Towards that, we used a MATLAB implementation of the numerical integration[3] for evaluating the probabilities of an object satisfying the lane-crossing query and range query for a simple case of a disk.

In the first settings, we evaluated the range query for a simple trajectory to detect how much the FB model reduces the location whereabouts. The setup for the experiment is shown in Figure 9. The GPS-based bead model returns

[3]We discretize both temporal and spatial axis and use numerical method to approximate the areas corresponding to the respective probabilities. The source code(s) and the data are publicly available at http://www.eecs.northwestern.edu/∼bvz686/FusedBeads



(a) GPS-based Bead Model (b) Fused Bead Model

Figure 9: Range Query

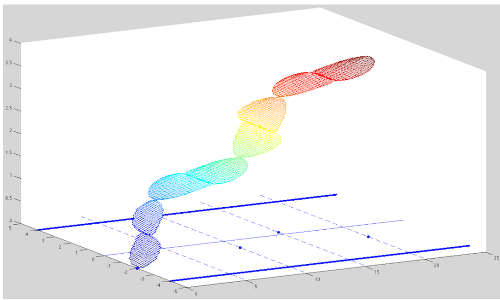


Figure 10: Lane-Crossing Visualization

true for the second bead(9a). However, the FB model returns false for the same trajectory(9b), due to the additional constraints provided by the roadside sensor data. The intersection between "global boundary" and query range for FB model is much smaller than that for GPS-based model.

In the second experimental settings, we investigated the impact of FB model on lane-crossing query in road network. To increase the number of possible lane-crossing instances, we simulated a vehicle moving along a road (cf. Figure 2) and having multiple crossings of the lane. The vehicle's motion has two direction-components: one parallel to the boundary of the road (i.e., lane), denoted as M_x, and the other one perpendicular to it, which is denoted as M_y. The width of each lane of the road segment was set to 4m. As a reference coordinate system, the central lane was set as x-axis so the full range for M_y is [-4m, 4m]. We considered densely depolyed sensors – located at every 10m along the road. Vehicle's GPS positions were also sampled every 1s and the movement along M_x was set to a constant speed, less than 50km/h. Vehicle's perpendicular movement M_y is generated by a random generator with uniform distribution given a movement interval. We generated two data sets. In the first one DS_1, M_y has a [-50%, 12.5%] movement interval, which means it will be uniformly distributed in the interval [-2m, 0.5m] – which implies that DS_1 contains more instances of boundary conditions. In data second dataset, DS_2, M_y has a full range of motion values – [-100%, 100%].

For each data set, we perform experiments with different trajectory length – 1km, 5km and 10km, on both GPS-based bead model and FB model. During the experiment, the series of FB is formed in spatial-temporal coordinate. Figure 10 is a visualization of our experimental setup, representing a snapshot of multiple FB on road network.

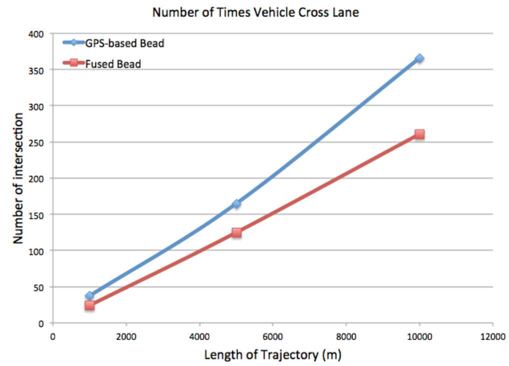


Figure 11: Lane-Crossing—Data Set One

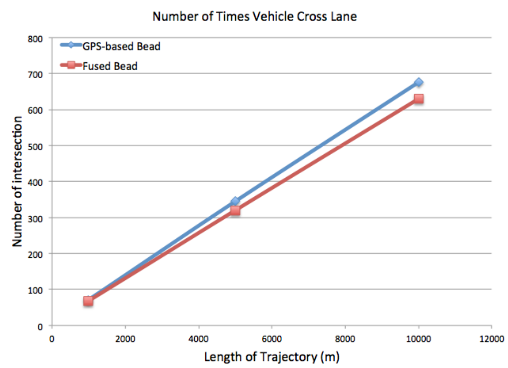


Figure 12: Lane-Crossing—Data Set Two

Figure 11 shows the benefit of FB. The number of lane-crossing incidents is reduced by around 30% by using FB model. The reductions are the result of correctly classification for those boundary scenarios that would misclassified as false positive by GPS-based bead model.

