Proximity-based Visualization of Movement Trace Data

Tarik Crnovrsanin*
University of California, Davis

Chris Muelder†
University of California, Davis

Carlos Correa‡
University of California, Davis

Kwan-Liu Ma§
University of California, Davis

ABSTRACT

The increasing availability of motion sensors and video cameras in living spaces has made possible the analysis of motion patterns and collective behavior in a number of situations. The visualization of this movement data, however, remains a challenge. Although maintaining the actual layout of the data space is often desirable, direct visualization of movement traces becomes cluttered and confusing as the spatial distribution of traces may be disparate and uneven. We present proximity-based visualization as a novel approach to the visualization of movement traces in an abstract space rather than the given spatial layout. This abstract space is obtained by considering proximity data, which is computed as the distance between entities and some number of important locations. These important locations can range from a single fixed point, to a moving point, several points, or even the proximities between the entities themselves. This creates a continuum of proximity spaces, ranging from the fixed absolute reference frame to completely relative reference frames. By combining these abstracted views with the concrete spatial views, we provide a way to mentally map the abstract spaces back to the real space. We demonstrate the effectiveness of this approach, and its applicability to visual analytics problems such as hazard prevention, migration patterns, and behavioral studies.

Keywords: Spatio-temporal visualization, proximity, linked views, principal component analysis, temporal trajectories, movement patterns.

1 INTRODUCTION

The visualization of movement traces in time is a challenging problem. Traditional approaches to the problem map 2D and 3D space directly and use time as either an extra dimension, with coordinated views, or animation. However, since the spatial distribution of traces may be uneven, static representations become cluttered and difficult to understand in regions of high density. Although maintaining the original spatial layout is often desirable, certain applications may consider the proximity of entities to a set of points of interest to be more important than their spatial coordinates in a given fixed frame. For example, the analysis of movement data in emergency situations may benefit from a visualization that highlights the distance of each entity to the epicenter of an event. Evacuation drills and procedures can be analyzed by considering the distance of moving entities with respect to the exits. Although spatial context information is lost, relative motion events, such as convergence or divergence of movement, or the appearance of trends and fluctuations, can still be visualized using derived quantities such as proximity. Other global patterns, such as congestion, symmetry, flocking, and repetitions, may be observable as well.

In this paper, we present the notion of *proximity-based visualization*, which attempts to improve the visualization of spatio-temporal

*tecrnovrsanin@ucdavis.edu

†muelder@cs.ucdavis.edu

‡correac@cs.ucdavis.edu

§ma@cs.ucdavis.edu

data where relative motion patterns are important. These visualizations differ from traditional approaches in that we derive a multi-dimensional data set consisting of proximity values, then map the derived data to an abstract space rather than the given spatial layout. This abstract space helps to prevent clutter and to disambiguate overlapping traces in high density areas. Since we want to preserve global trajectories as well, we require that the abstract space is continuous. That is, a path followed by an entity in the actual given space will result in a continuous path in the proximity space.

We present several visualization spaces, depicted as 2D plots. The first one projects the proximity data to a single dimension, while the other dimension represents time. These visualizations can represent the entire spatio-temporal data set in a single view. The second type of plot provides a richer spatial visualization, at the expense of losing the time dimension, which is represented as traces. This plot type is useful for representing proximity data with respect to multiple points, and uses affine and projective transformations (such as PCA) to map proximity data to a 2D abstract space.

The visualization of proximity data enables us to discover entities that exhibit interesting behaviors, such as suspects, witnesses or casualties. Furthermore, it provides a better space for describing the evacuation procedure and patterns such as congestion, density, and dispersion. To test our approach, we use an evacuation data set obtained from the VAST 2008 challenge [13] and the Elk habitat data set [25], and we discuss the strengths and limitations of each technique and how different patterns can be observed in an abstract space.

2 RELATED WORK

The main challenge when visualizing movement traces is the simultaneous depiction of temporal and spatial data. Despite the long history and ubiquity of maps as visual representations, the effective representation of time remains a challenge [23]. Initial attempts often enhance maps with arrows and lines [30] or using animation, such as the pioneering work by Moellering [22], and recent approaches as surveyed by Harrower [11] and MacEachrean [21]. More sophisticated approaches use the space-time cube metaphor, where two dimensions represent spatial coordinates while the third dimension represents time. Movement traces are therefore represented as line strips in a three-dimensional space [1, 17]. The ability to manipulate the parameters of the visualization and the variables mapped to the 2D map has enabled a myriad of new visualizations of geospatial data [19]. Tominski et al. extend the notion of space-time cube with 3D icons to represent multivariate time dependent data [27].

One of the problems of the above visualizations is the clutter that may occur due to the asymmetry of the spatial data. Some have dealt with this problem with abstraction [1], aggregation [27], and clustering [10] of the spatio-temporal data. As an alternative to the added complexity, Ivanov et al. decompose the spatio-temporal nature of data into several synchronized views. One view presents spatial information as traces, and a timeline view shows temporal information [15].

All of the above keep the spatial layout given by the data. However, most of these data can be visualized in an abstract space, using multi-variate visualization techniques [10]. These include scatter-plot matrices [5], parallel coordinates [14], star coordinates [16], radial layouts [26] and circle segments [2]. Although the origi-

nal spatial structure is lost, these visualizations allow the depiction of additional variables in a single 2D image, time included, and scale better to large data sets. Given the increasing complexity of high-dimensional data, these visualizations often combine data transformations to reduce the complexity, such as clustering [3], or to reduce the dimensionality to 2D, such as projections, principal component analysis [6] and self-organizing maps [18].

The advantages of abstract spaces have been exploited to represent time. Early hand-made illustrations of timelines and graphs [28] have inspired visual metaphors such as lifelines [24], the ThemeRiver [12], and History Flow[31] where curved lines of varying thickness depict a varying quantity over time. For spatio-temporal visualization, however, line graphs remain the predominant metaphor.

In our paper, we use line graphs to represent time on an abstract space rather than the given spatial layout. We aim to preserve the spatial relationships within the data as much as possible, while harnessing the scalability and expressive power of abstract representations. We therefore turn to *proximity data*, which quantifies the distance among entities of interest rather than the spatial coordinates in a fixed frame. The collection of these proximities is a multivariate data set that can be visualized in an abstract space, and help formulate questions about the relative motion of entities, regarding their patterns, the appearance of bottlenecks and the detection of suspicious activities. We make use of principal component analysis and affine transformations [9] to reduce the dimensionality of this space.

