Two Web-based Spatial Data Visualization and Mining Systems: Mapcube & Mapview

Chang-Tien Lu, Yufeng Kou, Hongjun Wang
Department of Computer Science
Virginia Tech
7054 Haycock Road
Falls Church, VA22043
{ctlu, ykou, howang1}@vt.edu

Shashi Shekhar, Pusheng Zhang, Rulin Liu
Department of Computer Science
University of Minnesota
200 Union Street, SE
Minneapolis, MN 55455
{shekhar, pusheng, rliu}@cs.umn.edu

1. Introduction

The explosive growth of spatial data obtained by government agencies and research institutes has created a need for next generation spatial analysis tools that can automatically transform the processed data into useful information and knowledge. Spatial data mining is concerned with the discovery of interesting and useful but implicit knowledge from spatial data. Visualization is the process of visually exploring data for pattern and trend analysis. Visualization and mining techniques allow organizations and companies to extract practical information from the vast amount of data they have gathered, thus helping them make more effective decisions. Classic statistical analysis methods and data mining techniques are very resource intensive and must be done off-line, not allowing users to interact with the results and change parameters on the fly. To address this problem, we have developed two web-based visualization and mining prototype systems, Mapcube & Mapview, for observing the summarization of spatiotemporal patterns and trends in transportation and census data. By integrating mining techniques and interactive graphic user interface, the spatial patterns and temporal trends can be identified easily and responsively. Mapcube is also designed for browsing the spatial-temporal dimension hierarchy via the integrated roll-up and drill-down operations. In addition, our systems support data visualization in a web-based environment. Users can conveniently access our system remotely with a web browser, thus facilitating the better utilization of spatial information for general public.

2. Mapcube

Mapcube is a high performance spatial visualization system for traffic data. It is a web-based application implemented in Java. Mapcube can be visited through web browser to display highway traffic information. Mapcube system provides several essential data cube visualization utilities for traffic pattern analysis, including traffic video, highway time map, traffic video comparison, and 2-D data cube map. Mapcube can also support spatial outlier detection, identifying abnormal traffic patterns from traffic data warehouse. These traffic patterns and rules can assist decision-making for transportation managers, commuting routes selection for commuters, and traffic model establishment for researchers and planners.

The concept of data cube is the engine behind the Mapcube. A data cube is used to generate the union of a set of alpha-numeric summary tables corresponding to a given hierarchy. Spatial data warehouses prefer browsing aggregated data in terms of albums rather than alpha-numberic summary tables. Extending the concept of data cube, Mapcube organizes the album of generated visualization using a given hierarchy to support browsing via roll-up, drill-down, and other operations on aggregation hierarchy. In traffic data warehouse, the measures are volume and occupancy, and the dimensions are time and space. Dimensions are hierarchical by nature. The time dimension **T** can be grouped into "hour", "date", "month" or "year", which forms a lattice structure. Similarly, space dimension **S** can be grouped into "Station", "County", "Highway", or "Region." Given the dimensions and hierarchy, the measures can be aggregated in different ways.

Figure 1 show the dimension hierarchy of Mapcube. The basic structure is a data cube, where $\mathbf{T_{TD}}$ represents the time of day, $\mathbf{T_{DW}}$ represents the day of the week, $\mathbf{T_{MY}}$, and **S** can represent a station or a group of stations. Each node is a visualization style. For example, the $\mathbf{ST_{TD}}$ node represents the daily traffic volume of each station. The $\mathbf{T_{TD}T_{DW}S}$ node represents the traffic volume at each station at different time of different days. The information is generated as a video. Because of the nature of data cubes, Mapcube can visualize various kinds of traffic data. The software only requires the space and time for each measure, such as volume, occupancy, and speed. These requirements are very simple and most highway monitoring systems should be able to satisfy them. Figure 2 provides an example of nodes in the Mapcube system. The three dimensions are Station **S**, Time of Day $\mathbf{T_{TD}}$, and Day of Week $\mathbf{T_{DW}}$, and the three pictures correspond to the three 2-D nodes, Time of Date–Day of Week ($\mathbf{T_{TD}T_{DW}}$), Day of Week–Station ($\mathbf{T_{DW}S}$), and Station–Time of Date ($\mathbf{ST_{TD}}$).

Traffic data is provided by The Traffic Management Center of the Minnesota Department of Transportation. It archives sensor network measurements from the highway system in the Minneapolis-St. Paul (Twin Cities) metropolitan area. For this demonstration, we use the traffic data for highways in the Twin Cities area from January 1997 to June 1997.