The result of a vehicle allowing full random movement perpendicular to the road is shown in Figure 12. Even though the number of boundary situations is less than the data set one, we could still see a reduction of intersections.

These boundary scenarios in data set two corresponding to the real world situation when a car's trajectory is slightly deflected from the center of the lane and quickly return back, mainly because of driving under the influence. The FB model will be able to differentiate these activities from lane-crossing.

In our second experiment, FB model is applied on range query.

We try to answer the following query:

$\mathbf{Q2}_u^r$: *Retrieve all the vehicles which have* $> \Theta$ $(0 < \Theta \leq 1)$ *probability of going through the range* R_c.

The vehicle's movement set up is similar to lane-crossing query and M_y range of values was set to be [-62.5%, 62.5%]. The query region R_c was set to be a circle with a 15m radius as shown in Figure 13.

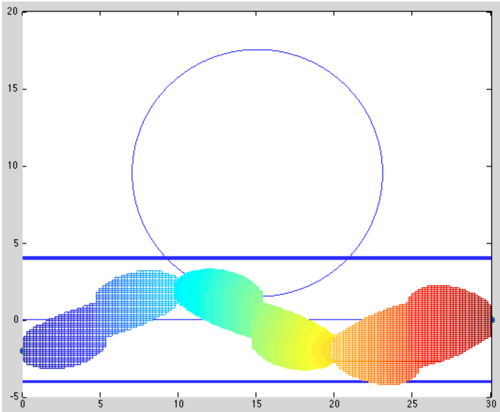


Figure 13: Range Query Setting

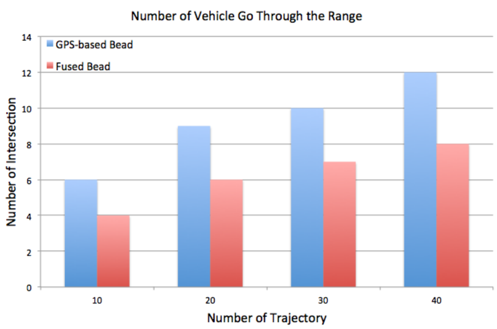


Figure 14: Range Query Result

In sequence, we let 10, 20, 30 and 40 vehicles move along the road, and the experimental results are shown in Figure 14, demonstrating that fewer "possible trajectories" would satisfy the range query under the FB model.

5. RELATED WORK

There are two main bodies of research literature that are related to, and were used as foundation for, our work.

The first one consists of results from GIS, MOD and spatio-temporal databases communities, where the problem of capturing the uncertainty of motion has been studied extensively. Starting with [9], and more recently [31], the issue of uncertain whereabouts from the perspective of probabilistic time geography has been tackled by a model of emanating cones-in-time, with a vertex at the last location sample. The 2D boundary of the possible locations of moving objects with bounded speed was formalized by an ellipse in [23], and its 2D+time version – beads – was presented in [10]. Subsequently, [17, 18] provided a full formalization of the beads model and also provided extensions to capture the impact of road networks [16]. Majority of the works dealing with uncertainty (either in free-space motion or road networks constrained) from MOD and spatio-temporal databases community have focused on efficient processing of the popular spatio-temporal queries (range, (k)NN, reverse-NN) under various models of uncertainty [8, 3].

Unlike majority of the works so far, in this paper we incorporated an additional source of location data – the roadside sensors, and considered the road network which has a width as a parameter, instead of simple edges.