Distance-based statistics and weighting has been explored to augment geographical visualization using secondary visual variables, such as color [8, 29]. The visualization of relative motion was explored by Laube et al. [20], who use pixel maps to depict the motion of entities over time. Weaver extended this idea to scatterplots, allowing the user to visualize different properties of relative motion such as time, azimuth, and speed [32]. Although these help the user discover motion patterns such as flocking, they may not represent global trajectories in an intuitive manner. In our paper, we show that global trajectories can be visualized in an abstract space while highlighting relative motion patterns and proximity information. In a recent article [7], Dodge et al. describe a taxonomy of movement patterns in terms of derived quantities from spatio-temporal data. Our paper explores one such derivative, namely the use of proximity or distance to reveal spatio-temporal patterns.
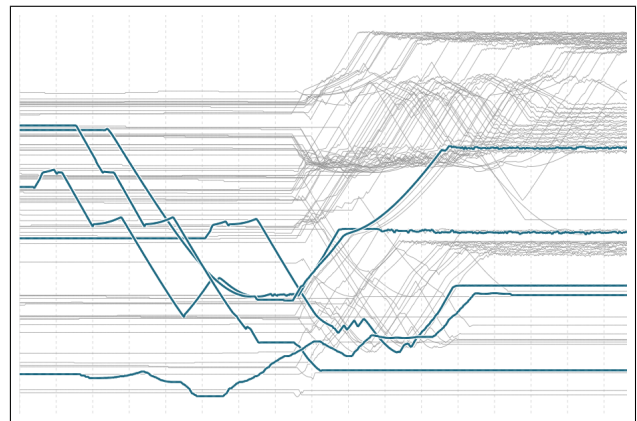
3 PROXIMITY-BASED VISUALIZATION

Traces can be defined as a representation of an object's movement through time. In general, the visualization of traces defines a direct mapping from the coordinates in a given reference frame to 2D or 3D coordinates in the visualization space. An example is shown in Figure 1(a), a visualization of a crowd simulation during an evacuation procedure following an incident (an explosion in one of the rooms). Squares represent the location of people, while lines represent the traces of movement over time.

A proximity-based visualization is an abstraction of the movement data by deriving relative information and using that information to obtain a novel view on the data. In particular, we consider *distances* as the derived data. Distance can be defined between moving entities and a fixed location, e.g. locations of doors or stairways in an office building. Distances can also be defined to entire regions, such as the distance to the closest road or water source when analyzing animal movement in the wild. In other cases, distances can represent abstract data, such as the similarity of two documents. Two documents are said to be "close" when they share a certain number of words or terms.

The advantage of using distance is an immediate reduction of dimensionality, that allows us to explore other dimensions, such as time, in a single 2D image. An example is shown in Figure 1(b),
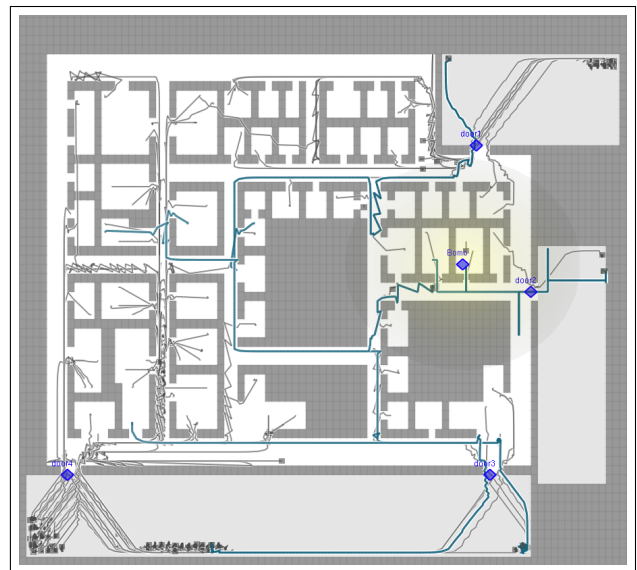
(a) Direct visualization of traces over time

(b) Proximity to a point of interest versus time

Figure 1: (a) Traditional visualization of the simulation of an evacuation in an office building after the detonation of an explosive in one of the rooms. Since time is not represented as its own dimension, it is not possible to understand what aspects of the event occurred concurrently. (b) Proximity-based visualization where traces are plotted as distance to the explosion (y-axis) vs. time (x-axis). Now, we can see the entire event at a glance. For example, we notice the motion of some entities before the detonation of the explosive, suggesting possible suspects and/or witnesses to the event.

where we plot the same information in Figure 1(a). In this case, each line is a movement trace that plots distance to the location of an explosion in a building (y-axis) versus time (x-axis). While traditional visualization of traces requires animation or switching between intervals to represent time, this variable is directly represented as its own dimension in proximity space. In this case, we can detect the moment when the evacuation occurs and identify entities that moved before the incident (highlighted in blue). Because we plot distance to an explosion, we can easily identify people that presumably died in the explosion or soon after, shown as steady lines near the x-axis.

In general, proximity-based visualizations are not limited to a single distance dimension. In many cases, it may be of interest to analyze the distance to multiple points of reference. Later on, we describe some of the resulting visualizations of multi-dimensional proximities.

(a) Spatial concentration   (b) Co-incidence

(c) Concurrence   (d) Trends

(e) Fluctuation   (f) Convergence
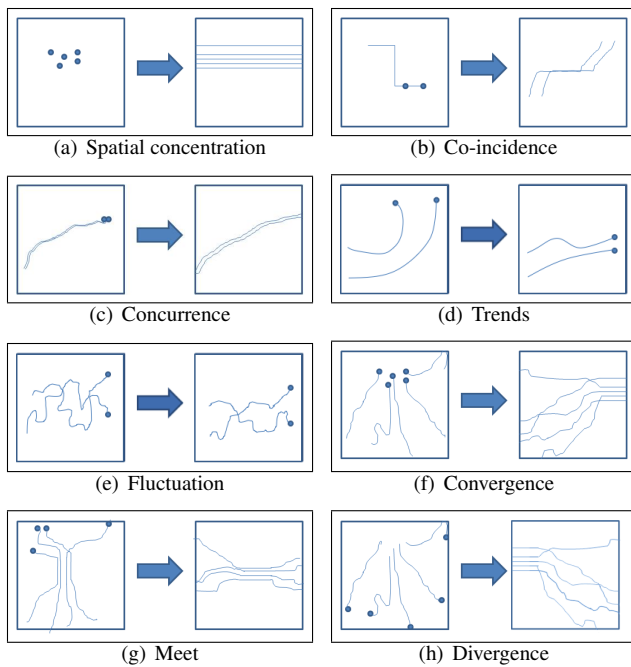
(g) Meet   (h) Divergence

Figure 2: Transition from normal movement data to proximity data. Each pattern is shown in absolute coordinates and mapped to a relative coordinate versus to time.