Software Architecture: As shown in Figure 3, Mapcube is a three-tier database application. It integrates GUI, web server and database server to provide a web-based visualization system. GUI is implemented in JAVA applets, accepting inputs from users and sending requests to web server. Implemented in Perl, CGI programs run on web server, accepting GUI request and sending SQL queries to DB server. MySQL is employed as database server to manage traffic data, accepting SQL queries and generating answers.

System Demonstration: Mapcube can display traffic with different color schemes. Darker color indicates heavier traffic flow on the specific highway. We will demonstrate four kinds of traffic visualization utilities as follows.

- **Traffic video:** As shown in Figure 4, users can specify the date, starting time and ending time to display traffic flow on one or several specific highways. Three kinds of traffic information, total volume, average volume, and occupancy, are provided,.
- **Highway time map** ($\mathbf{ST_{TD}}$): As shown in Figure 5, static traffic flow of one particular highway (a group of stations) can be illustrated for a particular date. User can specify the direction of traffic flow, such as northbound or southbound. An analysis chart of traffic flow information will be displayed beside the highway map, so that users can effectively identify traffic peak periods and locations. In addition, spatial outliers can be detected and displayed in the analysis chart.
- **Traffic video comparison ($\mathbf{T_{TD}T_{DW}S}$)**: Figure 6 shows a traffic video comparison for a specific highway or the whole highway network between two particular dates. Users can select to display traffic volume or occupancy and observe the traffic flow variation between these two dates.
- **2-D data cube map**: As shown in figure 7, users can select a specific dimension pair to show their relationship for a certain range of dates, such as Time of Date – Day of Week ($\mathbf{T_{TD}T_{DW}}$), Day of Week – Station ($\mathbf{T_{DW}S}$), and Station– Time of Date ($\mathbf{ST_{TD}}$). Figure 5 displays a $\mathbf{T_{DW}S}$ example of the weekday northbound average traffic volume for highway I-35W between January 1, 1997 and February 1, 1997.

Mapcube is available at: http://europa.nvc.cs.vt.edu/~ctlu/Project/Mapcube/mapcube.htm

3. Mapview

Spatial outliers are spatially referenced objects whose non-spatial attribute values are significantly different from those of other spatially referenced objects in their spatial neighborhoods. A spatial outlier is a local instability, or an extreme observation with respect to its neighboring values. Mapview is a web-based spatial analytical software, which is designed to facilitate the observation and discovery of spatial outliers for the US census data. Mapview supports the visualization of 11 different census attributes and provides the functionality of detecting local abnormality using various spatial outlier detection algorithms. Developed in Java, Mapview allows users to conveniently access it through web browser and to interactively define execution parameters. To overcome the deficiencies of the existing spatial outlier detection algorithms, we also propose two new algorithms, Iterative Z-value and Iterative Ratio, which can detect true outliers ignored by the existing algorithms and remove falsely detected spatial outliers.

The census data is the most detailed tabulation of American demographic data compiled by U.S. Census Bureau. It contains detailed data on population, race and ethnicity, age and sex, education, employment, income, poverty, housing, and many other attributes for more than 3000 counties in the United States. Our current Mapview demo version supports 11 attributes.

Software Architecture: The system has a three-tier architecture, GUI, Outlier detection algorithms and database files. The GUI draws a US map on a county level using the geographical coordination information of each county. The outlier detection algorithms receive user query from the GUI, calculate the outliers from data source, and return the result back to the GUI for display. There are three database files, including polygon file, county attribute file, and neighborhood relationship file.

System Demontration: Mapview can efficiently and effectively discover spatial outlier counties and mark them with distinguishable color. In addition, users can click each county to see its corresponding attribute value and those of its neighboring counties. In Figure 8, Cononino county in Arizona is selected and its population density value is displayed. The population densities of its six neighbor counties are shown as well. Before running the outlier detection algorithm, the system will request user input for the number of outliers, as shown in Figure 9. In Figure 10, the outlier counties are identified and marked in blue. Their attribute values and the attribute values of their neighboring counties are displayed in another window.

