A Real-time 3D Visualization Framework for Multimedia Data Management, Simulation, and Prediction: Case Study in Geospatial-temporal Biomedical Disease Surveillance Networks

[1]Nathaniel Rossol, [1]Irene Cheng, [2]Iqbal Jamal, [3]John Berezowski and [1]Anup Basu
[1]*Computing Science Department, University of Alberta, Canada*
[2]*AQL Management Consulting Inc., Edmonton, Canada*
[3]*Food Safety and Animal Health Division, Government of Alberta, Edmonton, Canada*

ABSTRACT

Geographic Information Systems (GISs), which map spatiotemporal event data on geographical maps, have proven to be useful in many applications. Time-based Geographic Information Systems (GISs) allow practitioners to visualize collected data in an intuitive way. However, while current GIS systems have proven to be useful in post hoc analysis and providing simple two-dimensional geographic visualizations, their design typically lacks the features necessary for highly targeted real-time surveillance with the goal of spread prevention. This paper outlines the design, implementation, and usage of a 3D framework for real-time geospatial temporal visualization. In our case study using livestock movements, we show that our framework is capable of tracking and simulating the spread of epidemic diseases. Although the application discussed in this paper relates to livestock disease, the proposed framework can be used to manage and visualize other types of high-dimensional multimedia data as well.

Keywords: multimedia data management and visualization, biomedical multimedia simulation and prediction, multimedia forensics, social network analysis

INTRODUCTION

Multimedia data streams used in real world applications have been growing at an astonishing rate. Often, cross data streams are collected from different sources without obvious correlations. Integration typically generates data of high dimensionality, which needs further processing to be useful. Although computer systems and artificial intelligence techniques can assist data processing and, to a certain extent, replace human intensive procedural tasks, incomplete data records and insufficient input from domain experts still pose challenges and make it impossible to implement a fully automatic simulation and prediction system. The goal of our visualization framework is to present cross data streams collected from multiple sources in an interactive and intuitive format, so that domain experts can navigate within the visualization space and identify trends that would not have been discovered using traditional analysis techniques. As a result, data record irregularities can be rectified and important expert input can be built in as system parameters, so that sophisticated machine learning and data mining algorithms can be applied.

In addition to simulating, tracking, and predicting diseases, our framework can be applied to analyze financial data, e.g., currency and interest rate movements, which have important impacts on economies. In this case, the visualization network parameters include the flows of import and export, national and domestic consumer indices. The duration of political instability, natural and seasonal disasters, e.g., hurricanes and other storms can be additional factors affecting the network flows. Another example would be tracking the source-destination and responsiveness of message exchanges in social webs. The visualization framework can also be used to improve traffic movements, and thus city planning. Traffic patterns, which can be denoted by the number of vehicles passing a certain bottleneck, can be plotted along a time scale. The time of the day and day of the week, as well as holiday seasons, can be used as filtering parameters to visualize selected data sets. Residential dispersion can be used to refine the analysis further. Consumer preferences and purchasing patterns are important analytic criteria in grocery chains. Identifying associations between product sale quantities and other factors, such as weather, holiday seasons, pay days, etc., helps stocking and pricing. In general, the proposed framework is useful in modeling flows over geographical areas, communication networks, business networks and human disease spread. Note that one common characteristic among these applications is that data or information flows can be modeled using *nodes* and *edges*, and a value can be assigned to each edge to reflect the probability of occurrence.

Although we use geospatial-temporal biomedical disease surveillance in our case study, the visualization framework can be applied in many applications. Visualization helps the understanding of data by making use of our human visual system's highly tuned ability and visual encoding, e.g., position, size, shape and color; to recognize patterns, identify trends, and discriminate exceptions, as well as making data more appealing to engage a larger audience in exploration and analysis (Heer, Bostock, & Ogievetsky, 2010). Observation and simulation of flows and trends are keys for proving scientific theories and discovering facts that the human brain would otherwise never imagine (Ailamaki, Kantere, & Dash, 2010). In response to these application demands, constant improvements of observational instruments and simulation mechanisms such as our visualization framework are launched in scientific research.

The rest of the paper is structured as follows: the Related Works Section reviews visualization and network analysis techniques; the Background Section introduces our case study and disease surveillance; the Design Section and the Implementation Section describe the challenges and technical details, as well as feedback from a case study. The final sections include the conclusion and describe future work.

RELATED WORKS

Visualization tools have become attractive to users because of their graphical presentations which often have higher expressive powers than simple text, images and videos alone. Properly structured graphics can convey the underlying associations between otherwise seemingly independent elements, and can perceptually attract more attention from the viewers (Compieta, Di-Martino, Bertolotto, Ferrucci, & Kechadi, 2007). The idea is to highlight possible associations between the displayed data and alert the viewers, if necessary, for more in-depth examinations, possibly through data mining or clustering (Bailey-Kellogg, Ramakrishnan, & Marathe, 2006). Due to the wide variations in design, implementation and even original purpose of the most closely related systems, there is no well defined state-of-the-art system by which to make a comprehensive comparison with our framework. Instead, this section describes various approaches similar to our framework, and then compares the unique features among them.

GeoTime (Proulx et al., 2007) is one such commercial analytics tool designed for processing temporal geospatial information. This tool takes in sequential event data and displays it over a map, with the vertical dimension indicating progression of time. GeoTime imports data using Excel, has a high degree of customizability, and facilitates detection of interesting events.

Another approach to deal with highly dimensional data is to arrange it into a hierarchical structure. Systems such as n-Vision (Beshers & Feiner, 1990) split the highly dimensional data space into a group of simpler, lower-dimensional subspaces. These subspaces are then arranged in a hierarchy to facilitate comprehension, although the ease of understanding in such visualizations naturally depends quite closely on the nature of the data and the partitioning used to generate the subspace hierarchy.

Not all visualization techniques are focused on just visualizing the raw data of interest. Some visualizations deal with high dimensional data by first processing it before visualization (Ferreira de Oliveira & Levkowitz, 2003). For example, some systems perform clustering or some other data mining task, and then visualize the corresponding processed data instead of the actual raw data itself (Gross, Sprenger, & Finger, 1995). While this approach aids in comprehension by breaking down large raw datasets into a more meaningful and structured form, the issue remains that the data pre-processing step might have removed or obfuscated patterns that were in the raw data that would have otherwise been of great interest or importance to the domain experts using the system.

