Visualizing uncertainty in spatio-temporal data

Ayush Shrestha
Department of Computer Science
Georgia State University
Atlanta, GA
ashrestha2@cs.gsu.edu

Ying Zhu
Department of Computer Science
Georgia State University
Atlanta, GA
yzhu@cs.gsu.edu

Ben Miller
Department of Communication
Georgia State University
Atlanta, GA
miller@gsu.edu

ABSTRACT

Analyzing the relationship between location and time in a spatio-temporal data is not trivial. It is even more challenging if the data contains uncertainty. In this paper, we present a new method that visualizes spatio-temporal data with uncertainty. This method is an extension of our 2D visualization technique called Storygraph, and it handles two types of data uncertainty: (1) the spatial and temporal uncertainty about an event; (2) the spatial and temporal uncertainty between two events. We applied this method to a case study that involves data extracted from witness testimonies and field reports containing uncertainties inherent to natural language.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Evaluation/methodology*

1. INTRODUCTION

The introduction of geo-location sensors in mobile devices and other commodity hardware has greatly aided in spatio-temporal data collection. As a result, novel and effective methods are needed to help analyze these great amounts of spatio-temporal data. Traditional methods like maps fail to show the temporal sequence of the *events*. An event in this paper refers to a row in the dataset having distinct time and location. If two events occur at the same location at different times, the markers will overlap, resulting in a single marker. Time series charts are helpful for presenting temporal information but difficult for analyzing spatial information. Other methods such as small multiples, animations, and 3D maps have significant drawbacks.

In our previous work, we introduced a technique called Storygraph [1] to address these issues. Storygraph is a 2D technique that visualizes both spatial and temporal components in an integrated graph. Our case studies demonstrated the benefits of this method on datasets containing precise geolocations and time such as military war logs [1] and software commit histories [2]. However, when applying our method to spatio-temporal data extracted from witness testimonies and field reports, we encountered problems of uncertainty in space and time. For example, our study of 511 interviews with first responders during the attack on World Trade Center (WTC) on September 11, 2001 showed that the narratives of these interviewees, who were trained to report incidences, still contained a fair amount of uncertainty in their descriptions of locations and times.

To address these issues, we developed a new version of Storygraph visualize uncertainty. In our revision, we begin by categorizing uncertainty into two categories: (1) event uncertainty and (2) between-event uncertainty. We designed our method to distinguish and visualize these two types of uncertainty. Event uncertainty is the spatio-temporal uncertainty about the event itself, including events with poorly specified spatial and/or temporal attributes. Between-event uncertainty is the uncertainty between two precisely recorded events, which we call them *key events*. This concept is in part influenced by Hagerstrand's Time Geography [3][4][5]. After specifying the key events, the between-event uncertainties are visualized as space-time prisms between the key events. Through this process, our visualization technique can be used to study the interactions between people (or characters) in both space and time.

The rest of the paper is organized as follows: Section 2 discusses related work in spatio-temporal and uncertainty visualization. Section 3 presents the mathematical model of Storygraph. Section 4, describes the classification of uncertainty. Section 5 discusses how uncertain events are visualized in Storygraph. Section 6 discusses how between-event uncertainty is visualized. Section 7 presents a case study featuring fire fighter interviews from WTC corpus. Section 8 concludes by summarizing our work and discussing future works.

2. RELATED WORK

Maps and time series charts are the most common visualization techniques to present spatial and temporal data sets. Other techniques include [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19]. These techniques, however, do not deal with uncertainties in spatial or temporal dimensions even though data collected from real world often contains various levels of uncertainty because of unreliable memory, unreliable source, or the inherent ambiguity of natural language.

Much work has been done in visualizing uncertainty [20]. Here, we focus on closely related work in spatial temporal data visualization. The most common method is to overlay uncertainty information on top of a map. For example, Love et al. [21] used color coding, displacement mapping, and bar glyph on a 3D map to visualize uncertainty. Some authors also used color to visualize probabilities on a 2D map [22][23]. Zuk et al. [24] used transparency, wire frame, or location shift to present uncertainties on 3D models. Some scientific visualization methods deal with location uncertainty by plotting multiple versions of the simulations

Shrestha, Ayush, Ying Zhu, and Ben Miller. "Visualizing uncertainty in spatio-temporal data." In Proceedings of the ACM SIGKDD Workshop on Interactive Data Exploration and Analytics, pp. 117-126. ACM, 2014.

or observations, which creates a spaghetti-like drawing of data points. Other methods use contour lines or sound to indicate uncertainty. However, most of the previous works are about visualizing uncertainty data associated with location and time rather than uncertainty in location and time themselves. For example, color coding, displacement mapping, and bar glyph on a map cannot show the area of possible (but uncertain) locations. Wire frame and transparency indicate the existence of uncertainty but not the possible range of uncertain locations or times. Pebesma et al. [23] used animation to show variability in time; but, with animation, users only see one image at a time, and it's difficult to conduct data analysis on a timeline [25][26]. Most importantly, previous methods have shown difficulty integrating spatial and temporal uncertainty in one view.

The main difference between our method and previous works is that, in our method, uncertainty information is not displayed on a map but on the more abstract Storygraph. The benefit is that it can visualize both spatial and temporal uncertainty in a single 2D view. Our method can clearly differentiate between uncertainty in location (spatial uncertainty), time (temporal uncertainty) as well as a combination of the two (spatio-temporal uncertainty). In other methods, such differences are not clearly distinguishable. Our method also visualizes between-event uncertainty, which is mostly ignored by other methods. Our between-event uncertainty visualization is influenced in part by Hagerstrand's Time Geography [3][4][5], a 3D map based visualization.

3. STORYGRAPH

Storygraph is a visualization technique that presents an integrated 2D view for spatio-temporal data [1]. It is a three-axis coordinate system with two parallel vertical axes for latitude and longitude and an orthogonal horizontal axis for time. Figure 1 illustrates the basic ideas of Storygraph.

The top sub-figure in Figure 1 shows 6 accidents marked on a map. Two accidents have been reported at each location at different times of the year. However, as shown in this figure, plotting these data points on a map results in overlapping markers. For the remaining non-overlapping markers, maps fail to show the temporal distance between these events. The sub-figure at the bottom shows the same events presented in Storygraph. Here, events are plotted on the location lines with no overlapping. In addition, Storygraph presents the temporal distance between the events. Figure 2 shows a Storygraph generated from the World Trade Center (WTC) corpus generated by our program. Few patterns that can be observed in this Storygraph are: (1) The points are clustered around location $(40.70, -74.00)$, (2) At times $t1$ - $t4$ and later on around $15:12$, there are events simultaneously taking place at many different locations.

Interpreting spatial information on Storygraph is not as intuitive as that on a map; however, analyzing temporal information on Storygraph is quite intuitive. The following analysis discusses the process of interpreting the spatial information on Storygraph.