Spatial Outlier Detection Algorithms Supported

- **Scatterplot** is a graph based outlier detection method. It shows attribute values on X-axis and the average of the attribute values in the neighborhood on the Y-axis. A least square regression line is used to identify outliers. Nodes far away from the regression line are flagged as spatial outliers.
- **Moran Scatterplot** is a plot of normalized attribute value against the neighborhood average of normalized attribute values. It contains four quadrants. The upper left and the lower right quadrants indicate a spatial association of dissimilar values: low values surrounded by high value neighbors and high values surrounded by low value neighbors. Spatial outliers can be identified from these two quadrants.
- **Z-value Approach** takes advantage of the standardized difference between the attribute value of a point and the average attribute value of its neighbors. These points with the standardized difference values greater than a pre-defined threshold will be flagged as spatial outliers.
- **Iterative Z-value Approach** is our proposed new algorithm. The key idea of iterative approach is to detect outliers one by one. After one outlier is detected, its attribute value will be substituted with the average attribute value of its neighbors before next iteration begins.

- **Iterative Ratio Approach** is similar to iterative Z-value approach, whereas it identifies outliers through the ratio between the attribute value of a point and the average attribute value of its neighbors. Those points with the ratios (or the inverse of ratios) greater than a pre-defined threshold will be selected as spatial outliers.

Mapview is available at: http://europa.nvc.cs.vt.edu/~ctlu/Project/Mapview/index.htm

Appendix I: Figures

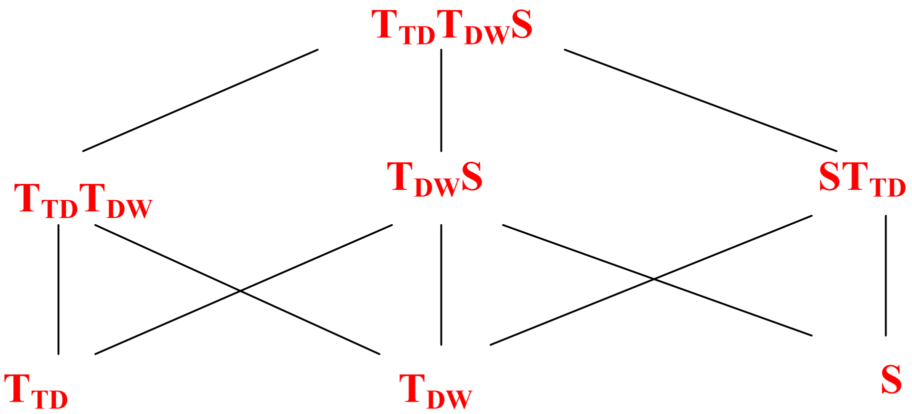

Figure 1: Dimension hierarchy of Mapcube.

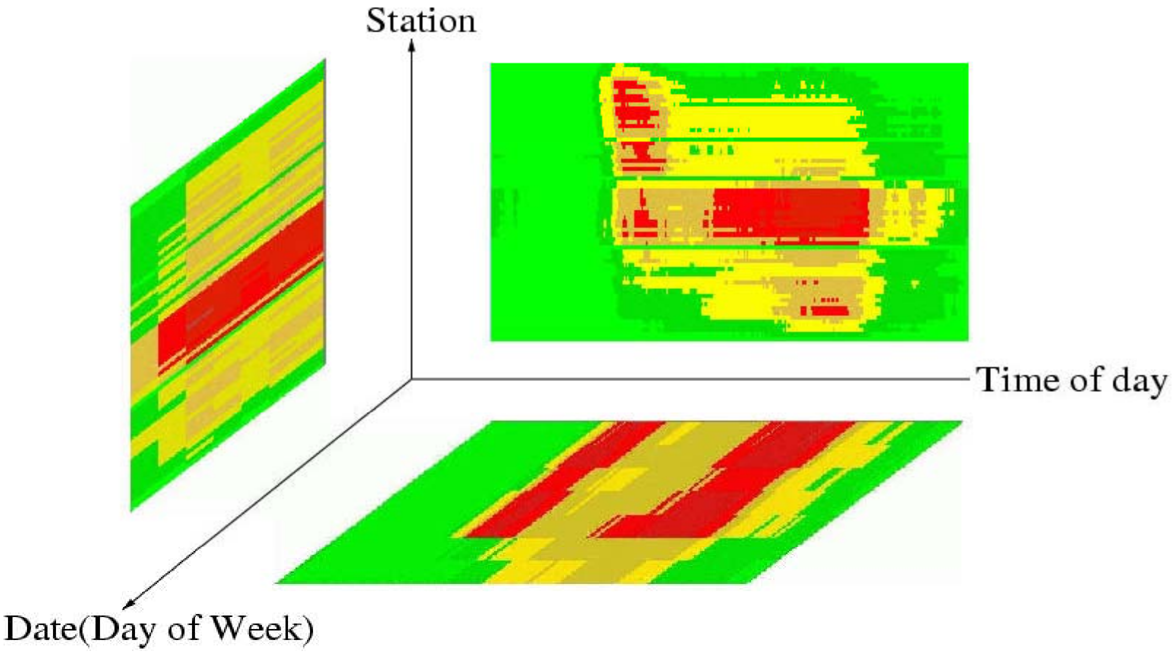

Figure 2: Traffic dimensions.

