Transportation data visualisation and mining for emergency management

Chang-Tien Lu* and Lakshmi N. Sripada

Department of Computer Science,
Virginia Polytechnic Institute and State University,
7054 Haycock Road, Falls Church, VA 22043, USA
E-mail: ctlu@vt.edu E-mail: lsripada@vt.edu
*Corresponding author

Shashi Shekhar and Rulin Liu

Department of Computer Science, University of Minnesota,
200 Union Street, SE, MN 55455, USA
E-mail: shekhar@cs.umn.edu E-mail: rliu@cs.umn.edu

Abstract: Emergency control systems influence the design, development, and maintenance of public infrastructure in a significant manner. These systems help to restore the infrastructure and sometimes prevent damage in cases of crisis. The road transportation system is an important part of public infrastructure and is one of the critical assets of a nation. Recovering from emergency situations on the road has been one of the major concerns of the traffic management personnel. Many software systems have been developed using advanced technologies to help in the detection of, prevention of and recovery from road crisis. We have developed the CubeView visualisation system based on the concepts of data-visualisation and data mining. Visualisation of loop-detector traffic data can help in the identification of potentially important patterns embedded in the data. CubeView helps in summarising the major trends in traffic patterns for both historical and current data. This paper presents the idea of applying the concepts of data-visualisation and outlier detection to road traffic data using the CubeView system for emergency situation control and management planning. CubeView is a web-based software system and is available at: http://europa.nvc.cs.vt.edu/~ctlu/Project/Mapcube/mapcube.htm.

Keywords: intelligent transportation system; emergency traffic management; data visualisation; spatial data mining.

Reference to this paper should be made as follows: Lu, C-T., Sripada, L.N., Shekhar, S. and Liu, R. (2005) 'Transportation data visualisation and mining for emergency management', *Int. J. Critical Infrastructures*, Vol. 1, Nos. 2/3, pp.170–194.

Biographical notes: Chang-Tien 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 a PhD degree in Computer Science from the University of Minnesota, Minneapolis, 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 is a member of the IEEE and ACM.

Lakshmi N. Sripada has a Master's degree in Information Systems from Virginia Polytechnic and State University (2004). Her research interests include data mining, data visualisation, spatial databases, and geography markup language (GML).

Shashi Shekhar is a professor at the University of Minnesota. He received a PhD in Computer Science from UC Berkeley in 1989. He was elected an IEEE fellow for contributions to spatial database storage methods, data mining, and geographic information systems (GIS). He has co-authored a popular textbook on spatial databases, is serving as a co-editor-in-chief of the Geoinformatica journal, and is a member of the NAS/NRC Mapping Sc. committee, and ACMGIS Workshop steering committee. He served on the Board of Directors of University Consortium on GIS and the editorial board of IEEE Transactions on Knowledge and Data Engineering.

Rulin Liu received an MS degree in Computer Engineering from the University of Minnesota, Twin Cities, in 2002. He is currently a software engineer at GE Healthcare. His area of work is integrated radiology information system (RIS) and picture archiving and communication system (PACS). His research interests include medical data visualisation and medical data mining.

1 Introduction

The transportation system plays a major role in the mobility and economic development of the country. According to the US Department of Transportation, an average person in the US travels approximately 17,000 miles annually and 12,600 ton-miles of freight are moved each year per person (US Department of Transportation Research and Special Programs Administration and VOLPE National Transportation Systems Center (1997). The statistical figures only further indicate the vastness of the infrastructure as it spreads across the entire sub-continent. Operations at this large a scale also carry with significant risk in case of damage to the system and the need to continuously monitor the system, so that emergency situations can be effectively controlled. The Departments of Transportation usually collect data related to the various traffic patterns on highways and other local roads using the sensor network embedded in the roadways. The data collected from these sensors is of great value for decision-making. However, it is usually stored in raw binary data format, and is not properly organised for analysis. These systems that aid in the decision-making for traffic control are called intelligent transportation systems (ITS), which use sensing, locating, computer, and communications technologies to model the operations on highways and transit (US Department of Transportation Research and Special Programs Administration and VOLPE National Transportation Systems Center, 1999). Huge amounts of data have been collected by various public infrastructure management agencies in the form of video and sensor measurements. This data usually is of little use without proper formatting and presentation. With the improvement in database technologies and various data processing techniques, many systems have been proposed and developed for presentation and analysis of such data. Concepts of database management, data mining, and data visualisation can be used for processing of information contained in these databases to discover trends that would otherwise not be evident.

Data mining has been emerging as a powerful tool for the processing and analysis of large databases. Concepts of data mining can be used for the analysis of the huge amount of traffic data collected by road sensors. Data mining is a process to extract implicit, nontrivial, previously unknown and potentially useful information (such as knowledge rules, constraints, regularities) from data in databases (Fayyad et al., 1996). One of the important tools of data mining is data visualisation. Visualisation helps in drawing meaningful conclusions from the data, which would otherwise be a cumbersome process. Data visualisation is the pictorial presentation of data, so that it can be easily interpreted. Effective data mining depends on having a human involved in the data exploration process. Visual data exploration integrates the human creativity and general knowledge with the computational power of computers which makes the process of knowledge discovery an easier process (Keim, 2001). In this paper we propose using the concepts of data mining and data visualisation for emergency control. For this purpose, a software tool for traffic data analysis has been developed – CubeView. CubeView was designed to help in the interpretation of traffic data and decision-making. It is a visualisation software package for observing the summarisation of spatial patterns and temporal trends embedded in traffic data. An important problem that arises during emergencies is the congestion due to evacuation from the crisis-hit spots. In addition, congestion can be a problem in case of emergencies that occur on the roads during day-to-day traffic flows, like accidents, repairs, and maintenance. In these cases, it is useful to continuously monitor the system and immediately identify the incident hot spots so that the necessary actions can be taken without any delay. The analysis of traffic flow can help in the prediction and control of problems occurring due to congestion on highways, thus helping in minimising the negative effects of such events.