Visual analytics (Lawton, 2009) has generated growing interest in the research community and in industry. An important contribution to the field of visual analytics was made by the National Visualization and Analytics Center in the United States (Thomas & Cook, 2005). Their approach focuses on information visualization relevant to homeland security and their research includes the state-of-the-art in the field, as it details the needs and expectations of stakeholders in order to map out a vision of future development. Visual analytics has four target areas: analytical reasoning techniques, visual representation and interaction techniques, data representation and transformation, and finally, production and presentation (Thomas & Cook, 2005). Such analytical reasoning techniques maximize the human capacity to perceive, understand, and reason. Significant research was done on the analysis of data that can be represented in terms of networks, investigation techniques such as graph partitioning, reorder-able matrix representations, and flow maps (Guo, 2007).

Other related work includes the development of EpiScanGIS which is a GIS-based research prototype designed to detect and visualize epidemics in order to aid decision-makers enact control measures (Reinhardt et al., 2008). It is similar to our proposed framework and implementation in that it is designed to visualize multiple layers of data in the same geographic visualization space, and can animate these visualizations within an adjustable time window in order to observe how they interact or change over time. In contrast to our proposed framework, EpiScanGS is currently designed primarily with human health in mind, and thus does not use network analysis or network visualization due to the inherent logistical realities of attempting to obtain a complete and accurate social network graph for a large human population. The system also does not provide a mechanism for simulating disease spread or explicitly tracing epidemics back to their possible source(s).

Joeseph Burns et al. (2006) in collaboration with the University of Michigan used ArcObjects (available online at: http://edndoc.esri.com/arcobjects/8.3/) to implement a GIS for epidemic disease surveillance in human populations. Although their system is similar to this paper's proposed framework in that it is designed to be capable of modeling network connections (in this case road networks) to visualize and even simulate disease spread over time, unlike this paper's proposed framework, the ArcObject's-based GIS does not allow for multiple visualization layers to be rendered and animated in the same space, and still does not provide any explicit mechanism for tracing disease propagation backwards through networks in order to identify potential sources. Also, much like EpiScanGIS, the ArcObjects GIS employs visually limited 2-dimensional visualizations, and is designed primarily with human epidemiology in mind.

In order to study the complex interactions that led to disease spread in the 1878 Yellow Fever epidemic of New Orleans, a 3-Dimensional GIS was implemented that was capable of geospatially mapping recorded mortalities over the Urban geography of New Orleans (Curtis, 2008). The system is capable of filtering datasets based on cultural groups and as a result patterns of disease spread through cultural clusters became quite apparent. The system is similar to this paper’s proposed framework in that it uses a 3D GIS visualization approach to visually enhance understanding beyond what is normally possible with only a standard 2D GIS visualization. The system also considers factors beyond simple proximity in determining disease spread, although it does not use any explicit network-based approach in either disease modeling or visualization, and simulation through such networks is therefore not part of the design, rather, the system is designed specifically for post hoc analysis and not real-time use by practitioners.

Related work was also recently achieved by M. Bigras-Poulin et al. (2006) who used networks to analyze the Danish cattle industry in order to examine the risk for disease spread. The approach was similar to the framework proposed in this paper as cattle operation premises were represented as nodes or vertices on the graph, and cattle movements or trade between premises served as connections or edges for the graph in their model. Although the authors were able to perform useful analysis with this approach, they stopped short of proposing or implementing a software framework suggesting how such data may be gathered and visualized in real-time to aid in epidemic disease management.

Data representation and transformation for more effective analysis is also an active area of research. For instance, in 2008, Heath, Vernon, &Webb investigated the temporal structure of UK cattle movement data using a network model in which nodes represent movements instead of locations, and links between nodes indicate sequences of movements. Their subsequent analysis demonstrated that this approach was especially effective in analyzing the effect of transmission rates on disease spread. However, once again, the authors stop short of proposing any multimedia software framework for visualizing or managing the data.

Datasets involved in disease surveillance are often quite large, and can typically involve data that are highly dimensional. Expressing large volumes of these data in a way that is understandable, meaningful, and from which useful knowledge can be extracted can often be a challenge (Reinhardt et al., 2008). As a result, GIS-based methods have become an increasingly popular choice for analysis because they offer a highly intuitive means by which to examine and visualize disease surveillance data (Boulos, 2004; Burns et al., 2006; Curtis, 2008; Janies et al., 2007; Jerrett et al., 2003; Reinhardt et al., 2008; (Zeng et al., 2004).

BACKGROUND OF DISEASE SURVEILLANCE

Geographic Information Systems (GISs)

A biomedical GIS visualization typically displays a variety of health surveillance data over a geographical overview of the area of interest. The data rendered on top of the geographical map can be as simple as two-dimensional shapes representing various important disease events or clusters (Reinhardt et al., 2008) or as complex as three-dimensional height-maps representing disease mortality rates within an urban environment (Curtis, 2008). Regardless of the specific visuals used to display the data, the key is that the visuals are spatially aligned with the underlying geography of the region in some meaningful way. This approach allows for non-obvious complex interactions, such as the manner by which environmental, climatic, or other spatial factors affect epidemic behaviours, to be more easily viewed and understood (Jerrett et al., 2003).

Network-based Approach to Modelling Disease Spread

Although traditionally epidemics in human populations were largely modeled via spatial and/or proximity measures, recent research has shown great promise in applying a network-based modeling approach through creating and analyzing a Disease Transmission Network (Luke & Harris, 2007). Network-based models have proven to be particularly effective in modeling the spread of human sexually transmitted diseases such as Syphilis and Human Immunodeficiency Virus (HIV) through social networks given that such diseases have more to do with contacts in one's social network rather than simple proximity (Rothenberg et al., 1998). Further research has concluded that even in the case of highly contagious diseases, social and cultural networks still play a large role in determining the risk of exposure (Curtis, 2008; Dunham, 2005).