The second body of works originates in the transporation and traffic management communities. Substantial efforts have been made to tackle the lane-crossing query and several works have focused on building novel system to overcome the shortcoming of single GPS receivers which yields unstable measurements with large uncertainty [4, 6]. Attempts have been made to acquire location data using commercially available smartphones [26], however, nearly 50% of the data failed to fall within the road network region. Other efforts include the use of integrated sensor like gyroscope to fill the unknown values between two GPS sample updates [28].

However, the works did not consider the uncertainty in-between consecutive GPS-based updates and sensor-based location detections.

Some of the works [6, 28], use map matching algorithms to determine which lane the vehicle belongs to and, subsequently, try to revise the measurement error using post-processing. However, the bead (or, space-time prism) model has not been exploited.

6. CONCLUDING REMARKS

We addressed the problem of combining the uncertain location data from two different sources: GPS on-board moving objects and roadside sensors. We proposed a formal model – fused bead – for the possible locations at given time-instant(s) when the location data from both sources is combined, and demonstrated that "two uncertainties are better than one", in the sense that fusing the data from both sources would narrow the possible whereabouts when compared to individual location data source. We analyzed the impact of the model on the basic spatio-temporal queries and we presented experimental observations illustrating the benefits of the fused bead approach.

There are a few directions that we plan to pursue in the near future. Firstly, we would like to work on generating efficient algorithms for *continuous* versions of the spatio-temporal queries for which we discussed the instantaneous variants here – and, of course, to extend the results to other popular queries such as (reverse) Nearest Neighbor. Our second avenue is to investigate the scalability and efficiency aspects of the query processing algorithms – for instance, as we observed in Section 3, one could rely on the "regular" space-time prisms for prunning, since the fused beads are always bound by the "regular" bead. A complementary objective is to extend the model/formalism so that it captures the uncertainty/imprecision in the very samples [24], not only the intermediate whereabouts, and to include other kinds of semantic information (e.g., type of a vehicle, size, etc.).

7. REFERENCES

[1] Southwest research institute, advanced traffic management systems. http://www.swri.org/4ORG/d10/its/atms/default.htm.

[2] T.-F. H. R. Center. *Traffic Detector Handbook*. U.S. Department of transportation, McLean, VA, third edition, volume i edition, 2006.

[3] R. Cheng, T. Emrich, H.-P. Kriegel, N. Mamoulis, M. Renz, G. Trajcevski, and A. Züfle. Managing uncertainty in spatial and spatio-temporal data. In *ICDE*, pages 1302–1305, 2014.

[4] T.-S. Dao, K. Y. Leung, C. M. Clark, and J. P. Huissoon. Markov-based lane positioning using intervehicle communication. *Trans. Intell. Transport. Sys.*, 8(4):641–650, Dec. 2007.

[5] Z. Ding. Utr-tree: An index structure for the full uncertain trajectories of network-constrained moving objects. In *MDM*, pages 33–40, 2008.

[6] J. Du and M. Barth. Next-generation automated vehicle location systems: Positioning at the lane level. *Intelligent Transportation Systems, IEEE Transactions on*, 9(1):48–57, March 2008.

[7] T. Emrich, H.-P. Kriegel, N. Mamoulis, M. Renz, and A. Züfle. Querying uncertain spatio-temporal data. In *ICDE*, pages 354–365, 2012.

[8] R. H. Güting and M. Schneider. *Moving Objects Databases*. Morgan Kaufmann, 2005.

[9] T. Hägerstrand. What about people in regional science? *Papers of the Regional Science Association*, 24:7–21, 1970.

[10] K. Hornsby and M. J. Egenhofer. Modeling moving objects over multiple granularities. *Ann. Math. Artif. Intell.*, 36(1-2):177–194, 2002.

[11] H. I. Inc. Vehicle detection using amr sensors. Technical report, Defense and Space Electronics Systems, 12001 Highway 55, Plymouth, MN 55441, 2005.