3.1 Movement patterns

One of the advantages of proximity-based visualizations is that they retain the general patterns in the data. As described by Dodge et al. [7], patterns emerge depending on the relative change of positions and velocities among moving patterns. Therefore, most of these are invariant to the frame of referenced used to define the motion, and can be identified (sometimes more easily) in proximity space. Figure 2 shows examples of movement patterns in a traditional visualization and their corresponding "signature" in proximity space.Should be noted, that this list is not a comprehensive, these pattern were chosen because they are common in proximity data and provide an example how patterns translate from one space to the other.

Movement patterns are generally any recognizable spatial and temporal regularity or any interesting relationship in a set of movement data. Spatial concentration is the concentration of moving entities at a certain instance in time. In proximity data, this is represented as entities that are near each other. For instance, people standing in the same room generates a spatial concentration pattern. Co-incidence is when two entities have the same positions in the same time or lag in time. In proximity data, this can be viewed as two lines that meet and coincide for a period of time. Concurrence is a set of entities exhibiting synchronous movement or at least similar motions, regardless of their relative speed. In proximity space, these tend to appear as curves that remain parallel. Trends are consistent changes in movement that can be predicted, while fluctuations are irregular changes in the movement. These patterns are retained in both spaces, because they relate to the derivatives of movement and not the frame of reference. The speed of a moving entity, and an overall trend, can be judged by the slope of its trajectory in proximity space. The higher the slope, the faster the entity moves. Convergence refers to the movement of a group of entities into the same location. In proximity data, this can be seen as tracks coming together to a point and possibly intersecting. A "meet" pattern is an extension of convergence, where after converging the entities stay together for some time. In proximity space, this pattern appears as the convergence pattern, but the converging point be-

comes a line. Divergence is the opposite of convergence and refers to the pattern formed when entities disperse from a common location.

3.2 Proximity continuum

When we consider proximities for visualizing movement data, we can imagine a continuum, spanning from the traditional representation in absolute coordinates, to visualizations that map the entire space in a relative frame of reference, i.e., proximity values. The continuum can also seen as a transition between a *spatial* frame of reference, where the visualization and the data spaces are mapped one to one, and an *abstract* space, where the visualization space is the result of a transformation from the data space, in this case, a transformation that derives proximity. Examples of these two extremes are the visualizations described in Figures 1(a) and 1(b). In between, we find visualizations that progressively trade elements in absolute coordinates to proximity. First, we have *single point abstraction*, where the locus of proximity is defined at a given point, without changing the actual locations of points. An example is shown in Figure 3, where we retain the map in absolute coordinates, but add views of the rest of the data relative to single points. We can see, in their respective plots, what characterizes the movement in their own proximity. For example, the entity on the left was closer to the bottleneck that lead towards the exit marked in green, and relatively far from the blue and red exit areas, which were closer to the zone where the incident occurred. Conversely, on the right, we see that most people ended up far from it, and a single entity seems to exit through the same area (blue line at the bottom). This type of proximity visualization enables the analyst to form hypotheses about individual behavior.

Next, we obtain visualizations that abstract entire regions. This enables the analyst to answer questions regarding crowd behavior and environmental factors in the movement of entities in physical regions. An example is shown in Figure 4, where we provide proximity visualizations for two hallways in the evacuation data set. We see that, for most part, people in these regions evacuated to the same area (green). However, we see one trajectory in the region to the right that evacuated to a different zone (marked up in red). In fact, we see that this trajectory is from a person that moved before the event occurred, and can be considered a suspect. Notice also the magnitude of the congestion in the hallway to the left compared to the hallway in the right.

4 PROXIMITY BASED VISUAL TECHNIQUES

We developed several techniques to visualize proximity information. One of them, as seen above, plots a single proximity quantity versus time. Proximity, is in general, a multi-dimensional collection of data. For this reason, we also present two techniques that handle multi-dimensional proximity in a 2D visualization. The effective visual analysis of these visualizations requires a series of operations, such as annotation and selection of entities and time, as described in the following sections.

4.1 1D and 2D Proximity Visualization

The traditional approach to dealing with movement data is to directly map the coordinates to space, as in Figure 1(a). One of the limitations of this approach is the inefficient use of space, which does not convey time very well. Even when trails are added, a static glance at the visualization does not explain concurrent events nor causal relationships. To get a good perspective of temporal patterns, the analyst must resort to back-and-forth interaction, with increasing cognitive load. Therefore, it becomes evident that mapping movement data to a one-dimensional proximity enables us to use an extra dimension as time.

Adding more proximities demand the use of visual representations for an extra dimension. When two proximities are visualized,
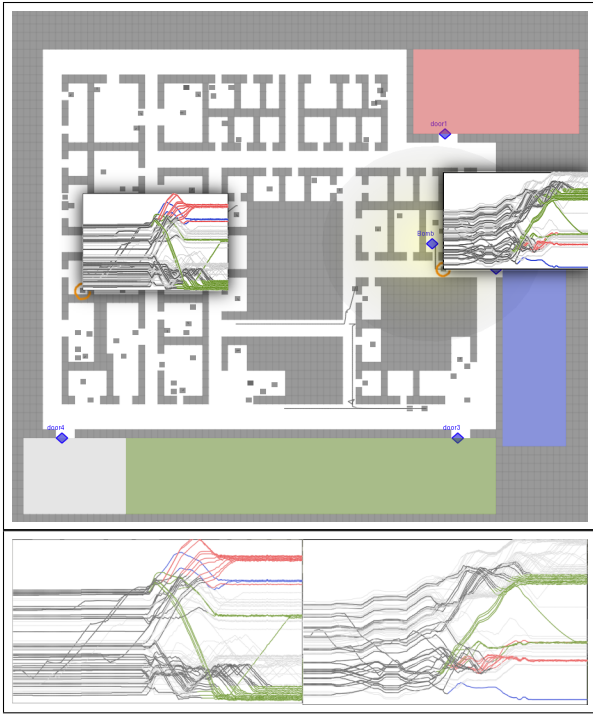
Figure 3: Object instancing. We can select an individual person and find the relative distance of everyone else with respect to him. In this case, we selected two people, the one on the right is a person near the event and the one on the left is a random individual chosen for comparison. To be able to discern the individual lines, the exits were colored.

one can use color to represent the second distance, while retaining the x-coordinate as time. This approach, however, does not extend to multi-dimensional proximities. For a general solution to the case of multiple proximities, we turn to projections.