The sections that follow will provide detailed description of how the various concepts of data mining and data visualisation can be used for crisis management. The rest of the paper is organised as follows. Section 2 gives an overview of the related work in data mining, data visualisation, and intelligent transportation systems. Section 3 provides an overall view of the CubeView system and its implementation. We will discuss the issues that arise during emergency events and, and describe our system architecture. Section 4 provides a description of the various views of CubeView system and discuss how CubeView can be used for emergency control and early warning. Section 5 addresses the limitations of the system and potential future directions. Finally, we summarise our work in Section 6.

2 Related work

Traffic management software tools have been developed for a number of state transportation agencies. These systems incorporate online analysis tools that use the data collected from sensor network to provide information to commuters and researchers. The Arizona Department of Transportation uses the freeway management system (FMS) for the analysis of data collected through the sensors embedded in the freeways of the state (Arizona Department of Transportation). FMS has been mainly designed to manage the congestion problems on the urban highway networks in the metropolitan Phoenix and Tucson areas. The Georgia Department of Transportation's Intelligent Transportation

System is called the Navigator (Georgia Department of Transportation). The transportation system developed by San Antonio District of the Texas Department of Transportation is called the TransGuide (San Antonio District of the Texas Department of Transportation). The state of California also has its own online information systems on transportation (San Diego District 11 Traffic Management Centre; Los Angeles Department of Transportation). While some of these systems provide a lot of features showing details about the highways on the maps, some systems provide only basic utilities like the current congestion and road condition without providing any analysis tools for processing historical data.

The system we have developed, CubeView, is based on the concepts of data mining and data visualisation. We introduce the basic concepts that have been used in the designing of the system and how these concepts aid in decision-making. Data mining has been defined as the process of extracting hidden information from large databases that can used for predicting future events (Thearling). The explosive growth in data and databases used in business management, government administration, and scientific data analysis has created a need for tools that can automatically transform the processed data into useful information and knowledge. Data mining allows organisations and companies to extract useful information from the vast amount of data they have gathered, thus helping them make more effective decisions. Two important features of data mining techniques are evident – automated prediction of trends and behaviours and automated discovery of previously unknown patterns. Both of these functionalities play an important role in the prediction and recovery from crisis situations on roadways. Prediction of trends based on analysis of past data can help traffic personnel predict the traffic patterns in case of emergencies. This is done in data mining by building models for historical data and using those models to predict future events. CubeView is such a model that has been developed based on the traffic data collected in the past and that can used to study the traffic flows in future.

Although data mining is concerned with database knowledge discovery in general, there is a specialised branch of database analysis that can be particularly used for traffic data – spatial data mining. Spatial data sets have an additional characteristic in addition to the normal data sets – spatial attribute. This attribute makes the data elements relate to each other based on the spatial characteristics. Traffic databases fall under the category of spatial databases, since the data collected is related to each other on the basis of location. Spatial data mining is concerned with the discovery of interesting and useful but implicit knowledge in spatial databases (Koperski et al., 1996; Koperski and Han, 1995; Chawla et al., 2001; Shekhar and Huang, 2001; Shekhar et al., 2002). With the huge amount of spatial data obtained from satellite images, medical images, and geographical information systems (GIS), it is a non-trivial task for humans to explore spatial data in detail. Spatial datasets and patterns are abundant in many application domains related to NASA, the Environmental Protection Agency, and the Department of Transportation. A key goal of spatial data mining is to partially automate knowledge discovery, i.e., search for pieces of information embedded in very large quantities of spatial data. The concepts of spatial data mining can be used for the information retrieval process from the traffic databases. However, it is not always an easy task to process spatial databases. Challenges in spatial data mining arise from the following issues. First, classical data mining is designed to process numbers and categories. In contrast, spatial data is more complex and includes extended objects such as points, lines, and polygons. Second, classical data mining works with explicit inputs, whereas spatial predicates and attributes are often implicit. Third, classical data mining treats each input independently of other inputs, while spatial patterns often exhibit continuity and high autocorrelation among nearby features. Such inter-relationships among data elements, although difficult to present and analyse, are very helpful in detecting abnormal situations. Data mining and knowledge discovery tasks can be classified into four general categories:

- dependency detection (e.g. association rules)
- class identification (e.g. classification, clustering)
- class description (e.g. concept generalisation)
- exception/outlier detection (Knorr and Ng, 1998).

Most research has concentrated on the first three categories, which correspond to patterns that apply to a large percentage of objects in a dataset. In contrast, outlier detection focuses on a very small percentage of data objects, which are often ignored as noise. An outlier in a set of data is an observation or a point that is considerably dissimilar to or inconsistent with the remainder of the data (Ramaswamy et al., 2000). This property of spatial databases is very important in analysing traffic databases and detecting abnormal and congestion points in traffic flow. Spatial data mining and thus outlier detection has been one of the significant data analysis techniques. It has been used for the study of data to find hidden trends and patterns. These techniques can prove to be useful in the analysis of traffic data, which can be helpful in controlling and preventing abnormal events in future.