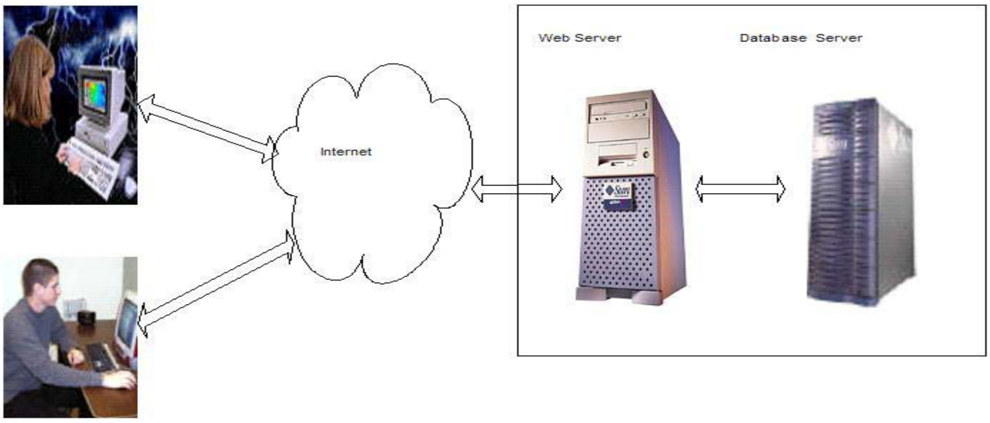

Figure 3: Three-tier architecture of Mapcube.

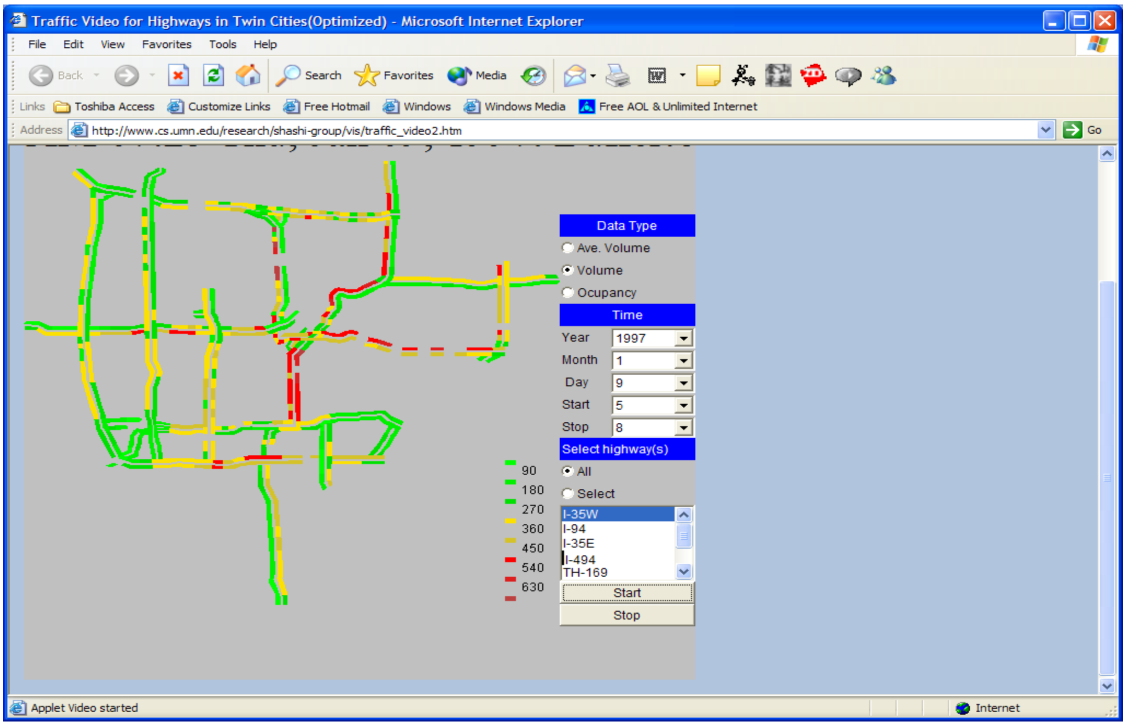

Figure 4: Traffic video.