4.2 Multi-dimensional proximity visualization

When we want to visualize the proximity of moving entities with respect to several points, we cannot longer rely on representing each proximity as its own spatial dimension. If we limit ourselves to 2D visualization spaces, we have two options. One, project a multi-dimensional proximity data point to a single dimension, and retain the other dimension for time. Two, project the proximity data point to a 2D point, and use time as traces. We explore these two in the following sections.

4.2.1 An Interpolation-based Technique

This mechanism projects a multi-dimensional proximity to a 2D point, and represents time as traces. Unlike traditional movement data visualizations, we can arrange the loci of proximities in abstract space and let the data points be arranged around these points. This can be obtained via a parameterization of the space so that the loci of proximity can be arranged freely in 2D. To do this, we turn to *mean value coordinates*, which parameterizes a data point as a function of the distance to a set of proximity centers, that is, for a data point p, the new point q in proximity space can be computed as:

$$q = \frac{\sum_i w_i \hat{c}_i}{\sum_i w_i} \quad (1)$$

where $\hat{c}_i$ are the positions of the proximity centers in the abstract space, and the weights are defined as

$$w_i = \frac{\tan(\alpha_{i-1}/2) + \tan(\alpha_i/2)}{||p - c_i||} \quad (2)$$
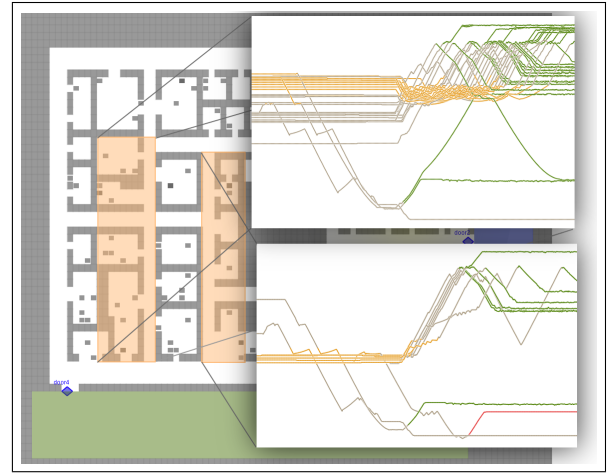

Figure 4: Region instancing. The application allows a region to be selected to be abstracted. In this example, we select two hallways and obtain proximity visualizations to compare the evacuation behavior in different regions. In this case, we see a much bigger congestion on the hallway in the left.

and α_i is the angle between the line segments $\overline{pc_{i-1}}$ and $\overline{pc_i}$ [9], where c_i are the original positions of the proximity centers in the physical space.

Then, a 2D plot is obtained by arranging the proximity points in the abstract space. This allows the user to "stretch" or "compress" entire regions to gain access to detail information that may be hidden due to clutter.

4.2.2 A PCA-Based Technique

Although the previous approach is intuitive, it requires user intervention to make efficient use of the 2D space. A general approach uses Principal Component Analysis (PCA) to reduce the dimensionality of the proximity data while guaranteeing maximum variance in their projections. Principal Component Analysis is a linear transformation that takes a set of n proximity vectors of m dimensions V_i, where $1 \leq i \leq n$, and calculates a new set of basis vectors B_j where $1 \leq j \leq m$. These basis vectors have the property that they are sorted according to which vectors point in the direction that the data has the most variance. That is, the projection of the data onto the first basis vector will have higher variance than the projection of the data onto the second basis vector. Therefore, to reduce dimensionality while preserving the maximum variance, we project the vectors V_i onto the first k basis vectors, where k is the number of dimensions available.

One alternative is to reduce the multi-dimensional proximity vector to a point in a single dimension, and retain time as another dimension in a 2D plot. The resulting plot is similar to the one-dimensional proximity, but the y-axis, instead of representing a distance to a single point, represents the projection of multiple proximities, such that variance is maximized.

Another alternative is to project the proximity data into 2D space, and trade the time dimension for traces. The result is similar to the visualization in absolute coordinates, but PCA ensures that the data points exhibit the largest variance in those two dimensions. This is important for avoiding excessive clutter and inefficient use of the 2D space. Because of the generality of this approach, we are not bound by a small number of proximity centers, but can also visualize the relative distance between every pair of entities, along with a number of fixed points of interest.

An example of 2D proximity visualizations is shown in Figure 5 for a number of trajectories of free-range animals, such as elk and deer [25]. On the left, we see the traditional visualization of deer

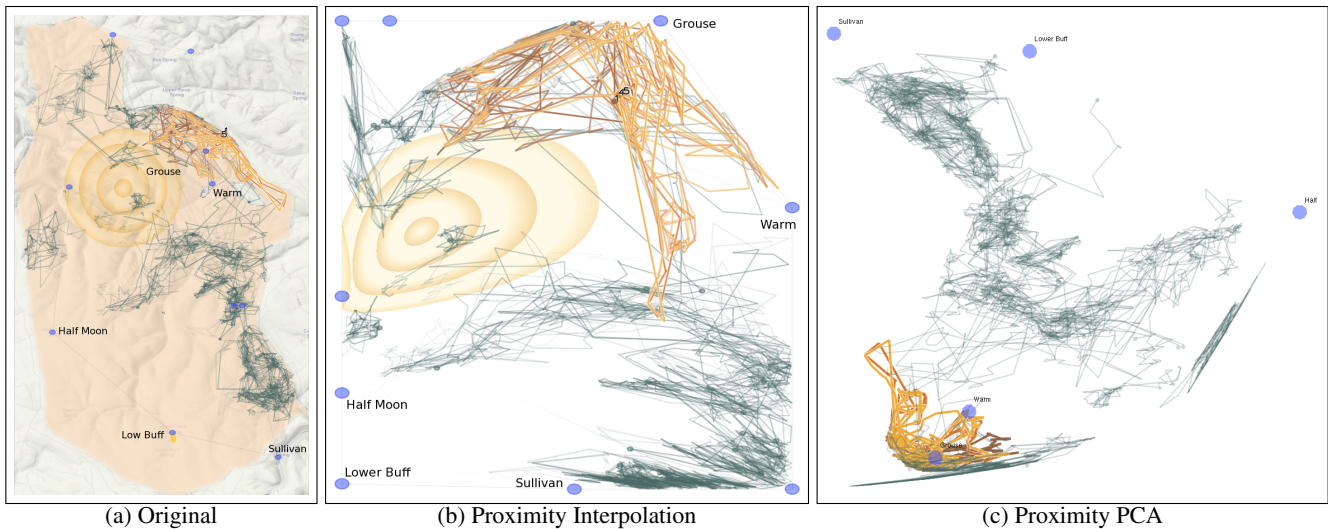| (a) Original | (b) Proximity Interpolation | (c) Proximity PCA |