Another important concept that has been used in the development of CubeView system is data visualisation. An important function of any emergency control and response system is to help in the detection of incident areas and give a detailed description of the situation to support the management personnel in decision-making. One of the techniques to help in the understanding and analysis of complex situations is data visualisation. Data visualisation is an important area of data analysis, where the data collected is summarised and presented in a visual form to aid in decision-making and to better grasp the critical issues of the situation. Visualisation techniques have been used in various fields of technology as an aid to decision-making and prove to be one of the important tools for clear communication and understanding of the situation. Visualisation for spatial databases is gaining importance due to its ability to detect abnormalities in data sets. These datasets are related to each other based on location and a deviation in behaviour from the neighbours helps in detecting the abnormal conditions.

CubeView system was designed based on such concepts and proves to be a useful tool for decision-support in cases of crisis management. It provides a pictorial representation of the results of analysis and trends and patterns that can be discovered. Data visualisation techniques effectively facilitate the task of spatial data-mining. The primary importance of user interaction and graphical representation of data, as part of the whole knowledge discovery process to make patterns more understandable by users has been emphasised (Fayyad et al., 1996, 1996). Many powerful visualisation tools have been developed in databases and data mining. Visualisation tools, such as VisDB (Keim and Kriegel, 1995), XGobi (Swayne et al., 1998) and XmdvTool (Rundensteiner et al., 2002), provide a set of predefined visualisations, e.g., histogram, scatter plots, and parallel coordinates, to explore multi-dimensional datasets. These views are augmented with brushing and zooming techniques. However, users cannot interactively construct and refine a wide range of displays to suit an analysis task for large spatial datasets. There have been many studies which focus on the visualisation of spatial data, though none of them address the constraints of large data sets on visualisation (Fayyad et al., 2001). Grinstein et al. (2001) provide a brief summary of the various methods of representing data in visual form and how it can be used in data mining. Effective visualisation is needed to explore large spatial datasets. This paper combines the techniques of data mining and spatial visualisation techniques to discover implicit patterns and relationships embedded in the traffic data, which can then be used for detecting and controlling emergency and abnormal situations.

3 Traffic visualisation and data mining for emergency control

3.1 Issues related to emergency control

Emergency events mostly occur without any warning signs and require immediate response in order to minimise the negative impact. Such events usually involve huge amounts of resources. Efficient use of limited resources is one of the important tasks in such situations. A good understanding of the situation and the events following is therefore a very important pre-requisite. However, proper management of such situations on the highways requires a few basic processes and systems to be set up to help the traffic personnel control the situation. The first important step towards the use of traffic data for emergency control is the formatting and cleaning up of data from its binary form. The data once formatted is more useful to transportation engineers and managers involved in decision-making. This data has therefore to be processed so that important conclusions can be drawn. The study of such data can help in predicting the traffic flow patterns in future. Although this process yields better results than skimming the raw data, it is not much effective in cases where the time available to study and analyse the data is short. In other words, during emergencies, the staff usually is over-worked and the people involved have very less time to spend on the analysis of data. In such situations it would be useful if the data can be readily presented to the users in a pictorial form and conclusions drawn from it. Visual presentation of data helps the decision-makers much more than mere numbers. Visual representation helps present a uniform picture to all the members involved. There is single access point to data and a single representation, which ensures that the data will not be misinterpreted.

There is another practical aspect of emergency management. Congestion is one of the major problems that arise due to emergency events. Evacuation is one of the major protective action option used in case of emergency situations (Rathi and Solanki, 1993). This creates a need for efficient traffic management and control in case of emergencies and the resulting congestions. Many studies have been conducted to estimate and simulate the evacuation time required in situations of emergency. The study and analysis of congestion patterns for the highways can help in developing an evacuation plan in cases of emergency when congestion is the main problem. According to the Arizona Department of Transportation, Transportation Technology Group, there are two types of congestion: recurring and non-recurring congestion (Arizona Department of Transportation and Transportation Technology group, 2002). Congestion occurs when the demand for highways or roads exceeds the capacity of the roads. Recurring congestion usually occurs during the morning and the evening commute when the number of highway users is large. Non-recurring congestion occurs due to unpredicted conditions such as accidents, disabled vehicles, adverse weather conditions, etc. Emergency traffic control and congestion belongs to the second category and the data collected through the sensors can be useful in managing such traffic.

CubeView visualisation can be used as a tool for the analysis and study of such congestions. It can help the traffic control department to estimate the time required to clear off the congestion by simulating various emergency situations based on the historical data collected. Also the analysts can determine alternative routes that can be used in case of emergencies. Such alternatives can be used by the emergency response teams to reach the hot spots without significant delays. They can also be helpful in redirecting the traffic from the incident region, thus avoiding over-congestion.