Similarly, research in the field of veterinary medicine has highlighted the merits of a network-based visualization approach to modeling disease spread. For example, research into the United Kingdom foot-and-mouth disease epidemic of 2001 revealed that the social business network between farms may be more important than simple proximity in determining disease spread, and that a lack of understanding of the livestock movement network at the time was a factor in the difficulties encountered whilst attempting to halt the spread of disease (Lawson & Zhou, 2005; Shirley & Rushton, 2005). Since then, networks and graph theory when applied to livestock movement data, have proven to be useful in the analysis and understanding of disease transmission in real-world animal movement networks (Bigras-Poulin, Thompson, Chriel, Mortensen, & Greiner, 2006; Shirley & Rushton, 2005; Staubach, Schmid, Knorr-Held, & Ziller, 2002).

Challenges and Motivations

Similar to other surveillance systems, the challenges of interpreting disease data are mainly due to the high data dimensionality and many hidden patterns. Technical knowledge combined with professional experience is necessary in order to understand and analyze the associations between data streams and predict future trends.

The current process for active livestock disease surveillance and epidemic prevention is tedious and often involves experts searching through archived records, many of which may not exist in consistently formatted digital databases. Furthermore, any suspected livestock disease outbreaks are typically only communicated to affected farms directly via telephone. However, this highly-manual procedure is not feasible when mega- to terabyte sized data management is involved. Moreover, the data formats range from manual records to video recordings and even satellite imagery. This significantly increases the complexity in data analysis. Association rules need to be applied before transforming the diversified data into meaningful information comprehensible to the users. An effective data visualization framework for domain experts thus plays an important role in evaluating multiple cross-correlated data streams, and reporting possible “suspicious events” for further investigation.

Many government agencies already actively update and maintain large disease surveillance and livestock movement databases (Zeng et al., 2004) so therefore, the potential exists to leverage online multimedia systems to aid in disease tracking and management. Naturally, these databases by themselves simply provide vast repositories of information and are not inherently useful without some form of analysis framework. Knowledge is typically gathered in a post hoc fashion by most current GIS applications as the system is used to find geospatial and/or temporal patterns and correlations well after the date on which the data were gathered (Curtis, 2008; Janies et al., 2007; Shirley & Rushton, 2005). Thus, the desire remains for a GIS framework that allows for disease surveillance practitioners to visualize, simulate, and predict the spread of disease through livestock movement networks in real-time. Real-time software tools to track diseases backwards to their potential source(s) are of particular interest.

SYSTEM DESIGN

In collaboration with the Alberta Veterinarian Surveillance Network, a framework was designed and a GIS-based system was implemented. The framework offers a uniquely targeted combination of features that distinguishes it from existing GIS systems in many key areas. Major contributions of the system include:

- A flexible graphical GIS interface that allows for 3D display to multiple platforms including web-based and stereoscopic displays.
- The ability to visualize and animate an arbitrarily-sized time window of movements between operations over time.
- The ability to select any time-continuous sub-network of the overall graph.
- The ability to simulate the forward spread over time.
- The ability to trace backwards through time in order to determine event sources and the ability to use different propagation or network prediction models via a flexible plug-in style interface.
- The framework is designed to maximize the ability of domain experts to understand and control network spread in a real-time visualization space.

The proposed framework was implemented and tested using real data provided by the Alberta Veterinary Surveillance Network (AVSN), for emerging diseases of livestock. The results were well received by our industrial partners as a model for future GIS-based online livestock disease surveillance systems.

Design overview

The proposed framework contains several components (Figure 1). The main visualization user interface component is designed to accept input from databases, as well as to dispatch visual information created by users to the databases.

There are two different plug-in style interfaces designed to maximize system flexibility and utility. The Disease Model Interface, allows for any compatible external disease modeller (i.e., a software module built that implements the interface specifications) to be used with the disease simulation features of the system. Likewise, in order to predict and visualize the livestock movement network data beyond what is available, the Network Prediction Interface allows for any compatible external prediction model to be used with the system. For testing purposes, a basic "worst case" disease modeller and a simple prediction model were implemented. The prediction model assumes fixed and repeating annual patterns. Under normal circumstances, only a percentage of contacted livestock are infected. The worst case model assumes 100% infection once in contact. The percentage parameter can be adjusted depending on the choice of modeler.

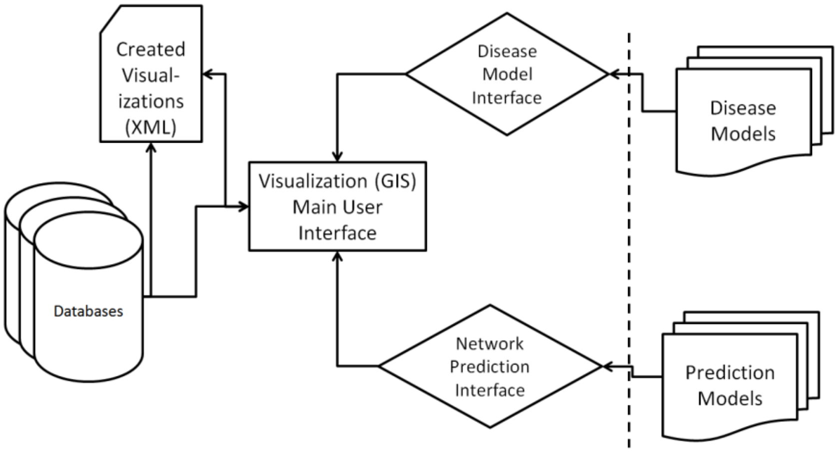


Figure 1. Conceptual overview of the proposed framework.

Since our framework is not designed for one specific application, a flexible plug-in interface allowing different prediction models and adjustable parameter values is important. As mentioned in the Challenges and Motivations subsection, high dimensionality and hidden patterns are two major issues to address. In the following, we will explain how our proposed framework is designed to handle dimensionality reduction and knowledge exploration using our geospatial-temporal visualization schema, and thus can support effective simulation and prediction in a real-time environment.