Figure 5: 2D proximity visualization. (a) Points of interest (blue circles) are placed on the physical space. (b) Proximity visualization with respect to the blue circles using interpolation. The user can "stretch" any part of the abstract space to highlight the trajectories of small regions. In this case, we see three trajectories merging. From (b), it is more clear that one of the three came from the left when merging. This is difficult to see in a traditional space due to the small size of this region (c) PCA-based proximity visualization. In this case, PCA automatically finds a dispersion of the 2D proximity data so that the overall variance is maximized, in the hopes of a better screen use. Here, a larger part of the screen is devoted to the trajectories, especially noticed for the regions in the middle. Compare this to the density in (a)

movement. We follow three trajectories, highlighted with orange, yellow and red. We can see that the particular region is small and the nature of the movement is difficult to see. In the middle, the data is plotted relative to a number of proximity centers representing water springs and other topographical features. Using interpolation, we map this data in a more space-efficient way. By distributing the proximity points, we see a more detailed view of the trajectories of interest. We see that the three trajectories merge, one of them (red) coming from the left, while the other two seem to coincide for a long period of time, suggesting some herding behavior. On the right, we plot the proximities using principal component analysis. When we use PCA, the trajectories are automatically arranged such that the space is better utilized in general. In this case, we can see more clearly the trajectories in the middle of the plot, which in 2D space clutter around small regions.

4.3 Annotation and selection

To aid in understanding the data in proximity space, we have built a prototype application that uses coordinates views of proximity and physical space, and provides coherent operations on both spaces, as described below:

Abstraction Selection. The user can select any level of abstraction for the proximity visualization. This idea builds upon the continuum of representations we have derived. In the most basic abstraction, the user selects a number of entities and enhances their representation with proximity plots. An example is shown in Figure 3 for an evacuation simulation. Here, we see two entities with their own proximity plots (as magnified at the bottom). We see that, for the entity on the left, most people exit through the same route this entity took and assembled in a region with people that evacuated through a different door (green traces). The entity on the right, however, took a route only shared by a single person (blue trace), away from what the majority of people did. Another selection is the proximity visualization of entire regions. In this case, the user "draws" a region of interest in the physical space and a new visualization is created that summarizes the proximity data within that region. An example is shown in Figure 4.

Region Annotation. Although distances are defined in 1D proximity plots in the y-axis, they may not be as intuitive in the 2D visualizations. To overcome this problem, we introduce proximity rings, which define concentric proximity regions from a point of interest in both the physical and abstract spaces. In proximity space, these rings appear as distorted regions, but they still convey the notion of proximity required to obtain accurate measures on the data, as seen in Figure 5.

Entity Annotation. We also let the analyst tag individual entities based on certain characteristics. For example, an analysis of the velocity of movement helps annotate those entities that stopped moving. In proximity space vs. time, these appear as traces that flatline. This type of tagging proved helpful to indicate the casualties in the evacuation simulation and identify any patterns surrounding those deaths.

Another tool is the annotation of entities based on the proximity, inclusion or intersection with a region in space. We let the analyst annotate regions in physical space with different colors and tag the traces (in both proximity and physical space) that are included in that region (so the color "shines" through the object), or that at some point in the past intersected that region (so the color is "stained" on the entity, as proposed by Bouvier and Oates [4]). Examples of this annotation are seen in Figures 3 and 4, where we tag some of the entities based on the exit they took out of the building.

Time Selection. We let the user select any time interval in the data, via an interactive timeline. The timeline itself also represents the trajectories as the first principal component of proximity against time.

5 CASE STUDIES

In order to demonstrate these techniques, we apply them to the evacuation data set from the VAST 2008 contest, consisting of a series of traces from a crowd simulation [13], and the Starkey habitat database for Ungulate research, consisting of traces of free-range animals in a small forest in northeast Oregon [25, 33].

5.1 VAST Evacuation Data

The VAST 2008 contest data set is a collection of traces from a simulation of an evacuation subsequent to an explosive in an of-
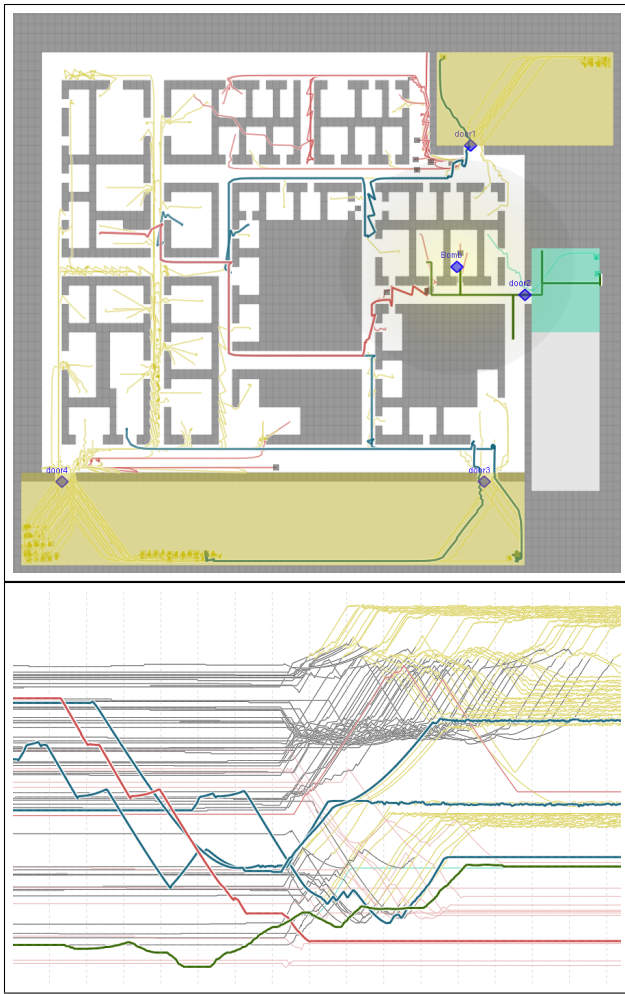
Figure 6: Proximity-visualization of evacuation simulation from the VAST 2008 data set [13]. Here we plot distance to the main event (explosion) vs. time. The green line denotes the main suspect. Highlighted (thicker) lines denote people of interest as they moved before the incident. The clutter in the upper side of the plot illustrates a congestion as lines converge (to one of the exits), and final divergence (yellow lines as they left the building). Red lines denote casualties. We see that most of the later deaths are from people that did not evacuate towards their nearest exit.