According to Pearce, the events of 9/11 turned the attention of USDOT to the Transportation Information Operations Center in Washington DC, which operated round the clock to keep the administration informed of the state of transportation in the two metro areas (Pearce, 2002). Such situations emphasise the importance of developing information processing systems that can help the administration in making informative decisions in case of emergency situations. These events have turned everyone's attention to the need for ensuring that the means of transportation and the agencies operating them are fully prepared to tackle threatening situations. This means that the people responsible for the smooth operation of this infrastructure be provided with the necessary tools to prevent, prepare for, respond to and recover from natural and man-made disasters. The related activities are therefore divided into six stages by the Office of Homeland Security (OHS): Detection, Preparedness, Prevention, Protection, Response and Recovery. Although, detection has been mentioned here in the context of detecting any suspicious things in case of emergency, detection of congestion points can be one of the functions of Traffic Control systems. Current data can be used to detect the points where congestion might occur.

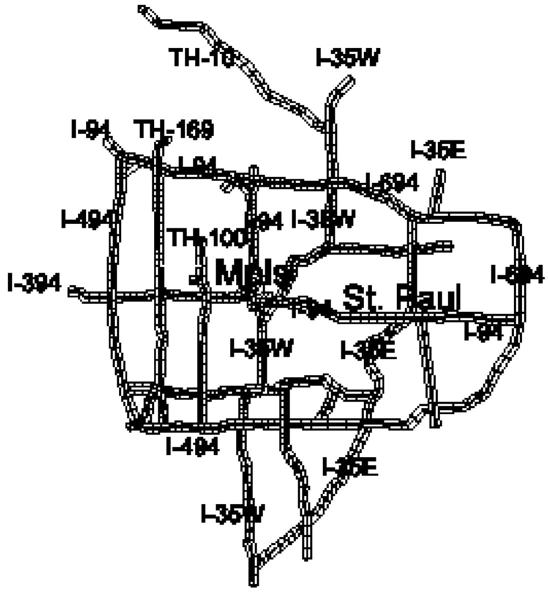
Effective and thorough planning is the first steps towards preparing to deal with emergency situations. Once the plans are prepared, they have to be understood by the personnel and practised regularly. CubeView can be an effective tool in preparing for the traffic congestion due to emergency situations. Based on the data collected on a day-to-day basis, an estimate can be made of the congestion conditions in case of emergencies. CubeView can be used to simulate situations of emergencies and used as a training tool for traffic personnel to find out alternative routes in case of emergencies. The whole process is depicted visually, which makes it easier for the users to understand the different scenarios and the effect of different alternatives. CubeView is also an effective tool for responding to the emergency situations, since it can give an idea of the distribution of traffic on various roads if real time data is fed into the system. The sections following will give a more detailed description of CubeView Visualisation system and its applications in traffic emergency control.

3.2 MNDOT system

A brief description of CubeView system and its origin would be helpful in understanding the working of the system and its relevance for crisis management. CubeView has been developed for the analysis of traffic data for Minnesota Department of Transportation. In 1995, the University of Minnesota, the Traffic Management Center (TMC) Freeway

Operations group started the development of a database to archive sensor network measurements from the freeway system in the Twin Cities (The Minnesota Department of Transportation's Traffic Management Center). The sensor network includes about nine hundred stations, each of which contains one to four loop detectors, depending on the number of lanes, as shown in Figure 1. Sensors embedded in the freeways and interstate monitor the occupancy and volume of traffic on the road. Figure 1 shows a map of the stations on highways within the Twin-Cities metropolitan area, where each polygon represents one station. The interstate freeways include I-35W, I35E, I-94, I-394, I-494, and I-694. The state trunk highways include TH-100, TH-169, TH-212, TH-252, TH-5, TH-55, TH-62, TH-65, and TH-77. I-494 and I-694 together to form a ring around the Twin-Cities. I-94 passes from East to Northwest, I-35W and I-35E are in South-North direction. Minneapolis downtown is located on the intersection of I-94, I-394, and I-35W, and Saint Paul downtown is located on the intersection of I-35E and I-94.

Figure 1 Detector map at the station level



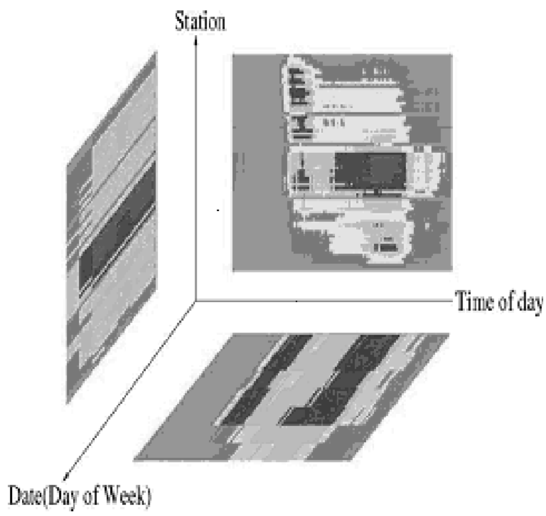
3.3 CubeView: an overview

High-performance visualisation techniques are becoming crucial as the amounts of traffic data collected by an ever-increasing network of sensors is becoming too large to be analysed manually. CubeView is a high performance spatial visualisation system for traffic data analysis. It is a multiple view system that has been defined as a system, which uses two or more distinct views to support the investigation of a single conceptual entity. Multiple views can provide utility in terms of minimising some of the cognitive overhead engendered by a single, complex view of data (Baldonado et al., 2000). The system presents information in various formats to assist transportation managers, traffic engineers, travellers and commuters and researchers and planners to observe and analyse traffic trends.

