COPLINK: Visualization for Crime Analysis

Hsinchun Chen, Homa Atabakhsh, Tim Petersen*, Jenny Schroeder*,
Ty Buetow, Luis Chaboya, Chris O'Toole, Michael Chau, Tom Cushna, Dan Casey*, Zan Huang
University of Arizona, AI Lab
{hchen, homa@bpa.arizona.edu}
http://ai.bpa.arizona.edu/
*Tucson Police Dept.

Abstract

We have applied various visualization techniques to assist crime analysis. These techniques include a hierarchical list, a hyperbolic tree, a time line tool, a periodic pattern tool, a GIS tool and a self-organizing map. We have integrated some of these techniques into a single tool in order to display different and synchronized views of the same data. This tool is flexible and scalable to allow for the incorporation of other visualization techniques in the future. This paper describes the various visualization techniques we have applied in the law enforcement domain and that we are integrating into our ongoing COPLINK project.

1. Introduction

Visualization techniques have been shown to be useful in various domains but have not been widely studied for applications in crime analysis where it is useful to identify the relationships among different entities such as people, vehicles, addresses and organizations. It is critical for crime analysts to retrieve, understand and analyze these relationships efficiently and effectively. There are some existing problems with visualization for crime analysis. Network charts are widely used by crime analysts but these are usually drawn manually. There are very few software applications available for drawing these networks and data exists in tabular format, which makes it difficult for crime analysts to see the big picture in order to draw conclusions easily. We have developed various visualization techniques to integrate into our ongoing COPLINK project (Chen et al., 2001). We describe each of these techniques and how we have incorporated some of them into one scalable visualization tool. We also explain some of the lessons learned when applying these techniques for the Tucson Police Department (TPD).

2. Criminal Relationship Visualization using a Hyperbolic Tree

In crime analysis, law enforcement officials have to process a large amount of criminal data and figure out their relationships. It is important to identify different associations among criminal entities. We proposed the use of a hyperbolic tree view and a hierarchical list view to visualize criminal relationships (Fig. 1) and developed a prototype system called the COPLINK Criminal Relationship Visualizer (Atabakhsh et al., 2001).

The hyperbolic tree is based on hyperbolic geometry (Coxeter, 1965). InXight (a spin-off from Xerox Parc) was first to use the hyperbolic tree for visualization of hierarchies (Lampings, 1995). It provides a convenient way to visualize exponentially growing trees (such as large hierarchies). This is an elegant way to see the big picture and the interesting details at the same time. When utilized to view relationships between different law enforcement entities (people, vehicle, location, organization), the user can search all entities having a relationship with a given search term using the concept space technique (Chen & Ng, 1995, Hauck et al., 2002), and view these relationships in the form of a hyperbolic tree as well as in a hierarchical tree structure.

An experiment was conducted to test the effectiveness and the efficiency of the hierarchical and hyperbolic views. The results show that the hyperbolic tree view is more effective (correctness of the answer) for an "identify" task (e.g. identify a person or a vehicle with a given set of characteristics) and more efficient (time to complete the task) for an "associate" task (find all entities associated with a given person or location). The participants generally thought it was easier to use the hierarchical list, with which they were more familiar (e.g.

hierarchical file systems). When asked about the usefulness of the two views, about half of the participants thought that the hyperbolic tree was more useful, while the other half thought otherwise. Our results indicate that both views can help in criminal relationship visualization. While the hyperbolic tree view performs better in some tasks, the users' experiences and preferences will impact the decision on choosing the visualization technique.

In addition to the hierarchal and hyperbolic views, we have developed other visualization techniques to use in law enforcement and crime analysis. From discussions with our partners at the Tucson Police Department, we have concluded that they would like to have various visualization tools since each tool has its own strengths and each can assist in solving a different type of problem. Other techniques we have experimented with are described in the next sections.

3. Spatio-Temporal Visualization for Crime Analysis

In order to visualize the data needed by crime analysts, we integrate three types of visualization techniques: a periodic view, a timeline view and a GIS view. Each technique has its own strength as follows: periodic visualization displays patterns with respect to time; timeline visualization displays characteristics of temporal data in a linear manner; GIS visualization displays information on a map and allows for spatial analysis of data. We combine these techniques into one tool, the Spatio-Temporal Visulizer (STV), to allow the same data to be examined from three different views simultaneously (Fig. 2).