[12] L. T. Inc. Trusense t-series, 2013. Online; accessed May 22, 2014.

[13] M. K. G. Institute. Big data: The next frontier for innovation, competition, and productivity, 2011.

[14] J. Jang, H. Kim, and H. Cho. Smart roadside server for driver assistance and safety warning: Framework and applications. In *CUTE 2010*, pages 1–5, Dec 2010.

[15] M. Koubarakis, T. Sellis, A. Frank, S. Grumbach, R. Güting, C. Jensen, N. Lorentzos, Y. Manolopoulos, E. Nardelli, B. Pernici, H.-J. Scheck, M. Scholl, B. Theodoulidis, and N. Tryfona, editors. *Spatio-Temporal Databases – the CHOROCHRONOS Approach*. Springer-Verlag, 2003.

[16] B. Kuijpers, H. J. Miller, T. Neutens, and W. Othman. Anchor uncertainty and space-time prisms on road networks. *International Journal of Geographical Information Science*, 24(8):1223–1248, 2010.

[17] B. Kuijpers and W. Othman. Modelling uncertainty on road networks via space-time prisms. *Int.l Journal on GIS*, 23(9), 2009.

[18] B. Kuijpers and W. Othman. Trajectory databases: data models, uncertainty and complete query languages. *Journal of Computer and System Sciences*, 2009. doi:10.1016/j.jcss.2009.10.002.

[19] R. Lange, H. Weinschrott, L. Geiger, A. Blessing, F. Dürr, K. Rothermel, and H. Schütze. On a generic uncertainty model for position information. In *QuaCon*, pages 76–87, 2009.

[20] T. Leonard. Delivering deeper insights with big data and real-time analytics, 2012.

[21] E. LLC. *Wieye - sensor board for wireless surveillance applications*. 401 North Coquillard Dr., South Bend, IN 46617, 2008.

[22] J. Niedermayer, A. Züfle, T. Emrich, M. Renz, N. Mamoulis, L. Chen, and H.-P. Kriegel. Probabilistic nearest neighbor queries on uncertain moving object trajectories. *PVLDB*, 7(3):205–216, 2013.

[23] D. Pfoser and C. S. Jensen. Capturing the uncertainty of moving objects representation. In *SSD*, 1999.

[24] D. Pfoser, N. Tryfona, and C. S. Jensen. Indeterminacy and spatiotemporal data: Basic definitions and case study. *GeoInformatica*, 9(3), 2005.

[25] J. Schiller and A. V. (editors). *Location-Based Services*. Morgan Kaufmann, 2004.

[26] Y. Sekimoto, Y. Matsubayashi, H. Yamada, R. Imai, T. Usui, and H. Kanasugi. Lightweight lane positioning of vehicles using a smartphone gps by monitoring the distance from the center line. In *ITSC 2012*, pages 1561–1565, Sept 2012.

[27] S. Shekhar and H. Xiong, editors. *Encyclopedia of GIS*. Springer, 2008.

[28] R. Toledo-Moreo, D. Betaille, and F. Peyret. Lane-level integrity provision for navigation and map matching with gnss, dead reckoning, and enhanced maps. *Intelligent Transportation Systems, IEEE Transactions on*, 11(1):100–112, March 2010.

[29] G. Trajcevski, R. Tamassia, I. Cruz, P. Scheuermann, D. Hartglass, and C. Zamierowski. Ranking continuous nearest neighbors for uncertain trajectories. *VLDB J.*, 20(5):767–791, 2011.

[30] G. Trajcevski, O. Wolfson, K. Hinrichs, and S. Chamberlain. Managing uncertainty in moving objects databases. *ACM Trans. Database Syst.*, 29(3), 2004.

[31] S. Winter and Z.-C. Yin. Directed movements in probabilistic time geography. *International Journal of Geographical Information Science*, 24(9), 2010.