Database Schema for Geospatial-Temporal Visualization

Data to be displayed by the network visualization component is classified into two tables just prior to data loading: one table for network vertices and the second for edges. The Vertices Table allows an arbitrarily large number of data columns (i.e., dimensions) but the framework requires at least one column which contains a unique identifier (primary key) for the vertex, and two more columns for its X and Y position in the visualization space. Similarly, the Edges Table can support an arbitrarily large number of data dimensions, with a minimum of three for a valid edge: the sender, receiver, and time of transaction. The sender and receiver are identified by their unique identifiers in the Vertices Table governed by the foreign key constraint. The sender, receiver, and time of transaction uniquely identify every row and thus collectively form a primary key for the Edges Table. A fourth, optional data dimension that is specific to the network visualization Edges Table is a column for transaction volume. If left empty, the framework will simply assume the transaction is of unit size (i.e., a size of 1).

Geospatial-temporal attributes play an important role in movement visualization. Our Vertices and Edges Tables design specification precisely captures the characteristics of network-based data visualization.

Reduction of High Dimensional Data

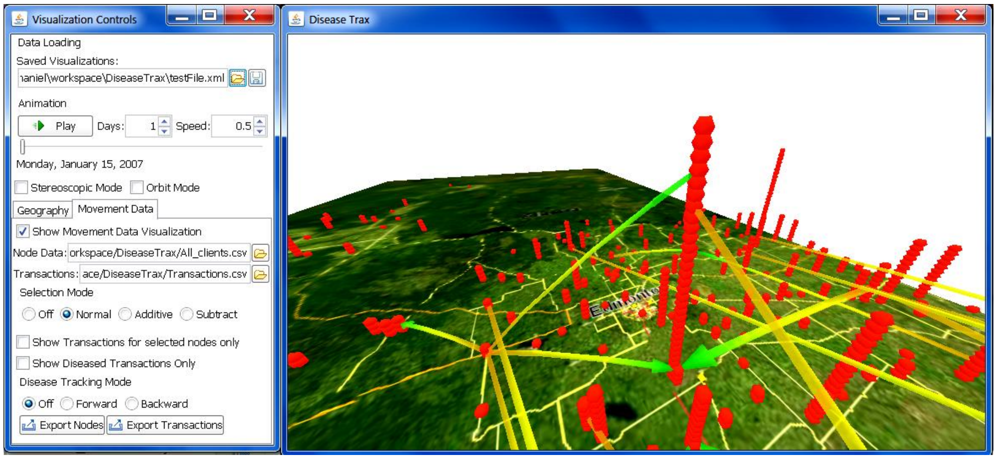
In the proposed framework, the management of high dimensional data occurs in two separate phases: preprocessing prior to visualization and during visualization. Due to the complex nature of integrating the visualization framework with an incomplete existing database, a flexible approach was chosen for the initial data population. Specifically, prior to being loaded by the network visualization component, the data must be pre-processed to create a data view in the simple two-table database schema specification described in the previous section. This specification approach allows for integration into any existing database setup as long as a basic conversion script is implemented to organize the data according to the defined format. A second major benefit of this preprocessing step is that it allows for integration of any other existing data dimensionality reduction solutions, e.g., KNIME (available at: http://www.knime.org), to be used prior to visualization. When implemented in a web-based solution, any preprocessing steps can be configured alongside the visualization interface, and have the new data loaded into the framework on the fly.

The second level of data dimensionality management occurs after the preprocessed data have been loaded. So long as the data loaded adhere to the format specifications, several tools are offered to the user to manipulate the data presentation both temporally and spatially. Reducing the data into manageable views is largely achieved by adjusting the size and position of a sliding window to only visualize data in the time-domain of interest. Similarly, if the focus is instead on a small spatial region over a large time interval, the framework allows for selections of spatial regions over any time window. In addition to the temporal and spatial mechanisms of manipulating the data presentation, the framework also includes or excludes data based on the results of the network prediction model plug-ins, which make use of the other data dimensions available. The framework is composed of multiple layers. These multimedia visualization layers (e.g., weather data, incident reports, etc.) can be manipulated by similar spatial and temporal tools so as to keep all layers fully aligned and consistent in both space and time.

Knowledge Exploration via Network Visualization

The main GIS visualization user interface is designed to allow users to navigate and explore the movement network data space over time in a geographical context. It is designed to support various visual display options including standard desktop computers, the World Wide Web, and specialized stereoscopic 3D projectors.

In our case study, cattle operations (which include farms, feedlots, slaughterhouses, etc.) are represented as vertices on the movement network (Figure 2), and cattle trade movements between operations are represented as directed edges (i.e., node connections or arrows). The network is rendered in three dimensions to provide different viewing perspectives, as many regions often have several cattle operations clustered closely together. In this framework, such clusters of cattle operations are arranged vertically in the third dimension so that complex transactions between large numbers of operations can be rendered simultaneously in the same visualization space with minimized edge intersections. At any point in time, even during animation of the network, the user may freely adjust their viewing position, angle, and zoom with mouse gestures.



Saved Visualizations: ıaniel\workspace\DiseaseTrax\testFile.xml
Days: 1 Speed: 0.5
Monday, January 15, 2007
☐ Stereoscopic Mode ☐ Orbit Mode
☑ Show Movement Data Visualization
Node Data: orkspace/DiseaseTrax/All_clients.csv
Transactions: ace/DiseaseTrax/Transactions.csv
Selection Mode
○ Off ● Normal ○ Additive ○ Subtract
☐ Show Transactions for selected nodes only
☐ Show Diseased Transactions Only
Disease Tracking Mode
● Off ○ Forward ○ Backward

Figure 2. User interface of the implemented framework with Alberta cattle movements shown in the visualization pane. The user may click on any cattle operation or transaction to bring up a large list of detailed information about it.

The visualization also includes an adjustable "time window" feature allowing the user to change the interval of time for viewing. By default, the cattle movement network for a single day is rendered as this is the finest level of granularity available in the dataset used, but the user may expand the time dimension to include weeks, months, an entire year, or any size in-between. As shown in Figure 3, the weekly view shows clearly the general trend that cattle tend to flow from north to south in Alberta from northern farms and feedlots to slaughterhouses and other businesses which tend to be located predominately in the southern part of the province. Animating this time window over the course of a year reveals other macro-scale patterns such as the seasonal increase of transaction volume during the spring and fall, and the relative lull in transactions (seen as a reduction in network density) during the summer months.