Based on [1], let α_{max} and α_{min} be the maximum and minimum latitude, and β_{max} and β_{min} be the maximum and minimum longitude. Likewise, let T_{max} and T_{min} be the maximum and the minimum timestamps.

The mapping function $f(\alpha, \beta, t) \rightarrow (x_{storygrah}, y_{storygraph})$ of event $E(l_\alpha, l_\beta, t)$ is given by:

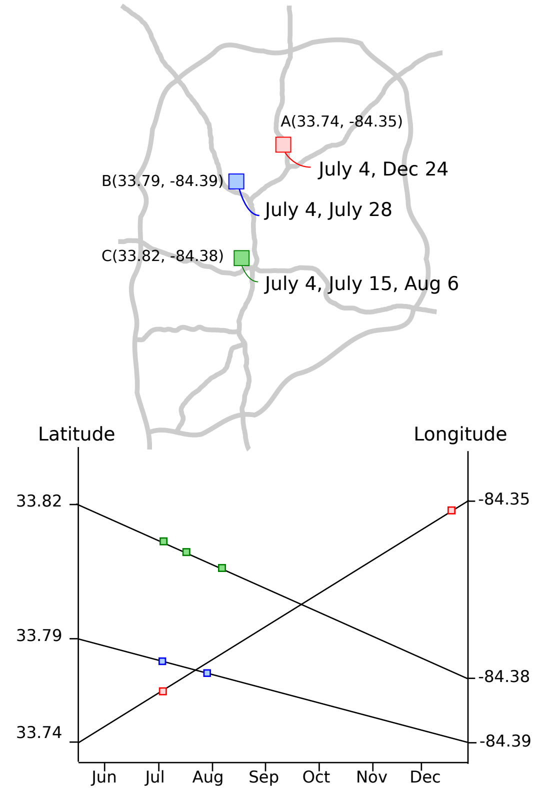

Figure 1: Example of Storygraph constructed from hypothetical accidents. Top: Outline map showing the major highways in Atlanta and hypothetical accident taking place at the junctions A, B, and C on the dates shown. Bottom: Same information plotted on Storygraph (not drawn to scale for illustrative purposes). Each location is represented as a line joining the latitude and longitude in the vertical axes. An event occurring at that location is represented by a point on the line. This representation allows users to see the temporal context of the events together with spatial context (i.e. when did most accidents take place? July-August in the figure above.)

$$y_{storygraph} = \frac{(\beta - \alpha)(x - T_{min})}{T_{max} - T_{min}} + \alpha \quad (1)$$

$$x_{storygraph} = t \quad (2)$$

Assuming $T_{min} = 0$ and $T_{max} = T$ without loss of generality, Equation 1 simplifies to

$$y = \frac{(\beta - \alpha)}{T} x + \alpha \quad (3)$$

Equation 3 is also the equation of the location line (Equation 1 rewritten in slope-intercept form).

In earlier sections, we discussed that a point on the Storygraph in the absence of location line can be mapped to range of locations in geographical space. Thus, the function f ceases to be one-to-one.

Lemma 3.1. *A point on a location line in Storygraph at*

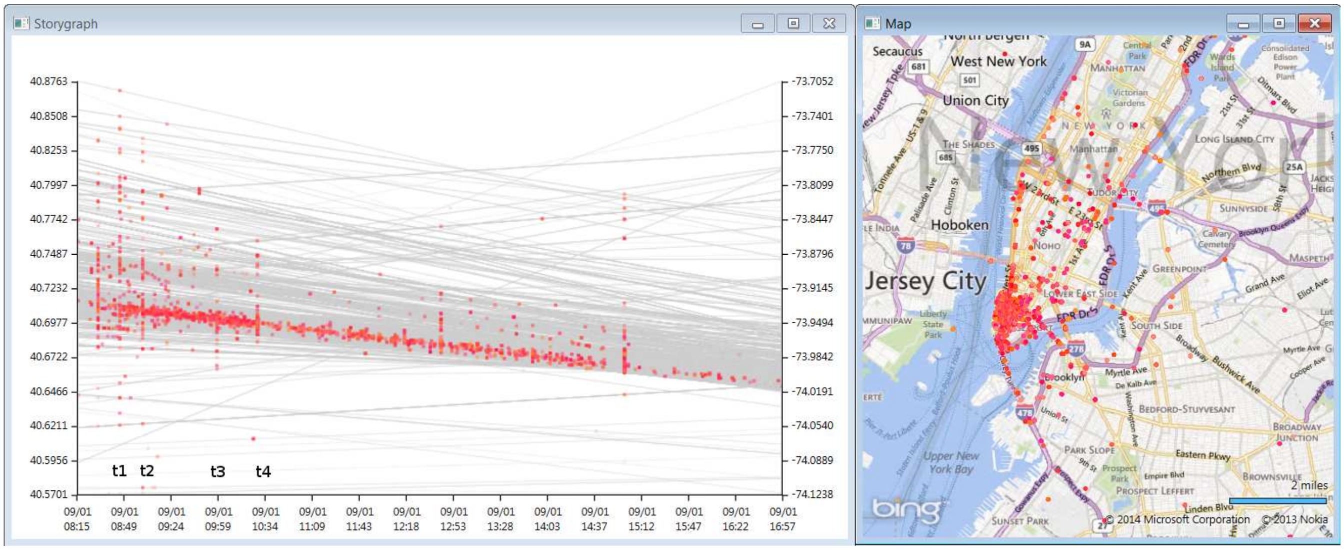

Figure 2: Left: Storygraph showing approximately 7000 events within 12 hours during 9/11 attack on WTC. Annotations $t1 - t4$ mark the key events: $t1(8:46)$, first plane crashes into the North Tower; $t2(9:03)$, second plane crashes into South Tower; $t3(9:59)$, South Tower collapses; $t4(10:28)$, North Tower collapses. At each of these times, events occurred simultaneously at multiple locations (marked by vertically aligned events). In addition, it can also be observed that the events clustered around the location $(40.70, -74.00)$. Right: Same set of events plotted on the map. Maps supplement Storygraphs as identifying locations on maps is relatively more intuitive.

time t corresponds to a precise point (geo-coordinate) on a map.