The STV is a data visualization tool built on top of our ongoing COPLINK project (Chen et al., 2001). COPLINK provides a one-stop data access and search capabilities through an easy to use user interface, for local law enforcement agencies such as the Tucson Police Department (TPD). STV is intended to enhance COPLINK by providing an interactive environment where analysts can load, save, and print police data in a dynamic fashion for exploration and dissemination.

3.1. Design and Functionality

STV is built into a Java applet in a modular fashion. This was done with the intent that other types of views would be added in the future with relatively little work by taking advantage of object-oriented inheritance. One key advantage of an applet is that no software needs to be installed or maintained on analysts' machines. Queries are performed through a servlet to an Oracle database. Results are stored by a controller class and accessed by each STV view. On the backend, JDBC is used to connect to the COPLINK database. STV overcomes some of the disadvantages of other existing crime visualization tools by viewing three synchronized perspectives on the same data.

The control panel maintains central control over temporal aspects of the data. The time-slider controls the range of time viewed. Thus, the data may span six years, but the timeslider may be narrowed to focus on one year, or one month. This time window into the data may then be moved like a typical slider to incorporate new data points and exclude others. This slider was inspired by Lifelines (Plaisant, 1996) and by (Richter, 1999). Granularity, referring to unit of time, is controlled through a drop down menu. Currently, years, months, weeks, and days are implemented. Changing this option has the effect of re-labeling the timeline and altering the periodic patterns being examined. The overall time bounds are controlled through a series of drop down menus. Thus, while all data points may lie in a particular time span, a user can narrow focus to a subset of data based on time bounds.

The main purpose of the periodic view is to give the crime analyst a quick and easy way to search for crime patterns. The periodic view used in STV is similar to the one used in ReCAP (Brown, 1998), but overcomes some of its shortcomings. The circle represents time in the granularity the user chooses. It may represent a year, month, week or day. Within the circle there are sectors dividing it into different time periods within the granularity selected. The analyst also has the ability to change the granularity of the sectors. For example, the circle could be set to year granularity and the sectors could be set to represent months, weeks, or even days. The advantage of this is that the analyst may see different patterns developing over the different time periods. Sectors are labeled to indicate their specific time interval. Data is represented using spikes within each time period. Rings with labels inside the circle represent quantity of data. Using the box plot method a crime analyst can easily determine if any spikes are outliers.

The timeline view is a 2D timeline with a hierarchical display of the data in the form of a tree. A specific time instant may be highlighted. When combined with the current granularity, all points in that time period are highlighted. For example, if the granularity is month and a point in June 1999 is selected, all data in June 1999 are highlighted. The tree view and timeline views of the data are coordinated such that expanding a node in the tree expands the data points viewed on the timeline. At the same time, data under a particular node in the tree is summarized in the timeline at that node's corresponding y-coordinate location. The time-slider controls the current timeframe viewed. This has the effect of allowing the user to slide across the timeline at various levels of detail. The tree view allows the user to see the data in a traditional and organized way.

The GIS view uses ESRI software (http://www.esri.com) and displays a map of the city of Tucson on which incidents can be represented as points of a specific color. The user can zoom in and out of the map. Zooming in allows for more streets to be displayed. Incidents may be selected by dragging a box around points on the map. This will narrow the information being displayed by all views, focusing on the selected incidents. The user can move backward and forward in the zoom history similar to an Internet browser. The view pronounces data points within the time period specified by the time-slider. Data points outside this period are faded. Data points highlighted in the timeline view are highlighted in the GIS view.

Although there are standalone visualization tools similar to each of the components implemented in STV, to the best of our knowledge there are very few tools that incorporate a timeline view simultaneously with a periodic tool as well as a GIS, to provide different and synchronized views of the same data set, as we have implemented in STV. Interaction between the three tools is one of the most important features of STV. Since each view is listening to the same data model, all views are reacting to the same events and settings. For example, when the time bounds are restricted, entities outside the time bounds are faded in the GIS view to represent their decreased importance, while the timeline and periodic views only display data that lies within the new bounds. STV views are visible concurrently, so users will not have to explicitly switch between views. Views can be enlarged and shrunk by use of split panes and scroll panes. Maps, tree structures and timelines appear intuitive to crime analysts who view screen shots of the tool, although no user study has been performed. The spiral view will require more explanation.