Besides adjusting the time window, the user may also view any desired "sub-network" of the movement network. Hence, selecting a single vertex, and setting the desired time window to one year, essentially allows the user to view exclusively the entire movement sub-network for a single cattle operation for a year. Selecting two vertices will likewise visualize all livestock movements involving the two selected operations for the chosen time window. The selection tool also includes additive and subtractive modes which allow for any vertices to be added or deleted from the selection sub-network at any time. This essentially allows for any arbitrary combination of vertices to form a sub-network of the overall network.

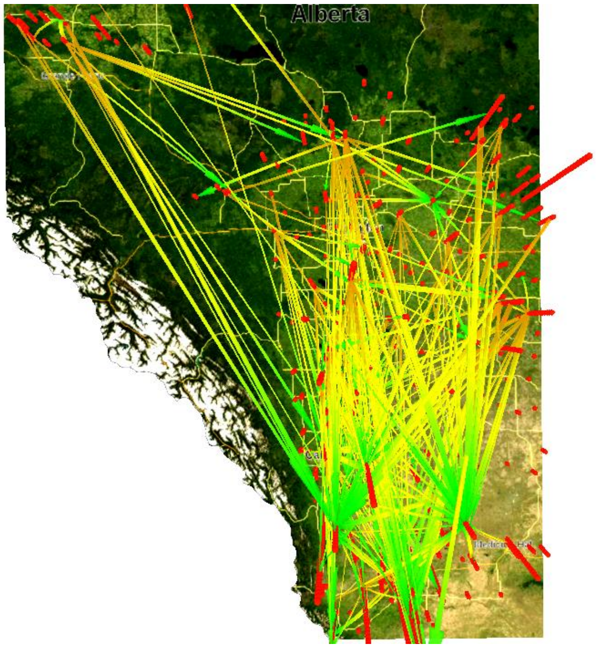
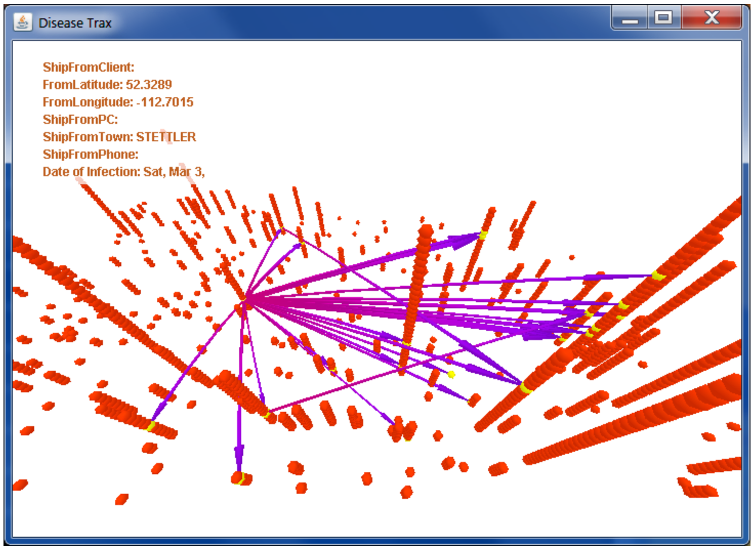


Figure 3. A zoomed-out top-down view of part of the cattle movement network across Alberta for one week in mid-January. Note that arrow heads are rendered in green and tails are rendered in orange, thus cattle movements flow from orange to green.

Propagation Simulation

One of the key features offered by the proposed framework is its ability to visualize a simulation of event spread, and also the ability to automatically track "backwards" in order to identify potential sources. Once the users enter the framework's Forward Disease Simulation Mode, they may select any livestock operations in the region (i.e., vertices on the graph) to mark them as the starting point for an infection. Using the disease model currently plugged into the framework (i.e., different diseases would likely necessitate different models) the framework will calculate how the disease will propagate over time. The infected or potentially exposed livestock operations and transactions will then be highlighted in a separate, visually distinguishing colour. A special visualization mode also exists that exclusively renders only the sub-network of vertices representing those operations marked as potentially exposed to disease by the disease model (Figure 4).



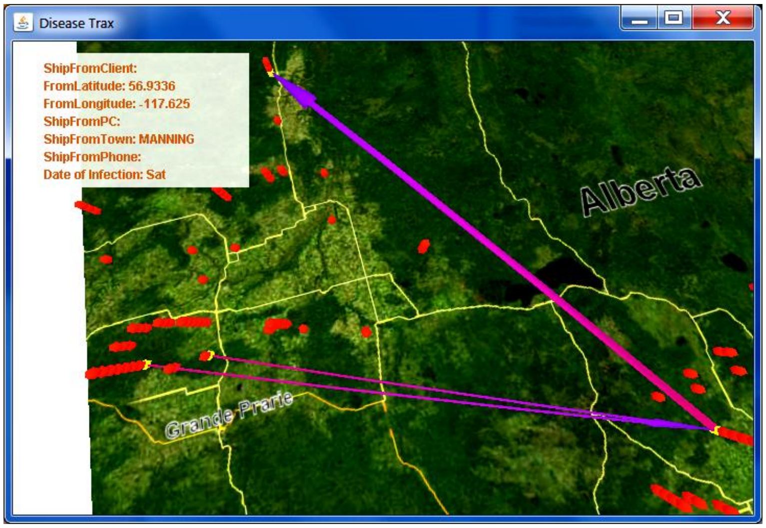
ShipFromClient:
FromLatitude: 52.3289
FromLongitude: -112.7015
ShipFromPC:
ShipFromTown: STETTLER
ShipFromPhone:
Date of Infection: Sat, Mar 3,

Figure 4. An example of the Disease Propagation system in use. The node in the centre-left has been identified as infected, and the system has simulated its probable spread through the network over the given time period according to the disease propagation model currently plugged-in. The option to disable rendering of the geography layers was also selected for visual clarity in this case. Information about the affected nodes can be revealed by clicking on them (personal information was removed from figure for privacy reasons).