PROOF. Setting $T = 0$ and $T = t$ Equation 3, we get the y_{lat} and y_{lng} of the Storygraph. Thus, geo-coordinates (α, β) can be obtained as

$$\alpha = y_{lat} \times \frac{\alpha_{max} - \alpha_{min}}{\alpha_{min} \times y_{max}} \tag{4}$$

$$\beta = y_{lng} \times \frac{\beta_{max} - \beta_{min}}{\beta_{min} \times y_{max}} \tag{5}$$

□

LEMMA 3.2. *Without location lines, a point on a Storygraph at time t corresponds to a line segment on a map.*

PROOF. We can rewrite equation (3) as

$$\beta = (1 - \frac{T}{x})\alpha + \frac{yT}{x} \tag{6}$$

Thus, a fixed point (x, y) on the Storygraph corresponds to many points (α, β) on the Cartesian map at time $t = x$: those $\alpha_{min} \leq \alpha \leq \alpha_{max}$ and $\beta_{min} \leq \beta \leq \beta_{max}$ satisfying (6). Plotting these values of (α, β) results in a line segment with non-positive slope since $x \leq T$ as illustrated in Figure 3. □

LEMMA 3.3. *Without location lines, a vertical line segment at time t on a Storygraph corresponds to an area on a map.*

PROOF. Consider a vertical line segment, with end coordinates (x, y_1) and (x, y_2), $y_1 \leq y_2$. Using 3.2, these extremes of the line segment in (6) we get two straight line equations

$$\beta = (1 - \frac{T}{x})\alpha + \frac{y_1 T}{x} \tag{7}$$

$$\beta = (1 - \frac{T}{x})\alpha + \frac{y_2 T}{x} \tag{8}$$

Hence the vertical line segment between (x, y_1) and (x, y_2) on the Storygraph corresponds to an area between two parallel lines (7) to (8) in the geographical space. As in Lemma 3.2, this area is also bounded by the maximum and minimum values of α and β – this results in a polygon as illustrated in Figure 4. □

LEMMA 3.4. *Without location lines, a vertical line segment at t on the Storygraph corresponds to a projected area, $A_{Storygraph} \geq A_{actual}$ in geographical space at t.*

PROOF. If the area on the plane is bounded by right rectangle, since $\forall\alpha : \alpha1 \leq \alpha \leq \alpha2$ and $\forall\beta : \beta1 \leq \beta \leq \beta2$, $A_{Storygraph} = A_{actual}$. For any other shape, the vertical line segment in the Storygraph represents a rectangular bounding box (from 3.3). Thus, $\exists\alpha : \alpha \in A_{Storygraph} - A_{actual}$. Hence, $A_{Storygraph} \geq A_{actual}$ □

COROLLARY 3.5. *Real-world area at time t maps to a vertical line segment in storygraph at time t.*

PROOF. Inverse of Lemma 3.4, when the exact coordinates of all the four corners are known

we can state that an area A_{actual} in the geographical space gets mapped to a line segment in the Storygraph orthogonal to the time axis. The area formed by this line segment l bounded by coordinates (α_1, β_1, t) and (α_2, β_2, t) is given by $A_{Storygraph} = (\alpha_2 - \alpha_1)^2 + (\beta_2 - \beta_1)^2$. Thus, $A_{Storygraph} \geq A_{actual}$. □

LEMMA 3.6. *Storygraph preserves spatial proximity for location lines but does not preserve spatio-temporal proximity for events.*

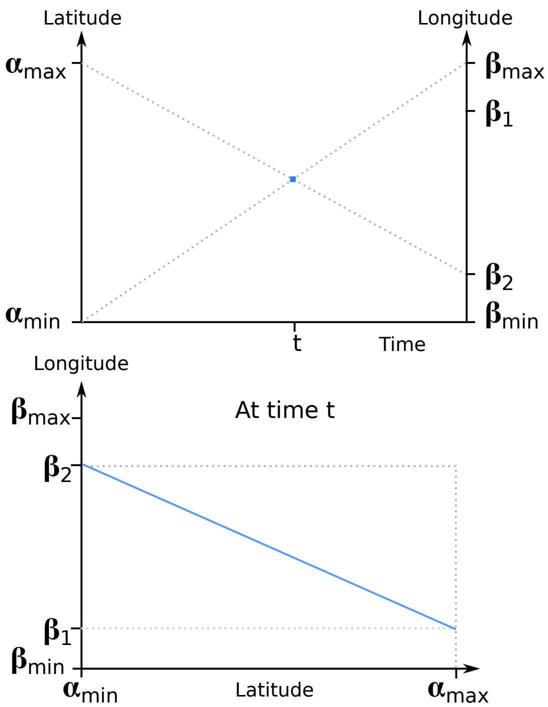

Figure 3: Top: A point in the Storygraph at time t and the corresponding location lines the point can belong to shaded. Bottom: The line segment generated in the Cartesian coordinate by mapping the point.

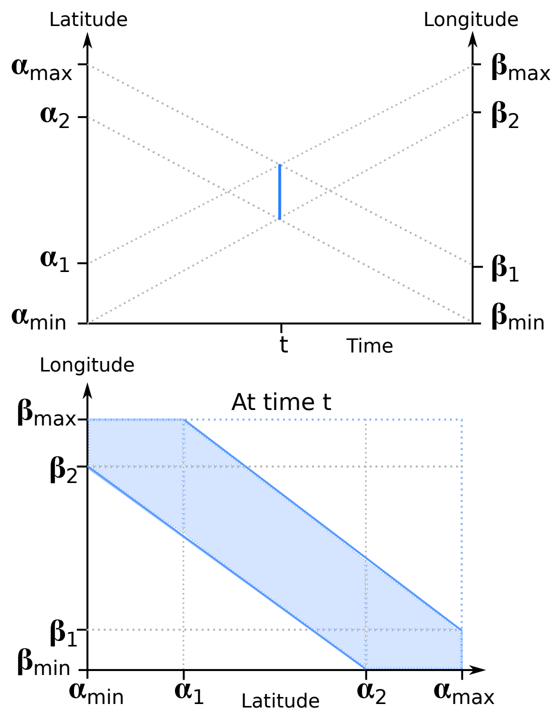

Figure 4: Top: A vertical line in Storygraph at time t and the corresponding location lines the line segment can belong to shaded. Bottom: The bounded region generated in the geographical space by mapping the line segment.

Proof. Two events close to each other in Storygraph may not be close to each other in geographical space. Consider two locations (α_1, β_1) and (α_2, β_2) in geographical space where $\alpha_1 \ll \alpha_2$. Since both of the axes are ordered in Storygraph, $\alpha_1 \ll \alpha_2$ holds true as well. □

4. CLASSIFICATION OF UNCERTAINTY

Different classifications of uncertainty have been proposed [27][28]; however, most of these classifications are about uncertainties introduced in scientific experiments or probabilistic models. In our case, uncertainties are introduced in narratives. Thus, we classify this kind of uncertainty into three categories:

Uncertainty about time and/or location of the event. This type of uncertainty is characterized by the presence of phrases denoting uncertainty before temporal or spatial description. An example is "I got there maybe around 11 am." The phrase 'maybe around' adds uncertainty to '11 am' in this example. Such uncertainties may also arise from ambiguity in language. For example, in "I was in Brooklyn when the plane hit the building," the word 'Brooklyn' does not give a precise location. We call these types of uncertainties as *event uncertainty* which can be further divided into three sub-categories:

- *Spatial uncertainty.* This category includes events that have precise time stamps but uncertain location.
- *Temporal uncertainty.* In addition to uncertain phrases (e.g. maybe, about), temporal uncertainty may come from the language itself. For example, in "I was at the station in the all day," the phrase "all day" without any modifier can refer to a wide range of time introducing uncertainty.
- *Spatio-temporal uncertainty.* This category includes events that have uncertainty in both time and location. For example, in "It was in the afternoon, I was heading south." The words 'afternoon' and 'south' are uncertain.