CubeView is a web-based application implemented in Java. It can be visited through web browser to display highway traffic information. The system provides several essential data cube visualisation utilities for traffic pattern analyses, including traffic video, highway time map, traffic video comparison, and 2D data cube map. It also supports spatial outlier detection, identifying abnormal traffic patterns in traffic data warehouses using visualisation techniques. These traffic patterns and rules can assist in decision-making for transportation managers, commuting routes selection for commuters, and traffic model establishment for researchers by providing a model to analyse the different alternatives available in the event of crisis. In the underlying data structure, the spatial data has been modelled as a spatial data warehouse to facilitate the query engine for the online analytical processing used in the visualisation software. The algorithms for filtering out the datasets to detect outliers have been tested on real-world traffic data sets.

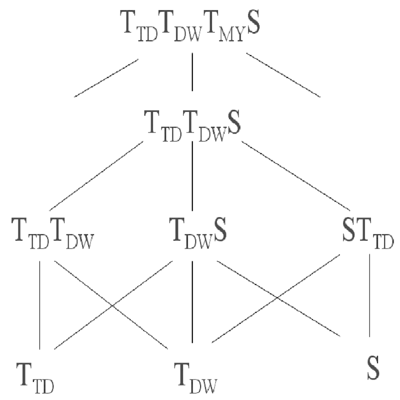
In general, the subjects of analysis in a multi-dimensional data model are a set of numeric measures. Each of the numeric measures is determined by a set of dimensions. In a traffic data warehouse, for example, the measures are volume and occupancy, and the dimensions are time and space. Dimensions are hierarchical by nature. These dimensions are shown in Figure 2. For example, the time dimensions can be grouped into ‘week’, ‘month’, and ‘season’, or ‘year’, which form a lattice structure indicating a partial order for the dimension. Similarly, the Space dimension can be grouped into ‘station’, ‘county’, ‘freeway’ or ‘region’. Given the dimensions and hierarchy, the measures can be aggregated in different ways. For example, for a particular highway and a chosen month, the weekly traffic volumes can be analysed.

Figure 2 Traffic dimensions



For the Twin Cities traffic data, we have the CubeView as a specialisation of general data cube operator as in Figure 3. In this figure, T_{TD} represents the time of a day, T_{DW} represents the day of a week, T_{MY} represents the month of a year and S represents the station. Each node is a data cube operation, and a view of the data. For example, S represents the traffic volume of each station of all the time. ST_{TD} represents daily traffic volume of each station. T_{TD} T_{DW} S represents traffic volume on each station of different time of a day, which is generated as a video in CubeView System. CubeView displays traffic with different colour schemes. Heavier traffic flows on a highway are indicated by darker colours. The basic concept behind CubeView is data cube. Because of the nature of the data cube, CubeView can analyse any traffic data. The software only requires space and time for each measure, like volume, occupancy and speed. Speed is derived from volume and occupancy. These requirements are very simple and most highway monitoring systems should be able to satisfy them.

Figure 3 Dimension hierarchy of CubeView



3.4 CubeView: system architecture

CubeView is a web-based java application with a three-tier structure. The three components of CubeView are: Graphic User Interface (GUI), web server, and database server. As discussed in Section 2.1, a data management tool for emergency control should provide for: collection of data at a singe source, provide an interface for extracting and processing the data, and provide interactive analysis tools. The three components in cube-view provide these three functionalities. The three components together provide a web-based visualisation system. GUI is implemented through JAVA applets, which accept input from users and sends requests to web server. GUI requests are sent to a web server that uses Perl and generates SQL queries for the database server. MySQL is the database server that manages the traffic data and processes the SQL queries and delivers query results.

The three layers of CubeView consist of:

- *Graphic user interface (GUI)*. This module provides the interface with which the user interacts. It accepts inputs from the users and displays the results. The GUI renders the highway map using geographic coordinate information of each station. It accepts queries from users and sends the queries to the middle level tier, which further requests the traffic data from the database. It colour codes the data according to the colour scheme and displays traffic video, outlier stations, and shows highway traffic volume maps for user-specified time, date and highway stations.

- *Data access middle tier*. The interface for the user and the Database containing the traffic data are connected through the web server. The web server has been implemented in Perl (Practical Extraction and Reporting Language) and CGI (Common Gateway Interface). The middle tier accepts requests from GUI and retrieves data from the database. It accepts requests from GUI and sends appropriate SQL (Structured Query Language) queries to the database to request the needed data. The results are sent back to GUI as strings.
- *Database tier*. All the data related to the traffic is contained in a database. My-SQL has been used as the database server. It stores the 5-minute interval road detection data and descriptions of each of the stations.

The raw data collected by the loop detectors is in binary format. This data is converted into text data using transformation programmes and later inserted into database servers. Three-tier architecture helps in the processing of both the current and the historic data. Historic data is stored in the databases and can be accessed by the users at any time. The simple web-based interface provides users with an easy access to the system. Easy access to the tool is one of the important requirements to be met by emergency management tools. Since crisis situations involve a large amount of mental and physical activities, it is important that the users of data have easy access to the data.