The test prototype aims to assist practitioners in understanding the manner in which livestock diseases may behave or spread throughout the region. For example, if a practitioner were to receive information that a serious disease was detected at a certain operation three days ago, using the framework will allow the user to view, in real-time, cattle movements that may have exposed other nearby or even remote operations to the disease. The practitioner may also use previous statistics to model where the disease is likely to spread in the near future. Ideally, this will reduce the risk of herd-to-herd disease transmission by isolating the sub-network of potentially exposed livestock operations (i.e., via quarantine or other measures) from the rest of the livestock movement network.

The Reverse Disease Tracking Mode was specifically identified as being a high priority by the epidemiologist practitioners involved in the research project. When the users enter this mode, they may select any livestock operation in the movement network to mark them as locations where a specific disease has been detected. From these starting points, the system will then run the plugged-in disease modeller “backwards” through time in order to mark all the livestock operations that may be a possible source of infection. For example, in Figure 5, the user has identified the cattle operation in the top-left corner of the visualization as being infected, and has chosen to identify all possible sources of infection within the past week. The system has identified three such locations, which are visually marked close to the bottom of the visualization viewport. Even though no cattle were moved explicitly from the two cattle operations in the bottom left-hand corner to the infected operation near the top, both sent cattle to an operation to the east which then later sent cattle to the infected operation in the north. Thus, all three operations are identified as a potential source of the infection. The practitioner may then click on any of the operations to bring up further details, including contact information, and likewise clicking on a transaction brings up all the relevant details available about the infected shipment.

By identifying multiple infected locations, and looking for overlaps in the locations identified as sources, a user can help narrow down or rank the likely candidates for an infection source.



ShipFromClient:
FromLatitude: 56.9336
FromLongitude: -117.625
ShipFromPC:
ShipFromTown: MANNING
ShipFromPhone:
Date of Infection: Sat

Figure 5. An example of the Reverse Disease Tracking Mode feature in use. The yellow vertex in the top left corner of the viewport was marked as a confirmed infected operation, and the three yellow vertices along the bottom are identified as being potential infection sources.

Network Prediction and Disease Model Interface Design

Both the network prediction and disease model interfaces are implemented by constructing a software module that adheres to the simple input-output specifications of the interfaces.

In the case of the Disease Model interface, the module is supplied three data arrays as input. The first two arrays contain all of the vertices and edges, respectively for the network including any and all extra data dimensions associated with each element. The third array is a list of all vertex-time index pairs that indicate which vertices are affected, and at what time. From these data, the spread models are expected to return the vertex and transactions list with new data columns appended onto them. In the case of the vertex list, each vertex now contains an element indicating what time the vertex was first exposed to the disease in question, or null if it was never exposed. Likewise, each transaction contains an additional data column specifying whether or not the transaction was potentially spreading exposed units. The framework is flexible in the sense that models are optionally allowed to output additional details such as the level of infection of a vertex, the stage of infection, etc., if the disease model computes such details. The disease modeller can be chosen by the user at runtime, and different modellers can be used for different types of diseases to account for differences in epidemic behaviours.

As mentioned previously, for the purposes of the case study, a basic worst-case model was implemented. Given a single starting point or a set of starting points, this model simply computes all the vertices in the graph that can possibly have received transactions either directly or indirectly from one of the starting points. The model iterates over each time step, in ascending chronological order and marks a vertex as exposed if it is a receiver of a transaction where the sender of said transaction has already been exposed. The pseudo-code for the algorithm is outlined below:

Assume *V* is the set of all vertices in the graph, *E* is the set of all edges, and *T* is the set of all time indices in ascending order, and $E(v_1,v_2,t_1)$ represents a directed edge from vertex v_1 to vertex v_2 at time index t_1.

```
for t=0, t < size(T), t++
     foreach v1 in V
          if isExposed(v1)
               then
               foreach v2 in V
                    if exists(E(v1,v2,T(t)))
                         then
                         setExposed(v2, true)
                    end if
               end foreach loop
          end if
     end foreach loop
end for loop
```

In reality, a contacted entity may not be infected. The worst case scenario assumes an infection rate of 100% once contacted. The percentage parameter can be adjusted depending on the model chosen. Disease modeller modules may load any additional model configuration parameters as desired from a configuration file.

The network prediction module interface is almost identical in design to the disease model interface except that in place of the list of vertex/time pairs, it instead takes as input a time value indicating up to what time index the modeller should stop creating the predicted network. The model then returns a list of predicted transactions between vertices up to the time limit specified.

IMPLEMENTATION AND CASE STUDY VALIDATION

The implementation of the framework presented in this paper was written as a Java application using JOGL: the Java Open Graphics Library (available online at https://jogl.dev.java.net/) for 3D visualization. Java was chosen as the programming language due to its cross-platform compatibly, the ease of deployment on the web, and strong flexible interfaces and dynamic library loading ideal for loading external components such as the disease or network prediction models during runtime. JOGL was chosen due to its built-in graphics hardware acceleration available on most major platforms, and its wide, flexible feature-set.

Real-Time Challenges and a Hybrid Caching Solution

Several design challenges needed to be overcome in implementing a visualization framework that could remain sufficiently responsive for real-time interactivity in spite of the large volumes of highly dimensional data present. In particular, the computational costs involved with rendering and animating a large number of transactions concurrently with different time windows was of particular concern, because although the number of vertices is known and constant, the number of transactions can be comparatively much larger. Assuming that there are n vertices in the network graph, and that there are t discrete time-steps, we can see that for each individual time-step there are a total of n^2 possible transactions. Therefore, given a time window of size 1, we have $n^2 \times t$ possible transactions for the entire dataset.

To keep the system performing in real-time, it becomes apparent that for large numbers of transactions, attempting to compute the position, scale, and direction of all 3D transaction meshes every frame, or after every parameter change, will quickly become too costly. Thus, a caching option for the 3D meshes in the visualization scene is required to keep the system responsive. One such caching approach that might be considered would be to pre-cache every scene for every possible time window configuration. However, in terms of the actual number of unique data views that can be created by the user, we can see that the total is actually significantly larger than just the value of t. Given that the upper bound of the sliding time window "u" can be of any value between 0 and t and likewise, the lower bound of the sliding time window "l" can be of any value between 0 and t as long as $l \leq u$, we can see that there are approximately up to $t^2/2$ possible unique views that can be created by the user just by using the time window feature alone. For large timescales involving several years of data, the value of t will be quite large and the space requirements for pre-caching a 3D scene mesh for every possible time view will quickly become unmanageable. This is complicated even further by all the possible unique spatial restrictions the user may configure let alone the additional possible views introduced by the disease or network prediction modellers.