Figure 5: Highway time map (ST_{TD}) with spatial outliers detected.

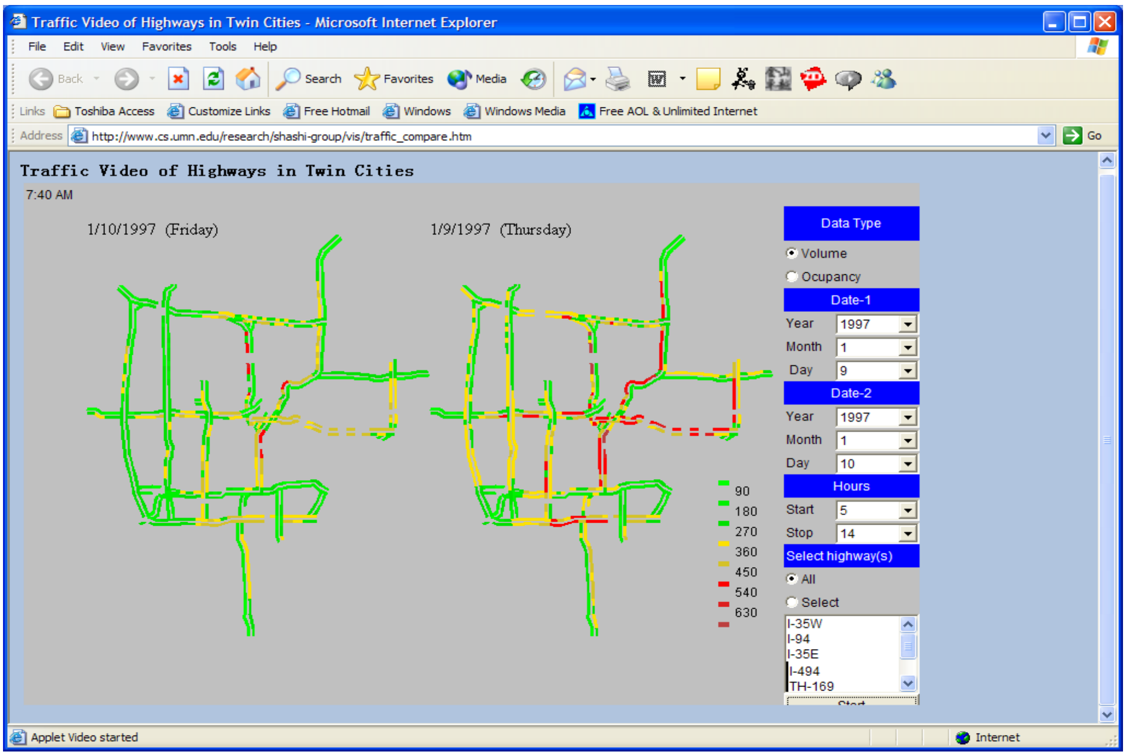

Figure 6: Traffic video comparison for two selected dates ($T_{TD}T_{DW}S$).