Uncertainty between two events. In "It was 8 in the morning I was at home. As soon as I heard about it, I reached the site at 10.", the first event ("at home") and the second event ("reached the site") are both certain. However, what happened between the two events is unknown. We call this type of uncertainty *between-even uncertainty*

Uncertainty about the even taking place. In the WTC corpus, we often encounter sentences like "I think Chief pulled me back". The word 'think' indicates an uncertainty about whether the event has ever happened. Detecting this type of uncertainty is difficult and beyond the scope of this paper. Instead, we focus only on visualizing event uncertainty and between-event uncertainty.

5. EVENT UNCERTAINTY

In this section, we discuss the extraction and visualization of event uncertainty. To extract event uncertainty, we compiled a list of English words that may indicate location uncertainty, such as "around," "near," "close to," "maybe," "perhaps," etc. We then gave each word an uncertainty

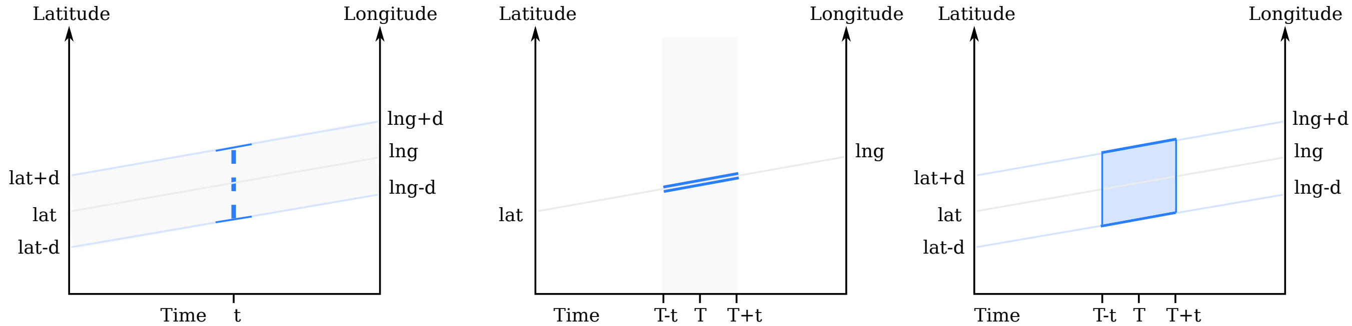

Figure 5: Three kinds of glyphs used to represent spatial, temporal and spatio-temporal uncertainty. Left: Dashed I-beam is used to represent spatial uncertainty. The slope of the top and bottom of the beam disambiguates the range of locations in the geographical space. Middle: parallel lines are used to denote the temporal uncertainty. Right: Box showing spatio-temporal uncertainty. The slope of the edges of the box maps to a fixed geographical area within a certain time.

score in the range of $1 - 100$ [29][30][31]. The same process was repeated for temporal information. We extracted the named entities from WTC corpus using Stanford NER [32] and time using SuTime [33]. TARSQI [34] was used to extract the temporal sequence of the events, and locations were geocoded using Google Maps API. The results were then verified and corrected.

In the WTC corpus, we observed all three types of event uncertainties: spatial, temporal, and spatio-temporal. Some key events with precise spatio-temporal information were used as anchor events. These include the first and second plane hitting the tower, and the plane crashing into the pentagon. These events were chosen as key events because all of the interviews described more local events in reference to these global events. Examples include "When the second plane hit the tower, I was running towards Vesey," and "I was at the station when the news about the first explosion was on TV." When considering these key events in the context of the first example, the time is certain but the location is uncertain. Additionally, in a sentence that references no key events like "When the EMS arrived at the scene, I began heading south", both location and time would be considered uncertain.

For each event, latitude, longitude, date/time, color, spatial uncertainty, and temporal uncertainty were fed to the visualization program, which then visualized the uncertainty information along with other information.

Spatial Uncertainty. Spatial uncertainty is visualized as a vertical dashed I-beam. From Corollary 3.5, we know that an area on a map corresponds to a line in Storygraph. The length of the I-beam is proportional to the area of possible locations. More importantly, the top and the bottom of the beam disambiguate the range of locations in geographical space. This is shown by the left sub-figure in Figure 5.

Temporal Uncertainty. We use sloped double lines to represent temporal uncertainty. Each double line is drawn along the location line for the corresponding event, which can be seen in the middle sub-figure in Figure 5. A double line indicates that the event happens at a particular location within a certain time frame. In contrast, a single solid line along the location line means that the character stayed at the specified location for a period of time. Through these representations, the two cases are visually distinct.

Spatio-temporal Uncertainty. We use a semi-transparent box to visualize spatio-temporal uncertainty, which means both location and time are uncertain. The sloped top and bottom sides of the box indicate the range of locations while the vertical sides of the boxes shows the temporal bound. The box is drawn as semi-transparent to prevent glyph occlusion. This is shown by the right sub-figure in Figure 5.

Figure 6 shows this concept applied to the events extracted from WTC corpus.

6. BETWEEN-EVENT UNCERTAINTY

The purpose of visualizing between-event uncertainty is to display the space-time constraints between two key events. Any activity takes place within a certain span of time and a certain geographical region. Individuals participating in these activities have to trade time for space or vice versa. For example, during a workday lunch hour a person could walk to a nearby restaurant for a longer meal or drive to distant restaurant for a shorter meal. Visualizing between-event uncertainty can assist planning, scheduling, analyzing possible overlapping in people's activities.

Our between-event visualization technique is partially based on Hagerstrand's Time Geography, a conceptual framework which focuses on constraints and trade-offs in the allocation of time among activities in space [5]. However, Time Geography is a map based 3D visualization. Therefore it suffers from the typical problems associated with 3D visualizations, such as 3D occlusion and difficulty of navigation. Besides, space and time are not well integrated in Time Geography. Our work is an attempt to address these issues.