fice building. The simulation tracks the movement of 82 people for about 15 minutes. One of the questions when analyzing these data set is to figure out where the bomb was believe to be detonated and characterize the events that unfolded afterwards, including the identification of any casualties, witnesses and suspects. To analyze this data, we turn to both traditional and proximity visualizations. The location of the bomb, a very specific detail, was first obtained by analyzing the traces in a traditional 2D animation. However, discovering patterns and tracking individuals became increasingly difficult by looking at the animation. We attribute this difficulty to the need to track multiple targets, which is known to be increasingly difficult as the number of targets increases (typically beyond 3 or 4 targets). Although some behavior is evident from the animation, such as a couple of casualties following the explosion, it is difficult to *account* for all behaviors that may have ensued. For this reason, we turn to proximity visualization. In this case, we plot distance to the explosion versus time. The result is seen in Figure 6. We see one person, identified as the main suspect, approaching the location of the bomb (presumably planting it) and then leaving (green line).
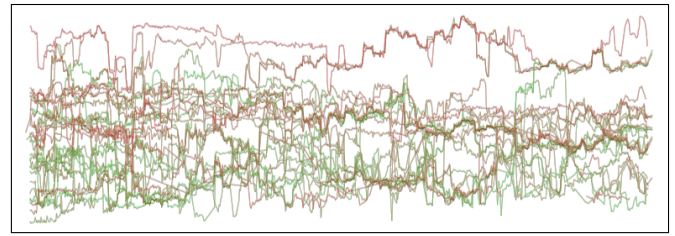
(a) Elk93 Data set
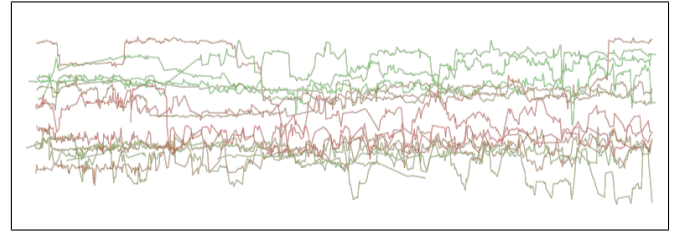
(b) Deer93 Data set
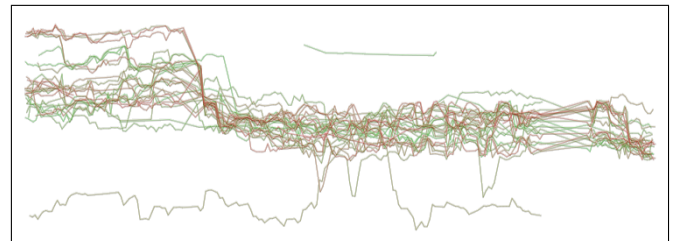
(c) Cattle95 Data set

Figure 7: Visualization of free-range animal movement [25], relative to distance to a water source, and separated per species.(a) Elk, showing convergence patterns. (b) Deer, showing a predominant territoriality, (c) Cattle, exhibiting herd behavior.

We can see when the evacuation takes place as most of the lines are still (horizontal lines) and then begin to disperse towards the nearest exit. This visualization steers our attention to three more lines, which exhibit movement prior to the event, which may suggest witnesses or suspects. By tagging the exit areas (two in yellow and one in teal, for the exit used by the prime suspect), we can get a glimpse of the overall evacuation procedure. We also detect those people that did not exit the building, here identified as casualties (red lines). From the distance of the red lines to the x-axis, we can formulate hypotheses about the cause of death. The nearer ones possibly due to the explosion, while the lines farther away correspond to deaths caused by smoke or fire.

By looking at the proximity plot we find a collection of convergence patterns that correspond to bottlenecks in one of the hallways. Within this bottleneck, we notice a red line, which seem to correspond to a person that could not get out the nearest exist (notice a congestion in the upper side of the plot), and dies when attempting to find another evacuation route. Except for this person, everyone who evacuated through this route made it safely to the outside. The same cannot be said about the upper exit, shown as yellow lines in the middle of the proximity plot. In this region we see a number of red lines, corresponding to casualties. We notice that some of them come from the upper size of the plot, suggesting that they did not pick the nearest exit for evacuation, resulting in death. As a static representation, a number of behaviors can be inferred by looking at proximity that cannot be inferred from the 2D traditional visualization due to the inability to depict time.

5.2 Free-range Animal Data

This data set consists of a series of movement tracks of elk, deer and cattle in a region in northeast Oregon [25]. The data set tracks in-

dividual animals using GIS (geographic information systems) and comprises a series of variables associated with vegetation, water, soils and roads. The goal of this data base is to develop models that help understand the behavior of free-range data in relation to the management of the habitat such as the effects of open roads through the environment. To analyze the behavior of individual groups of animals, we use a proximity visualization that plots distance to one of the water sources vs. time. Three examples can be seen in Figure 7, where we plot three subsets of the data, corresponding to elk movement in 1993 (Elk93), deer movement in 1993 (Deer93) and cattle movement in 1995 (Cattle95). We can see from these plots a series of typical behaviors for each species. Elk appear to roam through the entire reserve, as seen from the *fluctuation* in movement. We see a number of convergence patterns towards the middle of the time span, suggesting mating behavior, and divergence patterns towards the end. Deer, in contrast, do not roam and they appear more territorial. We can see this in the apparent constancy of their distances to the proximity locus. Cattle, on the other hand, exhibit a herd behavior, as seen by the concentration patterns and group trends in their movement. We also see a meet patterns, indicating a convergence point when cattle is herded from place to another, which coincided with a passage through one of the streams in the forest. Overall, proximity plots help us discover *global* patterns associated with each species, and steer our attention towards *local* details such as mating and grouping.