3.2. Lessons Learned

Although the STV tool has not yet been deployed at TPD, we have been able to receive feedback regarding the tool from ten TPD crime analysts and from a detective. It is important for the STV tool to be assessed by these sources because it is the detectives and crime analysts who will be the primary users of the STV tool. Comments made by the detective and analysts throughout the initial development are summarized below.

From our first meeting with analysts, the options to load, save and print projects were expressed as high priorities. Once implemented, projects no longer needed to be recreated each time a user logged onto COPLINK. Similarly, the ability to produce a hard copy of information is often very desirable. These functions enable users to more easily incorporate STV into their analysis. Potential users of STV at the TPD have indicated that the ability to expand and constrict the data being displayed is important in searching for different crime patterns. For instance, an analyst may begin with a large number of incidents being displayed and then narrow them down to relevant incidents, or vice versa. They feel that the STV tool does this quickly and efficiently by means of the control panel and the GIS view. The STV tool will also allow police managers, along with the help of analysts, to discuss ongoing problems and trends. The analysts indicated this as an important strength because quite often the Police Chiefs and managers will want to see different aspects of crime trends "on the fly". The potential users have noted that they can easily learn to switch between the three different tools integrated into STV. A final strength that cannot be overestimated is STV's ability to abstract away tedious details of database searches and displays. Computers are excellent at these types of processes. By shifting an analyst's focus from a low level of computer interaction to a much higher level of patterns, causes, and effects of crime, STV increases the efficiency of analysis.

While most of the feedback we received from TPD was favorable, users have indicated certain areas of potential improvement for STV. The biggest concern is the limited customization that the tool currently supports. For instance, crime analysts may wish to add a note and reference it to an incident that is being visualized. They may also wish to add events to the data set that are not present in the databases or to customize the colors and shapes. For example, officers may want to have all robberies displayed by a green triangle, and all homicides displayed by a red circle. Size of data points was also expressed as a concern. A problem common to virtually all visualization techniques is that of labeling. Analysts recommended a variety of labels for data points, from standard text labels to balloon labels that appear on mouse hovers. The size and content of labels were also of interest.

Crime analysts and detectives express a need for having the ability to view the same set of data from different angles using different kinds of visualization techniques. Our Spatio-temporal visualization tool allows for this flexibility by incorporating different visualization techniques into one easy to use interface. We have investigated other visualization tools in addition to the ones described in the previous sections, to integrate into STV. One of these techniques is a self-organizing map described in the next section.

3.3. Crime Visualization using a Self Organizing Map

We have used an artificial neural network based self-organizing map (SOM) developed by Kohonen (Kohonen, 1995) to cluster and visualize crime related data. Kohonen based the neural network on the associative neural properties of the brain. This network contains two layers of nodes: an input layer that represents the object features and a mapping layer in the shape of a two-dimensional grid. The mapping layer acts as a distribution layer to summarize general feature patterns in the collection of objects. Each node in the input layer corresponds to one of the features of an object. Each node in the mapping layer is connected to all input layer nodes with certain link weights. Thus, the mapping layer node can be also represented as a feature vector in which link weights correspond to feature values. The SOM has attracted substantial research interests in a wide range of applications. SOM is an unsupervised learning mechanism that clusters objects having multi-dimension attributes into a lower-dimension space, in which the distance between every pair of objects captures the multi-attribute similarity between them. The input for the SOM algorithm is a set of objects with multiple attributes. Each object is represented as a vector of the attribute values and is referred to as an input in subsequent description.

Our SOM tool (Fig. 3) is opened through a modified version of COPLINK Connect and Detect. Currently, query results performed in Connect are stored into a flat file on the server side. A Java bean gets passed the file from a JSP page. The SOM gets the information from the bean as an input stream. The applet for SOM, separate from the STV applet, processes the information and visualizes the results. Using the SOM, we can cluster crimes based on different attributes (such as male or female criminal, crime type, location, date and time etc.) and view the various clusters on the colored background. Figure 3 illustrates an example of data on bank robberies in Tucson. In this example, the different shades of color represent different months and the clusters represent year (these can be easily changed to other attributes i.e. any data field in the database). The label on each point indicates the crime type but can be easily changed to other attributes.