Figure 4 Overall view of the system architecture

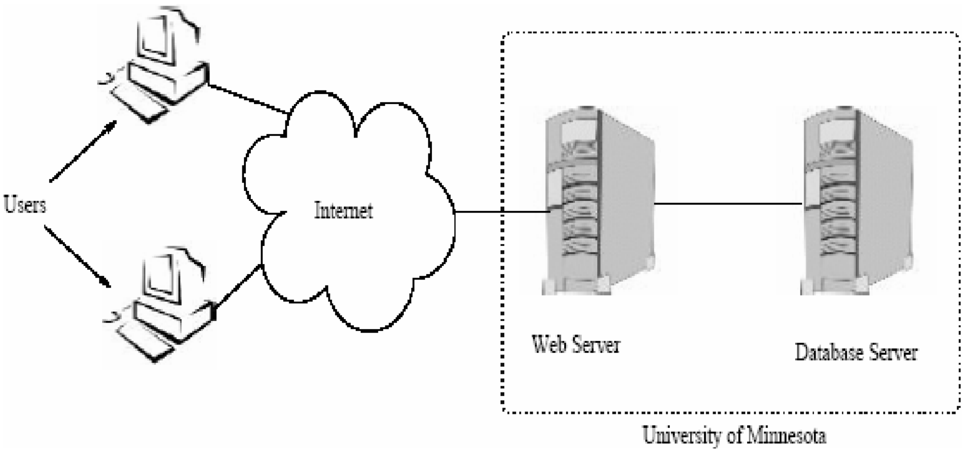
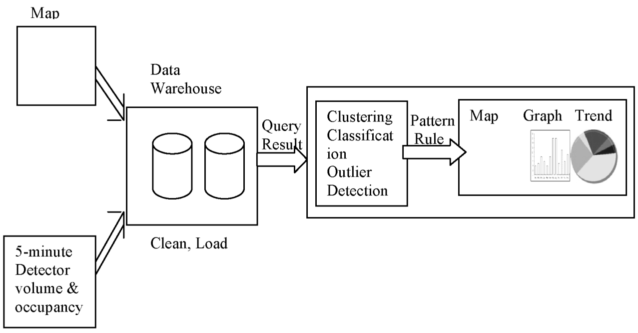


Figure 5 illustrates the data flows and major modules of CubeView system. The basic map and raw data are cleaned, transformed, and loaded into the data warehouse, which provides the multi-dimensional views and the OLAP operations for a variety of front end data mining analysis tools, e.g., classification, clustering, association rules. The discovered patterns or rules are then visually displayed as maps, tables, or charts for further interpretation. CubeView displays different levels of traffic volumes using different colours, which helps in easy identification of the traffic pattern. Use of archival traffic data for future emergency control purposes requires that the data should be analysed for relevant information and the information be presented in such a way that the users can easily draw conclusions from the patterns. The four kinds of traffic visualisation utilities provided by CubeView are:

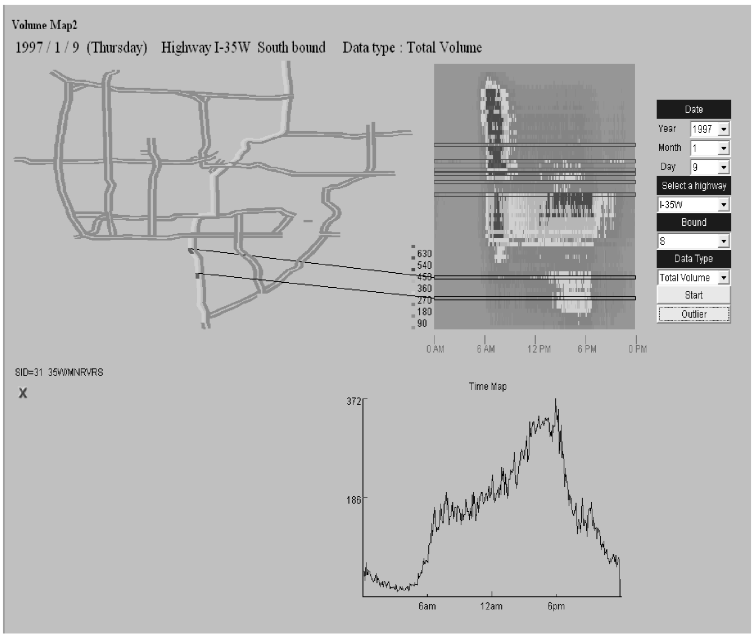
Figure 5 Data-flow and main modules in our system



Traffic video. This visualisation shows the changes in specific traffic patterns over a period of time in a video-like display. The changes are updates continuously, so that the users feel that there is flow in the traffic display. 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 (ST_{TD}). This view provides a detailed analysis of the traffic flow pattern on a particular highway for a particular day. As shown in Figure 6, static traffic flow for a selected highway (a group of stations) can be rendered for a chosen 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 quickly identify traffic peak periods and locations. In addition, spatial outliers (abnormal stations) can be detected and displayed in the analysis chart.

Figure 6 View showing the outliers and traffic patterns for each station on a highway. (Dimension ST_{TD})



Traffic video comparison (T_{TD} T_{DW} S). The traffic video can also be used to compare two different highways for their traffic patterns. CubeView supports 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 variations between these two dates.

2D data cube map. Users can select a specific dimension pair to show their relationship for a certain range of dates, such as Time of Day–Day of Week (T_{TD} T_{DW}), Day of week–station (T_{DW} S) and Station–Time of Day (S T_{TD}). These utilities provide the users with the independence to choose the dimensions on their own. Such type of functionalities is important so that users are provided with all the necessary tools for data analysis during emergencies.