Thus, to solve the dilemma, a hybrid caching solution was chosen whereby the three-dimensional positions, scales, and orientations of every transaction mesh were pre-cached only on a per-transaction basis. Each transaction mesh then contains a single Boolean flag indicating whether or not it should draw itself at the current time index, and these flags are toggled only on affected transactions during a change of a relevant visualization parameter. With this approach, the transaction caching step is linear in the number of transactions, as are the parameter updates, even in the rare worst case where all transactions must have their flags updated all at once.

Case Study and Results

The dataset used in the case study was tested on a standard 2.20GHz dual core personal computer with a 475 MHz graphics processor and 256MB of video RAM. The dataset contained 1,418 vertices with 6 dimensions per vertex, and 218,391 transactions with 14 dimensions per transaction, spread over 365 time-steps whereby each time-step corresponded to a single day of the year. Initial data loading and one-time pre-caching of the entire dataset took approximately 3 seconds to complete, with the majority of the time taken due to the unavoidable hard-disc read time rather than the pre-caching step. During normal operation with daily, weekly, or monthly time-windows (i.e., viewing up to 50,000 transactions at once), the visualization was able to maintain a frame-rate of between 30 to 40 frames per second regardless of other configuration parameters or animation settings. Visualization view updates due to new data or parameter configuration changes were all completed in less than 1 second.

Although the framework proposed in this paper can be used for multimedia data management and visualization in a wide range of applications, we chose disease surveillance of livestock production systems and the detection of possible outbreaks in our case study by examining historic trends and using disease spread models. The complex and high-dimensional multiple cross data streams collected over time and throughout the entire production continuum used in our case study was provided by the Alberta Veterinary Surveillance Network (AVSN), for emerging diseases of livestock.

Although AVSN databases contain ample data, there are few means by which to understand the data or draw conclusions from them. Previous efforts were made by AVSN to use statistical models and data mining in order to help detect alarming events, but these approaches are not useful for understanding and predicting long-term trends and behaviors. The proposed visualization framework was therefore developed and implemented with input and feedback from AVSN veterinarian epidemiologist practitioners for using multimedia data visualization and animation in order to practically and effectively perceive meaningful long-term patterns and trends. A similar computer-based real-time visualization framework is simply non-existent, as advised by the AVSN professionals.

The visualization framework was presented to several groups of scientists and middle managers of the Alberta Food Safety and Animal Health divisions. The senior scientists were interested in using the framework to simulate their assumptions about relationships in the data. From another perspective, the managers could see value for them to better understand these assumptions/relationships and then be able to communicate them to fellow managers and improve their deliberations about impacts of policy decisions. In general, the team of domain experts found the network-based visualization approach of live-stock production system data very helpful in the discovery of underlying patterns and interesting events, which may lead to the understanding of historic trends and the prediction of future disease outbreaks. The forward prediction and backward tracking both proved helpful in providing information for controlling disease spread.

Using the implemented visualization framework, the veterinarian epidemiologist practitioners were already able to make several preliminary discoveries and new knowledge gains:

- A large number of long-standing assumptions regarding the most at-risk regions for infection were validated and confirmed by the visualization framework, and several new smaller high-risk hotspots were shown as well.
- The visualization framework also clearly demonstrated that the overall risk of herd-to-herd secondary infections were surprisingly small as the vast majority of cattle operations appear only to buy or sell cattle, but not do both. Of the operations that did frequently both buy and sell cattle, they typically did not do both types of transactions inside small time windows or sell to a large number of clients all at once. However, whether this is a result of cattle operations failing to report certain transactions remains unknown until new larger and more robust datasets become available.
- In addition to the preliminary discoveries made by the intended use of the framework, a number of additional unexpected benefits of the system also manifested during the case study. Practitioners were quickly able to notice anomalies in the visualization where certain cattle operations appeared to sell large numbers of cattle without showing any source of obtaining them. Further investigation revealed that the operations involved, perhaps due to misunderstanding of proper procedures, had not submitted the proper paperwork for their cattle purchases. In fact, this issue turned out to be quite common amongst several operations across the province. Given the relative ease with which such operations could be identified by the visualization, remedial action could be undertaken immediately.

Overall, the users described the visualization interface as straightforward and easy to understand. As the system continues to be used by the practitioners over time, and as user familiarity increases, additional discoveries and further knowledge gains are expected.

FUTURE WORK

Future works include plans to implement and incorporate additional models into the system, including models for other applications that use geospatial-temporal data for simulation and predication. In addition, there are plans to incorporate additional visualization layers to examine any possible relationship with geography and environmental properties. Lastly, we will enhance the visualization further by distinguishing the various states. For example, “susceptible”, “exposed”, “infected”, “immune”, etc., which in our case study, a herd at a livestock operation might exist in.

CONCLUSION

This paper proposes the design of a multimedia data management and visualization framework for a 3D, online, real-time network-based GIS, and presents a case study where the framework proved effective in aiding veterinarian epidemiologists in understanding herd-to-herd disease spread through livestock movement networks with the ultimate goal of assisting in the management and prevention of such diseases. Through its network based visualization approach, the framework offers unique tools and approaches to movement visualization and simulation including the ability to model where current events will possibly spread in the near future, and predict where a known event may have come from in the recent past. The software implementation of the system was created with an interdisciplinary research team involving computing science researchers, veterinarian epidemiologist practitioners, and an industrial partner. The unique features of the framework are, to the best of the authors’ knowledge and the knowledge of the participating veterinarian epidemiologist practitioners, a novel and useful approach to multimedia utilization in the field of biomedical multimedia computing and multimedia forensics, particularly in the field of disease surveillance. Although the proposed framework was tested with the management, simulation and predication of events and visualization of high-dimensional multimedia biomedical data, it can be readily applied to other similar forms of complex multimedia data as well.