6.1 Space-time paths and space-time prisms

We adapted two important concepts from Time Geography: space-time paths and space-time prisms. Space-time path traces the movement of a character in space and time. Figure 7 shows an example of a space time path adapted from [5]. The base plane is the geographical space and the orthogonal axis is time. In this example, an individual travels from location 1 to 2, spends some time at 2 and then moves on to 3. The time and location of the starting or end point are known as *control points* or *key events.* The straight line segments connecting two control points are known as *path segments.* Path segments are represented by straight line segments for simplicity [35][36]. In our earlier work, we

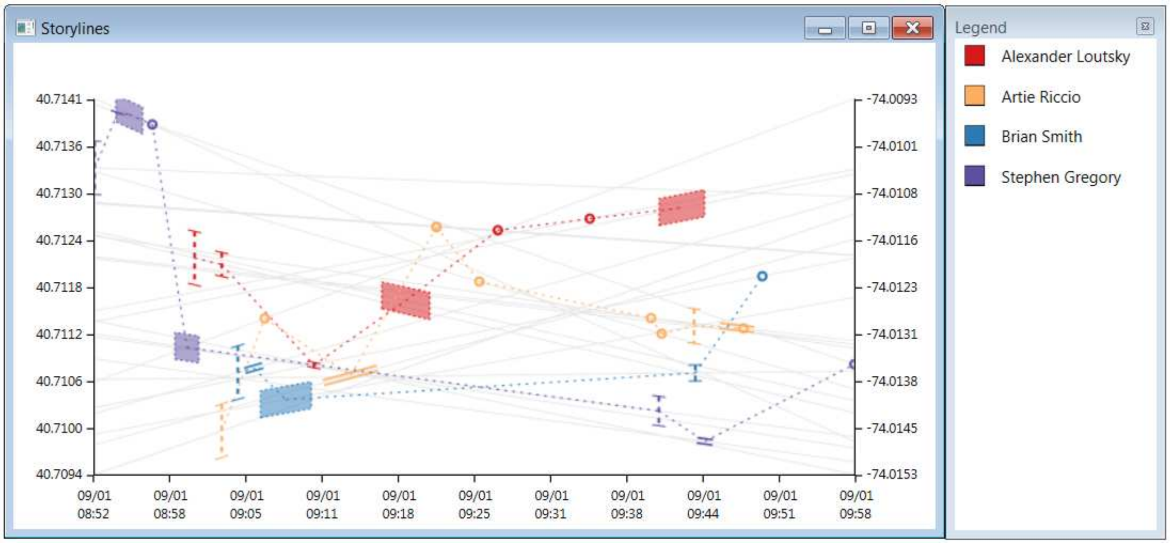

Figure 6: Storylines of four firefighters before the second tower collapsed along with event uncertainty. The dashed vertical I-beam shows the spatial uncertainty. The slope of the top and bottom portion of the beam shows the possible range of locations. The parallel lines show temporal uncertainty. The boxes represent spatio-temporal uncertainty and the circles show certain events.

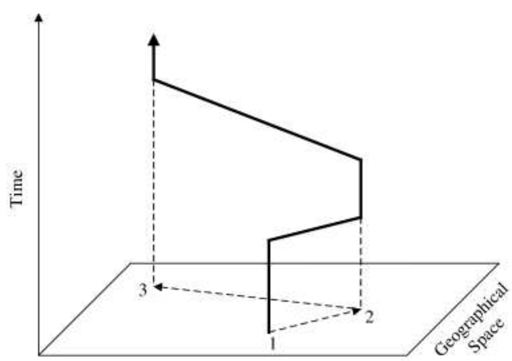

Figure 7: An example of space time path adapted from [5]. Space time paths trace the movement of an individual moving from one location to another. Space-time paths also show the amount of time spent at a location by the individual before moving to the next location.

adapted the concept of space-time paths in Storygraph using storylines[1]. Here, Storylines become space-time paths, connecting two consecutive key events via dotted line segment.

Space-time prisms extend space time paths to create a 3D space consisting of all the possible routes an individual can take while moving from one point to another. This space is known as the *potential path space*. The prism between $t1$ and $t2$ in Figure 8 demonstrates this concept. The slope of the edges of this prism is determined by the inverse of maximum velocity. That is, the possible paths are constrained by the maximum velocity of the individual, a fixed time frame, and fixed destinations. In our implementation, the maximum velocity is set by the user.

If an individual is at origin, o, at t_o and needs to reach destination, d, at t_d, the time budget is $T = t_d - t_o$. The path space from the origin under the time budget is shown by the red inverted dotted cone. This space shows all the possible paths and all the possible locations that can be reached within the time budget with maximum velocity v. Let this region be denoted by $R_o(T)$. Similarly, the blue dotted cone shows the path space towards d under the time budget. This 3D space gives all the locations from where d can be reached under time T. Let this region be $R_d(T)$. The intersection of these cones give the potential path space for individual traveling from o to d [37]. Hence,

$$R_{od}(T) = R_o(T) \cap R_d(T) \tag{9}$$

The projection of the space time prism on the geographical space, as shown by a gray circle in the figure, shows all the possible locations that the user can reach. This area is called the *potential path area*.

Given all the control points within a specific time window, τ, the construction of space-time prism requires the destination d to lie within the $R_o(T)$ and vice versa. Stating it formally,

$$\forall o, d \in \phi_\tau : (o \cap R_d(T) \neg \emptyset) \wedge (d \cap R_o(T) \neg \emptyset) \tag{10}$$

In Time Geography, space-time paths and space-time prisms are generally drawn inside a 3D space-time cube [38] (Figure 8). In our work, space-time paths and space-time prisms are drawn on Storygraph in a 2D view.

6.2 Visualizing between-event uncertainty

Storygraph draws space-time prisms based on Equations 9 and 10. From Corollary 3.5, we know that an area in the geographical space is mapped to a line in Storygraph. Thus starting from a location, $o(\alpha, \beta)$, at $t0$ and taking a snapshots of the potential path area at each time step we get a set of areas sequentially increasing at the rate of the velocity.

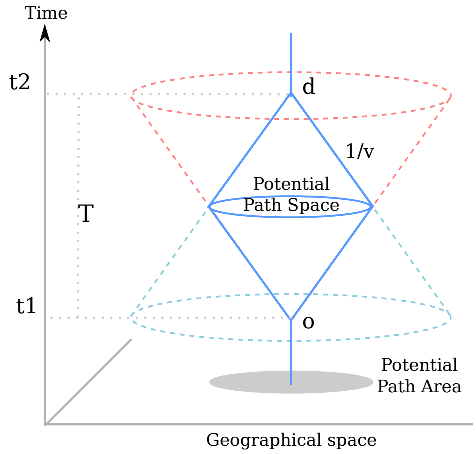

Figure 8: Space-time prism. In this figure, the individual is at location o at $t1$ needs to be at the same location d at $t2$ (o origin of travel). (S)he has the time budget of T. The red dotted cone shows the possible path space starting from o with the maximum velocity, v. Similarly the blue dotted cone shows the path space towards d. The intersection of these two cones gives the potential path space under the given time budget T. The potential path area is shown by the gray area on the geographic space.