4 Different views in CubeView system and emergency control

Different kinds of users can benefit from the analytical tools provided by the Cube View system. The traffic engineers through the outlier detection and analysis can analyse abnormal traffic flows. Travellers can observe the historical traffic flow for a specific event and avoid congested road segments.

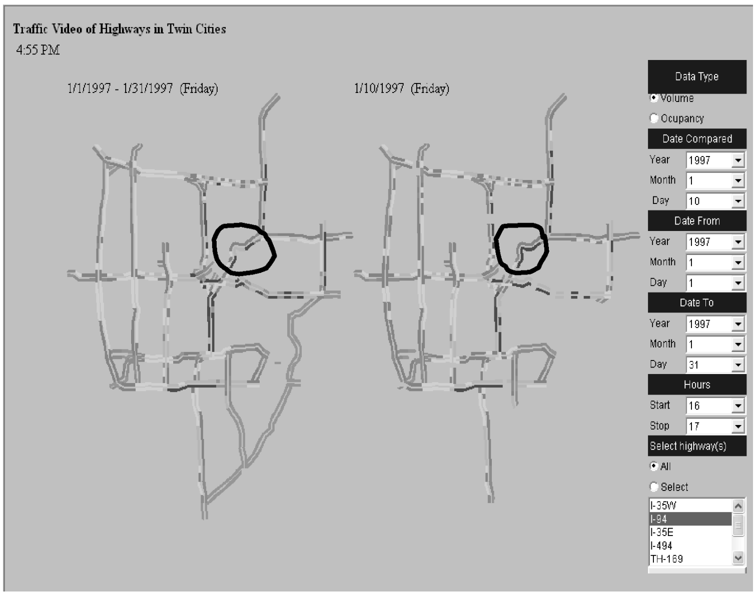
Each of the views that CubeView system presents can be helpful in a particular aspect of decision-making. Presently the CubeView system uses only historic data stored in databases, but it can also be used for analysing real-time data. CubeView thus proves to be an excellent tool for the analysis of current and historic traffic data. Historic data can be analysed to find out any exceptional cases of traffic flow and this experience can be used in dealing with the emergencies in future. The views provide summary reports for the particular days, months or weekdays and the highway or time. The different views provided are: date–time of the day, date–highway, highway–time of the day, date–map, highway–map, time of day–map, etc. Each view provides a summary of the traffic patterns, which can be helpful for the analysis of traffic patterns on particular highway, or on particular weekdays or particular time intervals. Some important terms that have been used in the system are: Total volume, average volume and occupancy. Total volume is measured as the number of vehicles passing through a station during a 5-minute time interval. Average volume is the total volume divided by the number of lanes at that station. Occupancy is the percentage of time during which a station is occupied by a vehicle. At regular intervals, this information is sent to the Traffic Management Center for operational purposes, e.g., ramp metre control, as well as research on traffic modelling and experiments. The different views in the system and their relation to emergency control measures can be summarised as follows:

4.1 Video map comparison

This view helps the users analyse and compare the traffic patterns by allowing them to simultaneously observe the traffic flow on two different dates.

Figure 7 shows the traffic flow on Friday, January 10, 1997 compared with the traffic pattern for the month of January 1997 during 4:55–5:00 PM. The view can be used to represent the volume or occupancy on the roads. Depending on the measure, different conclusions can be drawn. The areas in red colour show the presence of high level of traffic. This data can be useful for the decision makers in finding out the consequences of emergencies happening on the weekdays when the traffic is going to be high. The system provides a video like view of the traffic changes for the time interval. Users can choose a single date to be compared to the other date or an interval of days to be compared with a particular day. Such comparison tools are helpful in analysing how the patterns are typically different on particular weekdays compared to the other days of the week. This view on traffic pattern can help the users analyse the effect of emergencies that occur, on different routes and the roads that would get congested in such cases. This view can thus be used for the study of traffic both in the past and current situations. Analysts can view the flow of traffic in cases of past emergencies and can see the effect of congestions and the availability and usage of various roadways. This study can reveal any typical behaviours of motorists in cases of emergency in taking particular routes compared to others. In cases of current data analysis, researchers can for instance compare the present day patterns with the same weekdays in the past. This would help them to continuously monitor and study changing patterns. For instance in the comparison in Figure 7, the circled area on the right side as compared with the circled area on the left side shows that there is a part of I-94, that is congested on Fridays but less congested on other days. Thus traffic managers can plan accordingly in case anything happens on Fridays in that area. Such traffic patterns are not readily apparent when raw data is studied.