As suggested in one of their own studies, the researchers of this habitat found that elk and deer behave differently in relation to the vehicular traffic through the open roads that cross the reserve [33]. We can visualize this too with proximity plots. In this case, the distance metric is not with respect to a point, but to entire regions, such as open roads. This is depicted in Figure 8, where we show the map of the forest color coded with respect to distance to open roads (black is closest and yellow to white for farthest from the road). When we plot the proximity of elk (top) and deer (bottom) to roads, we see a predominance of elk to stay far from the roads, while deer seem to remain close. We even see a trend where elk move away to less trafficked areas that seems to coincide with a movement of deer towards the trafficked areas. At a glance, this visualization provides us global knowledge about the entire time, while still providing individual detail in the form of traces. This behavior was explained as *disturbance competition*, which suggests that the mere presence of population in a region (the elk away from traffic) annoys another population into leaving the area (the deer, who have no choice but to move towards the traffic areas) [33].

6 DISCUSSION AND FUTURE WORK

Exploring proximity based spaces has led to some interesting observations. We found that mapping points in a two dimensional space into an n-dimensional proximity space and back down to a two dimensional space did not necessarily reveal much new information. This is because the resulting visualization still requires to track movements in a 2D space, while time is only represented tacitly as traces. In fact, the visualization of proximity (or projections of proximity) vs. time seems the most effective. The gain in this new metaphor can be explained thanks to the *cognitive shift* that implies this new type of visualization. Freeing one dimension allows us to plot directly time in a single image, which does not require tracking moving patterns anymore. Therefore, the analyst is not bound to follow a reduced amount of targets, but can formulate hypotheses about a larger number of targets and the entire population. A natural extension, of course, is to extend a 2D proximity plot in a third dimension to account for time. However this implies a different cognitive load that prevent us from transferring the same skills used in a 2D visualization to 3D. For example, users now resort to rotation to disambiguate overlapping trajectories in space and time.

Coordinated views are also fundamental to deploy our tech-

niques into current analysis tools. The traditional visualization in absolute coordinates is a needed view that retains the context of the data. Operations that demand accuracy in terms of the original frame of reference (such as latitude and longitude for time of day or climate changes in a geographical data set), are more effective in their original space. Proximity-based tools, on the other hand, are more effective for operations that are *invariant* to the frame of reference, such as movement patterns, derivatives (velocity and acceleration) or the local impact of the environment (such as topography, proximity to vegetation, food sources or water).

One aspect of the use of proximity data is the pre-computation time. When the proximity loci are known, this transformation is negligible as it consists of a constant number of operations. A more costly operation is when we must compute the proximity of each pair of points, which requires $O(n^2)$ operations, where n is the number of data points. In addition to the cost of computing the proximity, PCA introduces an additional cost. For extremely large data sets, this may become prohibitive. One alternative is to explore the proximity continuum to find projections locally in regions of interest (in both space and time), without requiring to analyze the entire collection of points.

While we have applied the techniques in this paper to the tracking of people or animals, our techniques can be generalized to abstract motion. For example, we can define an abstract space for the organization of documents, such that documents lie near each other if they have some topic or context in common. The more similar two documents are, the closer they appear in this abstract space. Movement tracks can be associated with changes to the document. Therefore, emerging patterns in proximity space reveal trends in the evolution of a document collection.

7 CONCLUSION

We have presented a general notion of proximity-based visualization for the analysis of movement data. We have shown that derived quantities, such as proximity to a point or region, implies a dimensionality reduction that allows us to incorporate time as another dimension in a 2D plot, revealing global and local patterns difficult to see through traditional means. Through a number of examples we see that, despite exhibiting interesting patterns in space-time, proximity-based visualizations imply a change in the frame of reference, which may be difficult to comprehend in isolation. For this reason, we believe that these metaphors are better deployed in redundant coordinated views, where annotations and operations in proximity-space should be reflected in the original view of the physical space, and vice versa. The abstraction of spatio-temporal data into proximities proves to be a potential tool for discovering patterns that may be missed by looking at an animation, or that may be obscured due to clutter. Other derivative information, such as velocity and acceleration, are some of the dimensions that can complement proximities towards more effective analysis.

REFERENCES

[1] G. Andrienko, N. Andrienko, and S. Wrobel. Visual analytics tools for analysis of movement data. *SIGKDD Explor. Newsl.*, 9(2):38–46, 2007.

[2] M. Ankerst, D. A. Keim, and H.-P. Kriegel. Circle segments: A technique for visually exploring large multidimensional data sets. 1996.

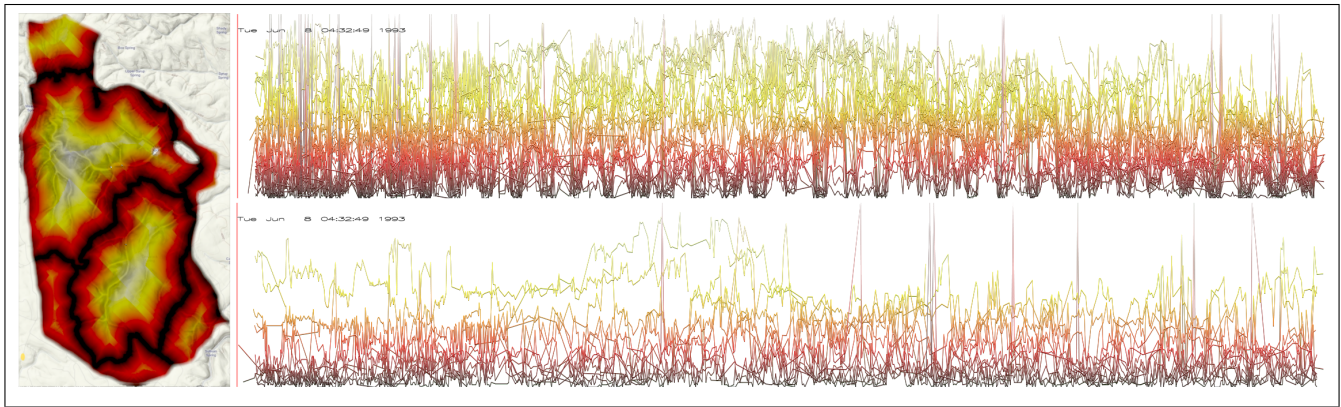[3] P. Berkhin. Survey of clustering data mining techniques. Technical report, Accrue Software, San Jose, CA, 2002.