.

4. Conclusions and Future Directions

Our partnership with the Tucson and Phoenix Police Departments for our ongoing COPLINK project, has shown that law enforcement is one application area that can benefit from better visualization techniques. We have applied various visualization methodologies to crime analysis. These techniques include a hierarchical display, a hyperbolic display, a timeline view, a GIS tool, a periodic pattern tool and a self-organizing map. Through our Spatio-Temporal Visualizer, we have shown that some of these views can be incorporated into one scalable and flexible tool that can display different views of the same data set simultaneously. There are many user interface features that are the wish lists of crime analysts who have seen this tool. For example, customization of color schemes, sorting by any of dozens of fields, and exotic labeling techniques are all seen as important features. In addition, analysts have expressed interest in automatic geographic hotspotting. Other analysis algorithms may also be applied. The Spatio-Temporal Visualization (STV) tool is scheduled to begin user studies at the TPD in March 2003. The plan is to have crime analysts use the STV tool in their daily activities in order to discover other strengths and areas for improvement. The experiences of crime analysts will provide valuable insights into future directions for the STV project. Other modules such as our hyperbolic tree crime relations visualizer and our SOM crime visualizer will also be incorporated into STV and tested at TPD.

References

Atabakhsh H., J. Schroeder, H. Chen, M. Chau, J. Xu, J. Zhang, and H. Bi (2001). "COPLINK Knowledge Management for Law Enforcement: Text Analysis, Visualization and Collaboration," in *Proceedings of the National Conference for Digital Government Research (dg.o 2001)*, Los Angeles, California, May 21-23, 2001.

Brown, Donald E. (1998). "The Regional Crime Analysis Program (RECAP): A Framework for Mining Data to Catch Criminals," *Proceedings from IEEE.*

Chen H., J. Schroeder, R. Hauck, L. Ridgeway, H. Atabakhsh, H. Gupta, C. Boarman, K. Rasmussen and A. Clements (2003). "COPLINK Connect: information and knowledge management for law enforcement," *Decision Support Systems (DSS)*, Special Issue "Digital Government: technologies and practices," 34(3), 271-285.

Chen, H., D. Zeng, H. Atabakhsh, W. Wyzga & J. Schroeder (2003). "COPLINK: Managing Law Enforcement Data and Knowledge," *Communications of the ACM,* pp 28-34.

Coxeter H. S. M. (1965), Non-Euclidean Geometry. University of Toronto Press. 1965.

Lamping J. (1995) A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. CHI proceedings 1995.

Kohonen, T., Self-Organizing Maps, Springer-Verlag, Berlin Heidelberg, 1995.

Plaisant, C., B. Milash, A. Rose, S. Widoff & B. Schneiderman (1996). "Lifelines: Visualizing Personal Histories," *ACM CHI '96 Conference Proceedings.* pp 221-227.

Richter H, J. Brotherton, G.D. Abowd & K. Truong (1999). "A Multi-Scale Timeline Slider for Stream Visualization and Control," *GVU Technical Report GIT-GVU-99-30.*

Acknowledgements

This project has primarily been funded by the following grants:

NSF, Digital Government Program, "COPLINK Center: Information and Knowledge Management for Law Enforcement," #9983304, July 2000-June 2003, NSF, Information Technology Research, "Developing A Collaborative Information and Knowledge Management Infrastructure," NSF/IIS #0114011, Sept. 2001-Aug. 2004. We would like to thank the following people for their support and assistance during the entire project development and evaluation process: all members of the University of Arizona Artificial Intelligence Lab staff and specifically past and present COPLINK team members, The Tucson and Phoenix Police Departments.