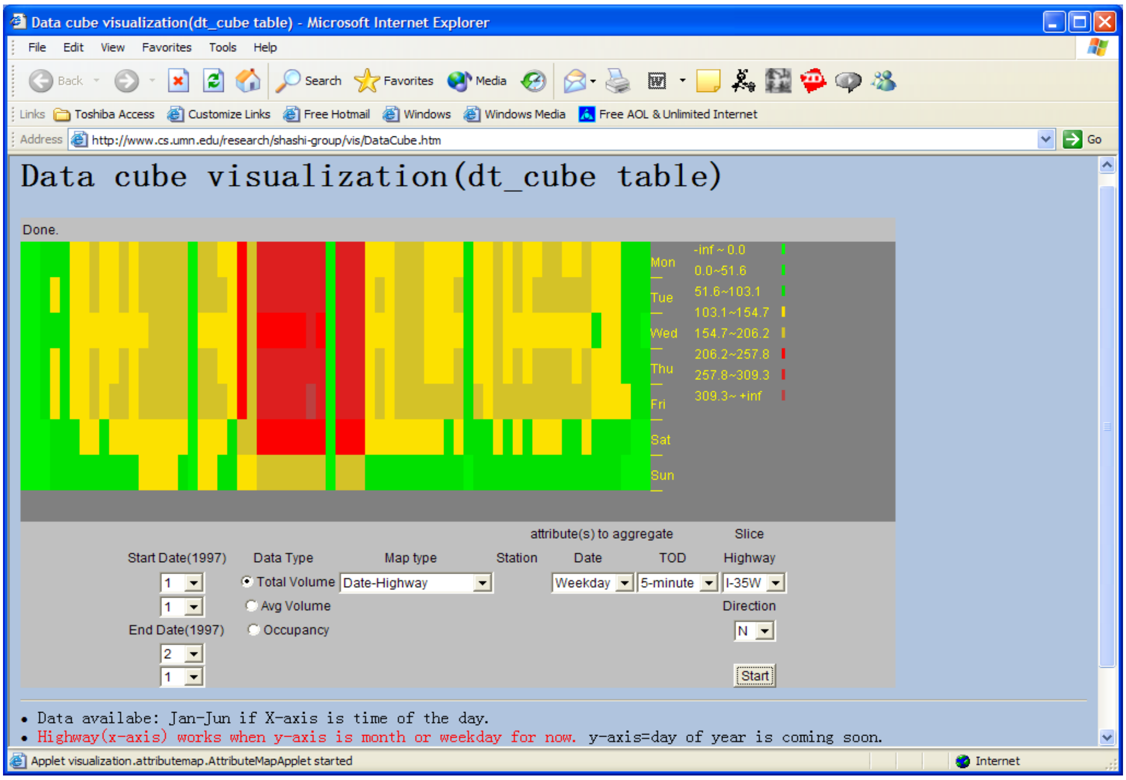

Figure 7: Example of 2-D data cube map ($T_{DW}S$).

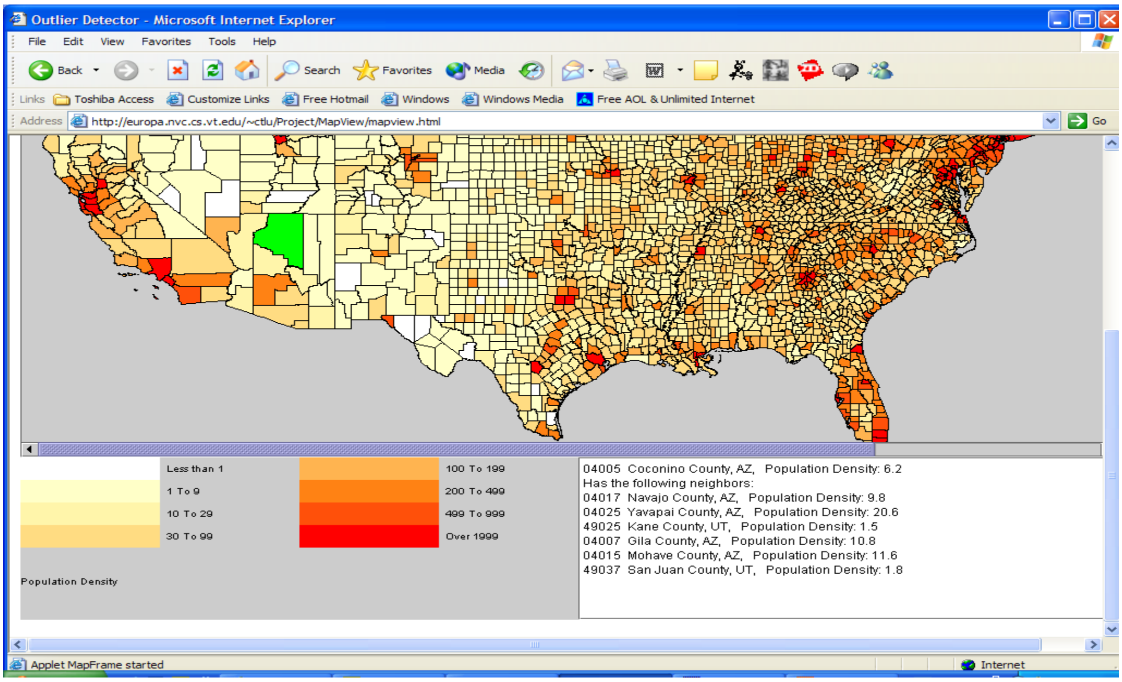

Figure 8: Population densities of Conconino County and its neighboring counties.