Figure 8: Visualization of movement traces with respect to traffic conditions. Here, we plot the animal traces for elk (top) and deer(bottom) in 1993 as the distance to the closest open road (color coded using a heat map, where black is closest to road and yellow to white is farthest). The distribution of this distance is seen on the left in the actual map of the forest. We see that elk increasingly move away from traffic while deer seem to stay close to traffic. This behavior can be explained as "disturbance competition", which suggests that the presence of a population in one area (elk) annoys another population into leaving that area (deer, which have to move away from elk into the high traffic areas).

[4] D. Bouvier and B. Oates. Evacuation traces mini challenge award: Innovative trace visualization staining for information discovery. In *Visual Analytics Science and Technology, 2008. VAST '08. IEEE Symposium on*, pages 219–220, Oct. 2008.

[5] D. B. Carr, R. J. Littlefield, and W. L. Nichloson. Scatterplot matrix techniques for large N. In *Proceedings of the Seventeenth Symposium on the interface of computer sciences and statistics on Computer science and statistics*, pages 297–306, 1986.

[6] J. Carroll and P. Arabie. Multidimensional scaling. *Annual Review of Psychology*, 31:607–649, 1980.

[7] S. Dodge, R. Weibel, and A.-K. Lautenschütz. Towards a taxonomy of movement patterns. *Information Visualization*, (7):240–252, 2008.

[8] J. Dykes and C. Brunsdon. Geographically Weighted Visualization: Interactive Graphics for Scale-Varying Exploratory Analysis. *IEEE Trans. on Visualization and Computer Graphics*, 13(6):1161–1168, 2007.

[9] M. S. Floater. Mean value coordinates. *Comput. Aided Geom. Des.*, 20(1):19–27, 2003.

[10] D. Guo, J. Chen, A. M. MacEachren, and K. Liao. A visualization system for space-time and multivariate patterns (VIS-STAMP). *IEEE Trans. on Visualization and Computer Graphics*, 12(6):1461–1474, 2006.

[11] M. Harrower. A look at the history and future of animated maps. *Cartographica*, 39(3):33–42, 2004.

[12] S. Havre, E. Hetzler, P. Whitney, and L. Nowell. ThemeRiver: Visualizing thematic changes in large document collections. *IEEE Trans. on Visualization and Computer Graphics*, 8(1):9–20, 2002.

[13] http://www.cs.umd.edu/hcil/VASTchallenge08/. IEEE VAST 2008 challenge, 2008.

[14] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proc. IEEE Visualization '90*, pages 361–378, 1990.

[15] Y. Ivanov, C. Wren, A. Sorokin, and I. Kaur. Visualizing the history of living spaces. *IEEE Trans. on Visualization and Computer Graphics*, 13(6):1153–1160, 2007.

[16] E. Kandogan. Visualizing multi-dimensional clusters, trends, and outliers using star coordinates. In *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 107–116. ACM, 2001.

[17] T. Kapler and W. Wright. Geo time information visualization. *Information Visualization*, 4(2):136–146, 2005.

[18] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, Sep 1990.

[19] M. J. Kraak. The space-time cube revisited from a geovisualization perspective. *Proc. of the 21st International Cartographic Conference*, 1988.

[20] P. Laube and S. Imfeld. Analyzing relative motion within groups of trackable moving point objects. In *GIScience '02: Proceedings of the Second International Conference on Geographic Information Science*, pages 132–144, 2002.

[21] A. M. MacEachren. *How Maps Work: Representation, Visualization, and Design*. The Guilford Press, 2nd ed. edition, 2004.

[22] H. Moellering. The potential uses of a computer animated film in the analysis of geographical patterns of traffic crashes. *Accident Analysis and Prevention*, 8(4):215–277, 1976.

[23] D. J. Peuquet. Making space for time: Issues in space-time data representation. *Geoinformatica*, 5(1):11–32, 2001.

[24] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman. Lifelines: visualizing personal histories. In *CHI '96: Proc. of the SIGCHI conference on Human factors in computing systems*, pages 221–228. ACM, 1996.

[25] M. M. Rowland, P. K. Coe, R. J. Stussy, A. A. Ager, N. J. Cimon, B. K. Johnson, and M. J. Wisdom. The starkey habitat database for ungulate research: construction, documentation, and use. Technical report, Gen. Tech. Rep. PNW-GTR-430. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, 1998.

[26] C. Tominski, J. Abello, and H. Schumann. Axes-based visualizations with radial layouts. In *SAC '04: Proceedings of the 2004 ACM symposium on Applied computing*, pages 1242–1247. ACM, 2004.

[27] C. Tominski, P. Schulze-Wollgast, and H. Schumann. 3d information visualization for time dependent data on maps. In *IV '05: Proceedings of the Ninth International Conference on Information Visualisation*, pages 175–181, 2005.

[28] E. Tufte. *Envisioning information*. Graphics Press, Cheshire, CT, USA, 1990.

[29] A. Unwin and H. Hoffman. New interactive graphics tools for exploratory analysis of spatial data. In *Innovations in GIS 5: selected papers from the fifth National Conference on GIS Research UK (GIS-RUK)*, pages 46–55, 1997.

[30] I. Vasiliev. Mapping time. *Cartographica*, 34(2):1–51, 1997.

[31] F. B. Viégas, M. Wattenberg, and K. Dave. Studying cooperation and conflict between authors with history flow visualizations. In *CHI '04: Proc. of the SIGCHI conference on Human factors in computing systems*, pages 575–582, 2004.

[32] C. Weaver. Cross-dimensional visual queries for interactive+animated analysis of movement. In *Workshop in Geospatial Visual Analytics*, 2008.

[33] M. J. Wisdom, N. J. Cimon, B. K. Johnson, Garton, E. O., and J. W. Thomas. Spatial partitioning by mule deer and elk in relation to traffic. *Trans. of the 69th North American Wildlife and Natural Resources Conference*, pages 509–530, 2004.