REFERENCES

Ailamaki, A., Kantere, V. & Dash, D. (2010). Managing Scientific Data. *Communications of the ACM*, *53*(8), 88-78.

Bailey-Kellogg, C., Ramakrishnan, N., & Marathe, M. V. (2006). Mining and Visualizing Spatial Interaction Paterns for Pandemic Response. *ACM SIGKDD Explorations Newsletter*. 80-82.

Beshers, C. G., & Feiner, S. K. (1990). Visualizing n-dimensional virtual worlds within n-Vision. *Computer Graphics, 24*(2), 37-38.

Bigras-Poulin, M., Thompson, R. A., Chriel, M., Mortensen, S., & Greiner, M. (2006). Network analysis of Danish cattle industry trade patterns as an evaluation of risk potential for disease spread. *Preventive Veterinary Medicine, 76*(1-2), 11-39.

Boulos, M. (2004). Descriptive review of geographic mapping of severe acute respiratory syndrome (SARS) on the Internet. *International Journal of Health Geographics*, *3*(1), 2.

Burns, J., Hatt, C., Brooks, C., Keefauver, E., Wells, E. V., Shuchman, R., & Wilson, M. L. (2006). Visualization and simulation of disease outbreaks: spatially-explicit applications using disease surveillance data. In *ESRI Users conference*, Retrieved September 24, 2010, from http://proceedings.esri.com/library/userconf/proc06/papers/papers/pap_1588.pdf

Compieta, P., Di-Martino, S., Bertolotto, M., Ferrucci, F., & Kechadi, T. (2007). Exploratory spatio-temporal data mining and visualization. *Journal of Visual Languages and Computing, 18*(3), 255-279.

Curtis, A. (2008). Three-dimensional visualization of cultural clusters in the 1878 yellow fever epidemic of New Orleans. *International Journal of Health Geographics, 7*(1), 47.

Dunham, J. B. (2005). An Agent-Based Spatially Explicit Epidemiological Model in MASON. *Journal of Artificial Societies and Social Simulation, 9*(1), 3.

Ferreira de Oliveira, M. C., & Levkowitz, H. (2003). From visual data exploration to visual data mining: a survey. *IEEE Transactions on Visualization and Computer Graphics, 9(*3) 378-394.

Gross, M. H., Sprenger, T. C., & Finger, J. (1995). Visualizing information on a sphere. In *Proceedings, IEEE Symposium on Information Visualization*, (pp. 11-16).

Guo, D. (2007). Visual analytics of spatial interaction patterns for pandemic decision support. *International Journal of Geographical Information Science, 21*(8), 859-877.

Heath, F. M., Vernon, M. C., & Webb, C. R. (2008). Construction of networks with intrinsic temporal structure from UK cattle movement data. *BMC Veterinary Research, 4*(1), 11.

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A Tour Through the Visualization Zoo. *Communications of the ACM, 53*(8), 59-67.

Janies, D., Hill, A. W., Guralnick, R., Habib, F., Waltari, E., & Wheeler, W. C. (2007). Genomic analysis and geographic visualization of the spread of avian influenza (H5N1). *Systematic Biology, 56*, 321-329.

Jerrett, M., Burnett, R., Goldberg, M., Sears, M., Krewski, D., Catalan, R., Kanaroglou, P., Giovis, C., & Finkelstein, N. (2003). Spatial Analysis for Environmental Health Research: Concepts, Methods, and Examples. *Journal of Toxicology and Environmental Health, Part A: Current Issues, 66*(19), 1783-1810.

Lawson, A. B., & Zhou, H. (2005). Spatial statistical modeling of disease outbreaks with particular reference to the UK foot and mouth disease (FMD) epidemic of 2001. *Preventive Veterinary Medicine, Proceedings of GISVET '04 - GISVET '04, 71*(3-4), 141-156.

Lawton, G. (2009). Users take a close look at visual analytics. *IEEE Computer Magazine, 42*(2), 19-22.

Luke, D. A., & Harris, J. K. (2007). Network analysis in public health: History, methods, and applications. *Annual Review of Public Health*, *28*, 16.1-16.25.

Proulx, P., Chien, L., Harper, R., Schroh D., Kapler, D., Jonker, D., & Wright, W. (2007). nSpace and GeoTime: A VAST 2006 case study. *IEEE Computer Graphics and Applications, 27*(5), 46-56.

Reinhardt, M., Elias, J., Albert, J., Frosch, M., Harmsen, D., & Vogel, U. (2008). EpiScanGIS: An online geographic surveillance system for meningococcal disease. *International Journal of Health Geographics, 7*(1), 33.

Rothenberg, R. B., Potterat, J. J., Woodhouse, D. E., Muth, S. Q., Darrow, W. W., & Klovdahl, A. S. (1998). Social network dynamics and HIV transmission. *AIDS, 12*(12), 1529-1536.

Rothenberg, R. B., Sterk, C., Toomey, K. E., Potterat, J. J., Johnson, D. Schrader, M., & Hatch, S. (1998). Using social network and ethnographic tools to evaluate syphilis transmission. *Sexually Transmitted Diseases, 25*(3), 154-160.

Shirley, M. D. F., & Rushton, S. P. (2005). Where diseases and networks collide: lessons to be learnt from a study of the 2001 foot-and-mouth disease epidemic. *Epidemiology and Infection, 133*, 1023-1032

Staubach, C., Schmid, V., Knorr-Held, L., & Ziller, M. (2002). A Bayesian model for spatial wildlife disease prevalence data. *Preventive Veterinary Medicine, 56*(1), 75-87.

Thomas, J. J. & Cook, K. A. (2005). *Illuminating the Path: the research and development agenda for visual analytics*. National Visualization and Analytics Center.

Zeng, D., Chen, H., Tseng, C., Larson, C. A., Eidson, M., Gotham, I., Lynch, C., & Ascher, M. (2004). Towards a national infectious disease information infrastructure: A case study in West Nile virus and botulism. In *dg.o '04: Vol. 262. Proceedings of the 2004 Annual National Conference on Digital Government Research* (pp. 1-10). Seattle, Washington: Digital Government Society of North America.