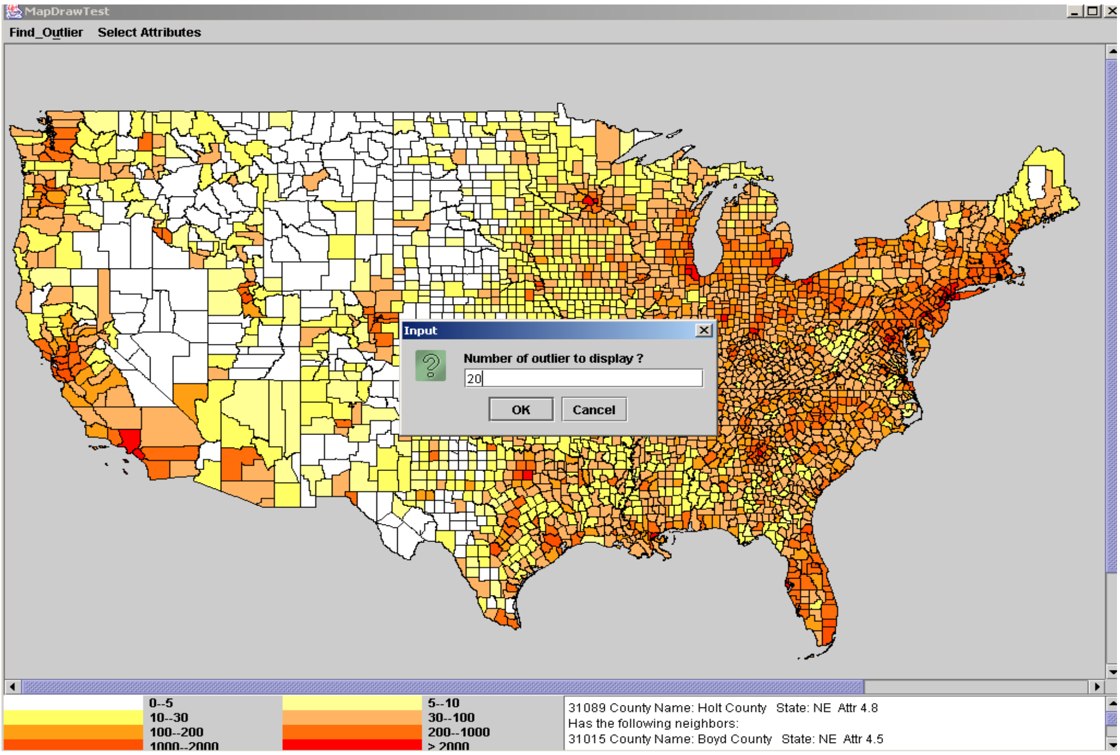

Figure 9: User input for the number of spatial outliers.