The top sub-figure in Figure 9 shows an individual at point (α, β) at $t0$ and his/her possible path area after each time step $t1 - t5$. The figure simplifies the drawing of the potential path areas by representing them with squares rather than circles. The bounding of the actual potential path by squares introduces some uncertainty itself [20] but greatly simplifies the drawing and the calculations.

Hence, if the time step, $\Delta t \to 0$, then conical region $R_o(T)$ would be reduced to a triangular region in Storygraph. This region is shown by the area enveloped by two gray lines in the bottom sub-figure in Figure 9.

Figure 10 shows the result of mapping the space-time prism in Figure 8 in Storygraph. The mapping process resembles the drawing of the space-time prism inside the space-time cube. Given two control points, maximum velocity and a time budget, these parameters are plugged into Equation 10 to check whether the control points satisfy this criteria. If the criteria is satisfied, we compute the extents (lat_{max}, lng_{max}) and (lat_{min}, lng_{min}) of the $R_o(T)$ with the following sets of equations,

$$lat_{max} = max_{lat_r}[\sqrt{(lat_r - lat_o)^2 + (lng_r - lng_o)^2} = vT] \quad (11)$$

$$lng_{max} = max_{lng_r}[\sqrt{(lat_r - lat_o)^2 + (lng_r - lng_o)^2} = vT] \quad (12)$$

$$lat_{min} = min_{lat_r}[\sqrt{(lat_r - lat_o)^2 + (lng_r - lng_o)^2} = vT] \quad (13)$$

$$lng_{min} = min_{lng_r}[\sqrt{(lat_r - lat_o)^2 + (lng_r - lng_o)^2} = vT] \quad (14)$$

Similarly, the extents for the $R_d(T)$ is calculated. Finally, $R_{od}(T)$ is obtained from the intersection of these regions.

6.3 Intersections of prisms in Storygraph

Space-time paths and prims are both based on the movement data of characters. Given a dataset containing the movement data of two or more individuals, it is likely that the space-time prisms will overlap. However, since Storygraph does not preserve event proximity (Lemma 3.6), it is important to note that these overlaps may not necessarily mean that these prisms intersect in geographical space.

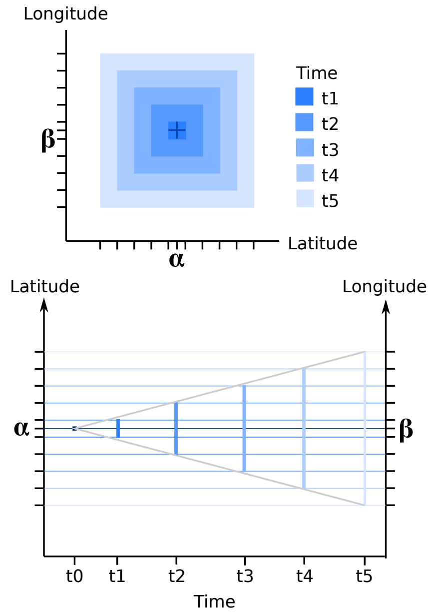

Figure 9: Above: Starting from the origin of travel, α, β, at $t0$, the potential path areas in each time step $t1 - t5$ (assuming a certain velocity and regular time intervals). Below: The same data plotted in Storygraph. Each potential path area is mapped to a line segments in Storygraph. For continuous time, this would result in an area enveloped by two gray lines. The slope of the gray lines is equal to the maximum velocity.

Hence, given a point p and a prism $R_a(T)$ in the Storygraph, we first establish the conditions for a valid point-prism intersection. Building on this, we then present the validity of intersection between two prisms.

Let $R_a(T) : R_a(T) = R_o(T) \cup R_d(T)$, be all the possible locations that the individual can travel within the time budget T with a maximum velocity v. Then following cases for point-prism intersection could arise:

1. The point is not inside the prism but the location line is inside $R_a(T)$. This case implies that the event occurred within the geographical bounds but the individual may not have been involved in the event due to the travel constraints.

2. The point is inside the prism but the location line is not inside $R_a(T)$. This implies that the event occurred within the time span, T, but at some other location $\notin R_a(T)$.

3. The point is inside the prism and location line is inside $R_a(T)$. This is the only case where the individual could have been involved in the event.

THEOREM 6.1. *For a valid point-prism intersection, the point should be inside the prism and the location line should lie inside $R_a(T)$.*

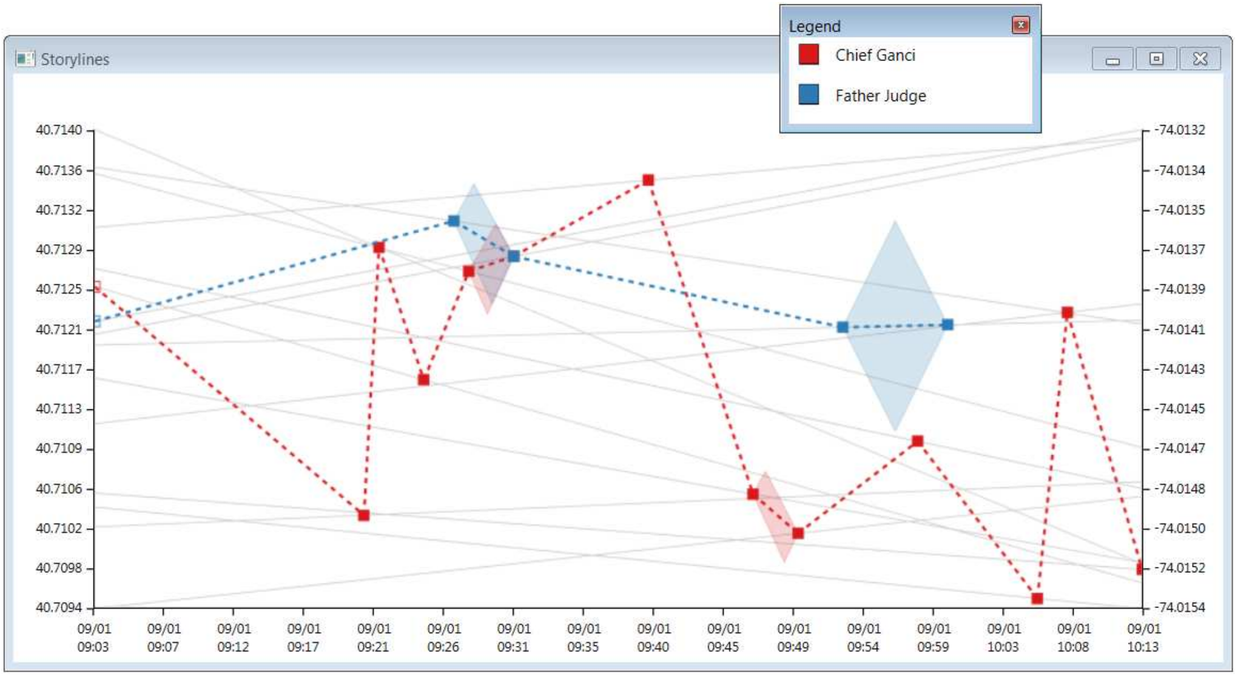

Figure 11: Storylines of WTC victims Chief Ganci and Father Judge. Father Judge was officially identified to be the first victim of the incident. Only a few key points have prisms between them. For other key points, the distance between them cannot be travelled within the given time at a velocity set by the user. This could either mean missing data points, change in velocity, or data reporting error.