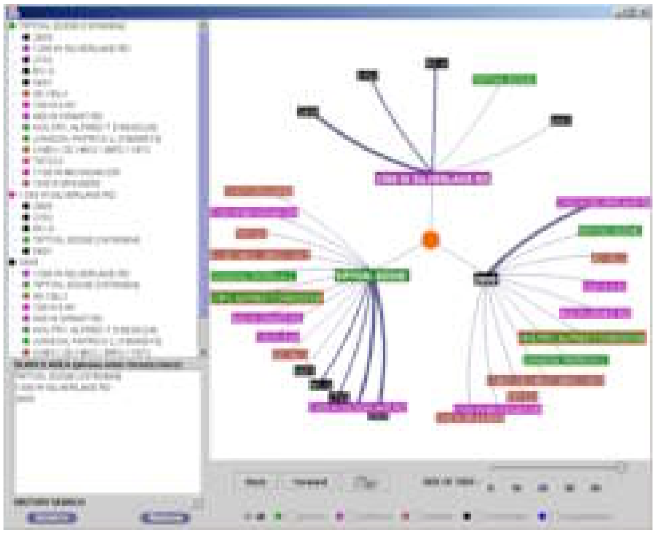

Fig. 1: COPLINK Criminal Relationship Visualizer.
Displays a hierarchical as well as a hyperbolic view of relationships between criminal entities (person, location, crime type, vehicle). The different entity types are color-coded. The hyperbolic view can be rotated and translated if the focus needs to be changed. The slider can be used to reduce the number of nodes displayed to avoid clutter.

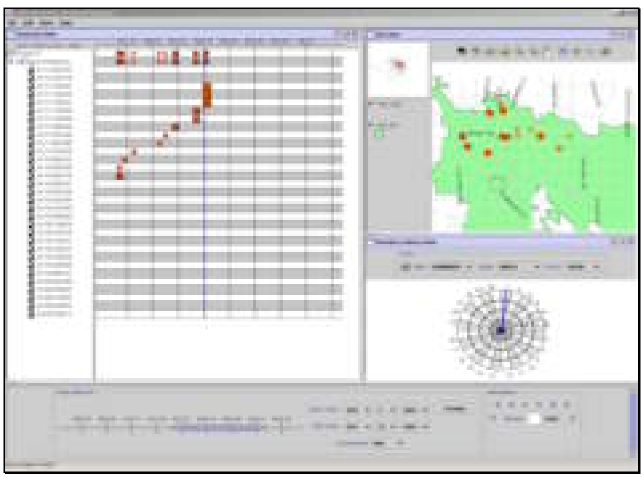

Fig. 2: COPLINK Spatio-Temporal Visualizer for Crime Analysis.
This tool presents a synchronized view of the same data using three different visualization techniques:
a timeline, a GIS and a periodic pattern tool. The focus can be switched easily to any of the three tools by moving the tool to the left pane that has a bigger display area. The time-slider (bottom) controls the range of time viewed. Data points highlighted in the timeline view are also highlighted in the GIS, those outside the time frame selected, appear faded in the GIS.

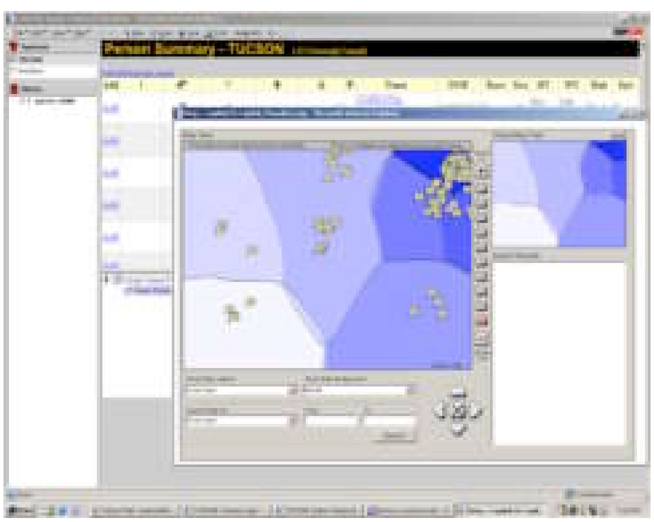

Fig. 3: COPLINK Self Organizing Map, for crime analysis.
Here the SOM is shown in front of the COPLINK Connect/Detect interface which is displaying details about a person selected in the SOM. In this example of data on bank robberies in Tucson, the different shades of color represent different months and the clusters represent year (these can be easily changed to other attributes i.e. any data field in the database). The label on each point indicates the crime type but can be easily changed to other attributes.

.