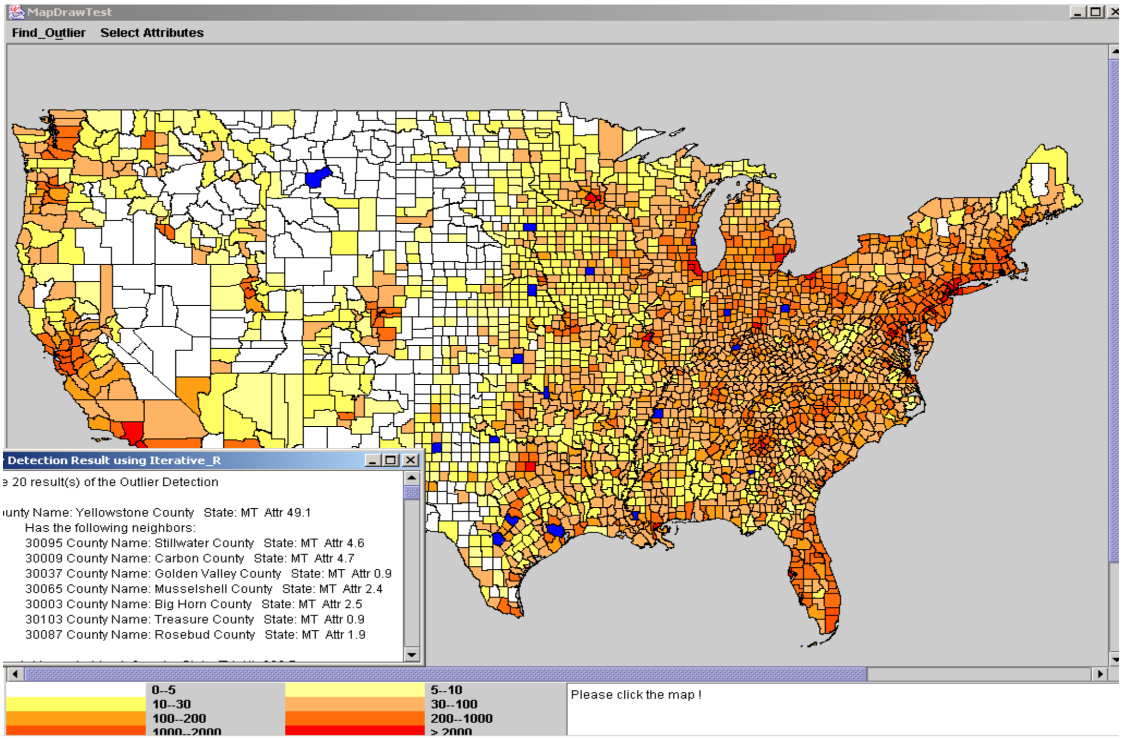

Figure 10: The detected 20 spatial outlier counties (marked in blue) and their corresponding attribute values.

Appendix II: Related Publications

1. Mapcube

[1] S. Shekhar, C.T. Lu, P. Zhang, Detecting Graph-based Spatial Outliers: Algorithms and Applications, *Proceedings of the Seventh ACM International Conference on Knowledge Discovery and Data Mining(ACM SIGKDD)*, 2001.

[2] S. Shekhar, C.T. Lu, P. Zhang, CubeView: A System for Traffic Data Visualization, *Proceedings of the Fifth IEEE International Conference on Intelligent Transportation Systems*, 2002.

[3] S. Shekhar, C.T. Lu, P. Zhang, Rulin Liu, Data Mining for Selective Visualization of Large Spatial Datasets, *Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence*, 2002.

2. Mapview

[1] C.T. Lu, Dechang Chen, Yufeng Kou, Algorithms for Spatial Outlier Detection, *IEEE International Conference on Data Mining*, 2003.

[2] C.T. Lu, Dechang Chen, Yufeng Kou, Detecting Spatial Outliers with Multiple Attribute, *IEEE International Conference on Tools with Artificial Intelligence*, 2003.

[3] S. Shekhar, C.T. Lu, and P. Zhang, Detecting Graph-based Spatial Outlier. *Intelligent Data Analysis: An International Journal*, 6(5):451-468, 2002.

Appendix III: Team Members

1. **Mapcube team:**
    - Shashi Shekhar (Professor, Department of Computer Science, University of Minnesota)
    - Chang-Tien Lu (Assistant Prof., Department of Computer Science, Virginia Tech)
    - Pusheng Zhang (Ph.D. Candidate, Department of Computer Science, University of Minnesota)
    - Rulin Liu (M.S. student, Department of Computer Science, University of Minnesota)

2. **Mapview team:**
    - Chang-Tien Lu (Assistant Professor, Department of Computer Science, Virginia Tech)
    - Yufeng Kou (Ph.D. Candidate, Department of Computer Science, Virginia Tech)
    - Hongjun Wang (M.S. student, Department of Computer Science, Virginia Tech)

Appendix IV: Presenter's Information

Chang-Tien(C.T.) Lu, Assistant Professor
Department of Computer Science
Virginia Tech
Northern Virginia Center
7054 Haycock Road, Room 308
Falls Church, VA 22043, USA

URL: http://people.cs.vt.edu/~ctlu
Email: ctlu@vt.edu
Phone: (703) 538-8373
Fax: (703) 538-8348

Short Bio:

Chang-Tien(C.T.) Lu received the BS degree in Computer Science and Engineering from the Tatung Institute of Technology, Taipei, Taiwan, in 1991, the MS degree in Computer Science from the Georgia Institute of Technology, Atlanta, GA, in 1996, and the Ph.D. degree in Computer Science from the University of Minnesota, Minneapolis, MN, in 2001. He is currently an assistant professor in the Department of Computer Science at Virginia Polytechnic Institute and State University. His research interests include spatial database, data mining, data warehousing, and geographic information systems. He served on the program committee and was publications chair for the 15th IEEE International Conference on Tools with Artificial Intelligence in 2003. He also served on the program committees for the 11th ACM International Symposium on Advances in Geographic Information Systems in 2003, the Third IEEE International Conference on Data Mining in 2003, the First ACM International Workshop on Multimedia Databases in 2003, and the 2003 International Conference on Parallel Processing.