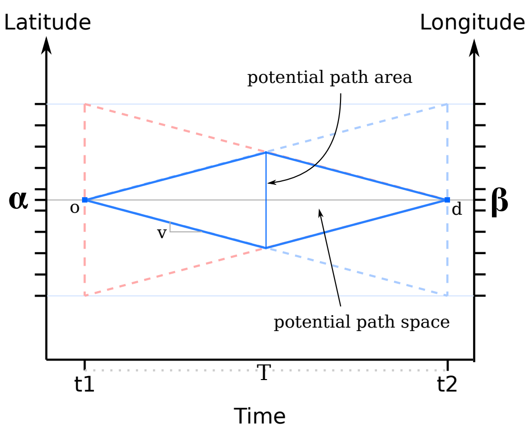

Figure 10: The space-time prism shown in Figure 8 drawn in Storygraph.

PROOF. Assume that this is not a valid intersection. It means that the point representing the event is either spatially or temporally incorrect. This is temporally incorrect because for a point to lie inside the prism, it has to occur within the time budget. This is spatially incorrect since $R_a(T)$ defines the maximum distance an individual can travel at within a time T. $\square$

Hence, given two prisms, $P1$ and $P2$, the prism-prism intersection is only valid if there exists a point p on location line l such that $l \in R_a^{P1}(T) \wedge l \in R_a^{P2}(T) \wedge p \in P1 \cap P2$.

7. CASE STUDY: WTC 9/11

In the immediate aftermath of the attacks in New York on September 11, 2001, the NYC Fire Department convened a task force to interview first responders to the affected areas. These 511 interviews, conducted in the two months following the attacks, were later released by the New York Times. Each interview was conducted by staff from the New York Fire Department assigned to the task force and ran anywhere from 8-20 minutes with the aim to elicit from first responders their activities on September 11. The language of the reports is typical of event interviews and oral histories. Despite having a population with high area knowledge and normalized reporting practices, locations and times were predominately referred to referentially. Known individuals seen by the interviewee are named, but most are either not named or referred to solely by rank. The primary reason to visualize this data is to enable historians and investigators to identify accurate and inaccurate information and to allow for more ready recognition of corroborating evidence. When viewed as a corpus rather than separate interviews, it becomes possible to identify overlaps in the reported events of the witnesses. The challenge posed to this task by the referential language usage of the witnesses is pervasive in oral history and other investigatory work reliant on interviewing.

Event Uncertainty Visualization. Time, location, and characters (or people) in this corpus were extracted using Java code and the aforementioned natural language processing tools. Each event was given an uncertainty score using the method described in Section 5. We first drew a Storygraph without uncertainty information (Figure 2). In this figure, key events – such as when the first and second plane hit and when the towers collapsed – are shown by $t1 - t4$. There are many co-occurring events around 15 : 00 hrs, but the causes of these patterns are not yet clear.

Next, we plotted the storylines of four fire fighters before the South Tower collapsed with event uncertainty (Figure 6). It should be noted that two storylines crossing does not necessarily mean the two characters encounter each other; rather, it only means that two people were moving in directions diagonal to each other. One limitation of using uncertainty glyphs is that they might result in occlusion and ambiguity for large datasets. When the dataset is large, the bigger glyphs (e.g. the ones representing spatio-temporal uncertainty) could occlude the smaller ones.

Between-event Uncertainty Visualization. Figure 11 visualizes the between-event uncertainty for two victims: Father Judge and Chief Ganci. The space-time prisms in Storygraph enable users to see the possibilities of individuals encountering each other between key events. There are two patterns in this figure: (1) the prisms are only present between some key events, and (2) some prisms overlap. The first pattern indicates that locations of the two events are too far apart in that it would be impossible for a person to cover that distance at the maximum velocity. It does not necessarily mean that part of the story is false; rather, it may be the result of missing information between two events or uncertainty in the events themselves. From overlapping prisms (from Theorem 6.1), we can also deduce that Chief Ganci and Father Judge might have encountered each other within that time frame and region.

8. CONCLUSION

In this paper, we presented a new method for visualizing uncertainty in spatio-temporal data set. This method is an extension of our previous work Storygraph, a visualization technique for displaying spatio-temporal data sets in an integrated 2D view. Our method can visualize both temporal-spatial uncertainty about an event and the uncertainty between events. This extended method provides more accurate and faithful visualization of spatio-temporal data sets with inherent uncertainties. In addition, between-event uncertainty visualization can help users analyze the feasibility of spatio-temporal events and possible encounters between multiple characters. We demonstrated this method in a case study.

In the future we plan to conduct user studies to evaluate the effectiveness of this method and compare it with other methods. We also plan to investigate new methods to analyze and visualize uncertainty in the identification of people or groups.

9. ACKNOWLEDGEMENT

This project is supported in part by National Science Foundation Grant No. 1209172. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

10. REFERENCES

[1] Ayush Shrestha, Ying Zhu, Ben Miller, and Yi Zhao. Storygraph: Telling stories from spatio-temporal data. In *Advances in Visual Computing*, pages 693–702. Springer, 2013.

[2] A. Shrestha, Ying Zhu, and B. Miller. Visualizing time and geography of open source software with storygraph. In *Software Visualization (VISSOFT), 2013 First IEEE Working Conference on*, pages 1–4, Sept 2013.

[3] Torsten Hägerstrand. Reflections on "what about people in regional science?". In *Papers of the Regional Science Association*, volume 66, pages 1–6. Springer, 1989.

[4] Nigel Thrift. An introduction to time-geography. Geo Abstracts, University of East Anglia, 1977.

[5] Harvey J Miller. A measurement theory for time geography. *Geographical analysis*, 37(1):17–45, 2005.

[6] Wolfgang Aigner, Silvia Miksch, Wolfgang Muller, Heidrun Schumann, and Christian Tominski. Visualizing time-oriented data - a systematic view. *Computers and Graphics*, 31(3):401 – 409, 2007.

[7] W. Aigner, S. Miksch, W. Muller, H. Schumann, and C. Tominski. Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics*, 14(1):47 –60, jan.-feb. 2008.

[8] D. Fisher, A. Hoff, G. Robertson, and M. Hurst. Narratives: A visualization to track narrative events as they develop. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology*, pages 115 –122, oct. 2008.

[9] K. Vrotsou, J. Johansson, and M. Cooper. ActiviTree: Interactive visual exploration of sequences in event-based data using graph similarity. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):945 –952, nov.-dec. 2009.

[10] W. Javed, B. McDonnel, and N. Elmqvist. Graphical perception of multiple time series. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):927 –934, nov.-dec. 2010.

[11] T. Gschwandtner, W. Aigner, K. Kaiser, S. Miksch, and A. Seyfang. CareCruiser: Exploring and visualizing plans, events, and effects interactively. In *Proceedings of IEEE Pacific Visualization Symposium (PacificVis)*, pages 43 –50, march 2011.

[12] Zhao Geng, ZhenMin Peng, R.S. Laramee, R. Walker, and J.C. Roberts. Angular histograms: Frequency-based visualizations for large, high dimensional data. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2572 –2580, dec. 2011.

[13] Jian Zhao, F. Chevalier, E. Pietriga, and R. Balakrishnan. Exploratory analysis of time-series with ChronoLenses. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2422 –2431, dec. 2011.

[14] M. Krstajic, E. Bertini, and D. Keim. CloudLines: Compact display of event episodes in multiple time-series. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2432 –2439, dec. 2011.

[15] R. Maciejewski, S. Rudolph, R. Hafen, A. Abusalah, M. Yakout, M. Ouzzani, W.S. Cleveland, S.J. Grannis, and D.S. Ebert. A visual analytics approach to understanding spatiotemporal hotspots. *IEEE Transactions on Visualization and Computer Graphics*, 16(2):205 –220, march-april 2010.

[16] Ali Asgary, Alireza Ghaffari, and Jason Levy. Spatial

and temporal analyses of structural fire incidents and their causes: A case of Toronto, Canada. *Fire Safety Journal*, 45(1):44 – 57, 2010.

[17] Charlotte Plug, Jianhong (Cecilia) Xia, and Craig Caulfield. Spatial and temporal visualisation techniques for crash analysis. *Accident Analysis and Prevention*, 43(6):1937 – 1946, 2011.

[18] Gennady Andrienko, Natalia Andrienko, Martin Mladenov, Michael Mock, and Christian Politz. Identifying place histories from activity traces with an eye to parameter impact. *IEEE Transactions on Visualization and Computer Graphics*, 18(5):675–688, May 2012.

[19] SungYe Kim, R. Maciejewski, A. Malik, Yun Jang, D.S. Ebert, and T. Isenberg. Bristle maps: A multivariate abstraction technique for geovisualization. *Visualization and Computer Graphics, IEEE Transactions on*, 19(9):1438–1454, Sept 2013.

[20] K. Brodlie, R. A. Osorio, and A. Lopes. A review of uncertainty in data visualization. In *Expanding the Frontiers of Visual Analytics and Visualization*. Springer, London, 2012.

[21] Alison Love, Alex Pang, and David Kao. Visualizing spatial multivalue data. *IEEE Computer Graphics and Applications*, 25(3):69–79, 2005.

[22] Lydia Gerharz, Edzer Pebesma, and Harald Hecking. Visualizing uncertainty in spatio-temporal data. In *Spatial Accuracy 2010*, pages 169–172, 2010.

[23] Edzer J. Pebesma, Kor de Jonga, and David Briggs. Interactive visualization of uncertain spatial and spatio-temporal data under different scenarios: an air quality example. *International Journal of Geographical Information Science*, 21(5):515–527, 2007.

[24] T. Zuk, S. Carpendale, and W. D. Glanzman. Visualizing temporal uncertainty in 3d virtual reconstructions. In *Proceedings of the 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST 2005)*, pages 99–106, 2005.

[25] Mark Harrower. The cognitive limits of animated maps. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 42(4):349–357, 2007.

[26] Kirk Goldsberry and Sarah Battersby. Issues of change detection in animated choropleth maps. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 44(3):201–215, 2009.

[27] Kristin Potter, Paul Rosen, and Chris R Johnson. From quantification to visualization: A taxonomy of uncertainty visualization approaches. In *Uncertainty Quantification in Scientific Computing*, pages 226–249. Springer, 2012.

[28] Osman Abul, Francesco Bonchi, and Mirco Nanni. Never walk alone: Uncertainty for anonymity in moving objects databases. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pages 376–385. Ieee, 2008.

[29] Lee D Erman, Frederick Hayes-Roth, Victor R Lesser, and D Raj Reddy. The hearsay-ii speech-understanding system: Integrating knowledge to resolve uncertainty. *ACM Computing Surveys (CSUR)*, 12(2):213–253, 1980.

[30] Marie-Catherine de Marneffe, Christopher D. Manning, and Christopher Potts. Veridicality and utterance understanding. In *Proceedings of the 2011 IEEE Fifth International Conference on Semantic Computing*, ICSC '11, pages 430–437, Washington, DC, USA, 2011. IEEE Computer Society.

[31] Roser Sauri and James Pustejovsky. Factbank: a corpus annotated with event factuality. *Language Resources and Evaluation*, 43(3):227–268, 2009.

[32] Jenny Rose Finkel, Trond Grenager, and Christopher Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 363–370. Association for Computational Linguistics, 2005.

[33] Angel X Chang and Christopher Manning. Sutime: A library for recognizing and normalizing time expressions. In *LREC*, pages 3735–3740, 2012.

[34] Marc Verhagen, Inderjeet Mani, Roser Sauri, Robert Knippen, Seok Bae Jang, Jessica Littman, Anna Rumshisky, John Phillips, and James Pustejovsky. Automating temporal annotation with tarsqi. In *Proceedings of the ACL 2005 on Interactive poster and demonstration sessions*, pages 81–84. Association for Computational Linguistics, 2005.

[35] Dieter Pfoser and Christian S Jensen. Capturing the uncertainty of moving-object representations. In *Advances in Spatial Databases*, pages 111–131. Springer, 1999.

[36] José Moreira, Cristina Ribeiro, and Jean-Marc Saglio. Representation and manipulation of moving points: an extended data model for location estimation. *Cartography and Geographic Information Science*, 26(2):109–124, 1999.

[37] Tijs Neutens, Nico Weghe, Frank Witlox, and Philippe Maeyer. A three-dimensional network-based spaceâĂŞtime prism. *Journal of Geographical Systems*, 10(1):89–107, 2008.

[38] Menno-Jan Kraak. The space-time cube revisited from a geovisualization perspective. In *Proc. 21st International Cartographic Conference*, pages 1988–1996, 2003.