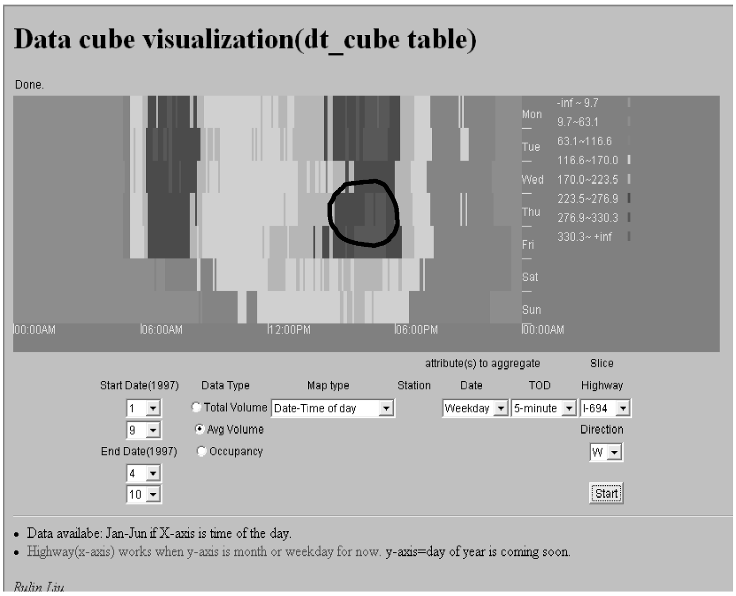
Figure 7 Traffic video comparison. (Dimension T_{TD} T_{DW} S)



4.2 *Time of the day and day of the week comparison*

This dimension is to show comparisons based on the time of the day and the days in the week for a particular highway. For instance, Figure 8 shows the traffic patterns for weekdays, during the period January 9–April 10 1997 on I-694W for different times during the day. Each weekday is represented as a row on the Y-axis and the time intervals are represented on the X-axis. Red colour indicates higher traffic volume. The view can be used for representing different measures of traffic patterns such as average volume, total volume and occupancy. Figure 8 shows the comparison of average traffic volume for different weekdays. We can see from the figure that the traffic is particularly high during the time interval 4–5 pm on weekdays, while it is not that high during the weekends. Also the traffic is particularly high on all days except Wednesdays during this time interval. Rush hour starts early on Fridays as compared to other weekdays. This analysis can be helpful in assessing the highways for traffic patterns based on the weekdays. Presenting the data in visual form helps in discovering many new characteristics of traffic flow that would otherwise have been very difficult to analyse. Based on this type of information, the traffic personnel can decide the amount of resources; they would require if any mishap occurs on a particular weekday at a particular time. Based on the day and the time of the day, the resource requirements would vary and the information obtained from the study of the traffic patterns can help in planning for the events. When used for analysing the current data, such type of analysis can help in predicting congestions and emergency situations. The visualisation would show unusually high amounts of traffic on such times, which can be easily detected when a comparison with other weekdays is presented.

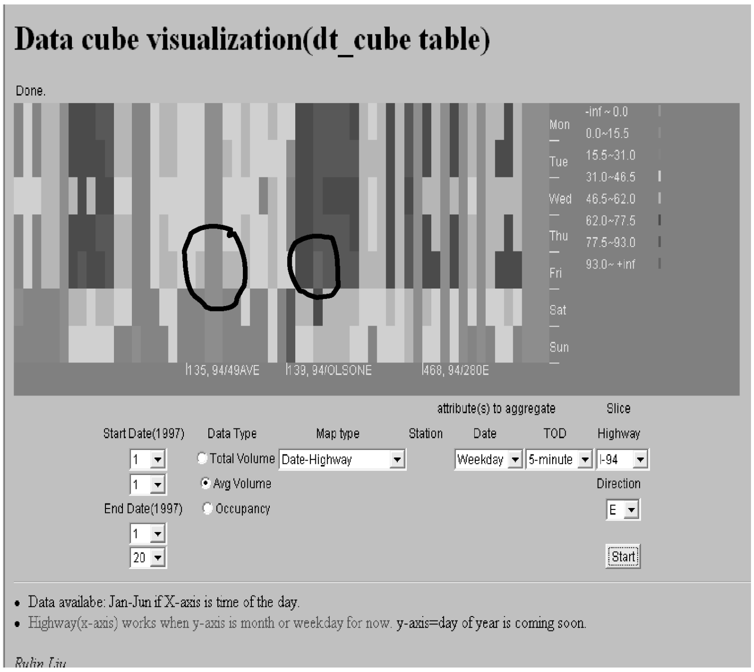
Figure 8 2D representation of time and traffic volume. (Dimension: T_{TD} T_{DW})



4.3 Day of the week and station comparison

This dimension provides a combination of the day of the week and the traffic for different stations on a particular highway. Figure 9 shows the comparison of traffic patterns on different days of the week for I-94 East during the time period January 1–20 1997. X-axis represents the different stations on the highway and Y-axis shows the weekdays. User can choose a particular highway and a particular measure for traffic pattern such as volume or occupancy. The figure shows the average volume for different stations and weekdays. The view can also be used to represent total volume or occupancy for different weekdays and stations. Thus different patterns can be studied as needed. This view provides a much closer look at the traffic flows on a particular highway. The points of congestion can be identified in this view. For instance, from Figure 9, it can be seen, that 94/OSLONE station on the I-94E is particularly congested on Fridays. The study of the patterns on a particular highway can be useful to manage the traffic on that highway based on the day on which any emergency situation occurs. The figure below for instance, shows that the point 94/OLSONE is particularly congested on Fridays. On the other hand, the point 94/49AVE is never congested. This information can help traffic personnel to estimate the resources they might require to deal with emergencies that might occur at various places on the highway. The pictorial representation of the traffic on the highway is very detailed and can show the points of congestion and potential points of congestion in case of emergency. An analysis of the past data thus helps in the identification of potential trouble spots. An analysis of current data would help in immediately identifying congestion points and outliers.

Figure 9 2D representation of traffic at different points on a particular highway. (Dimension T_{DW} S)

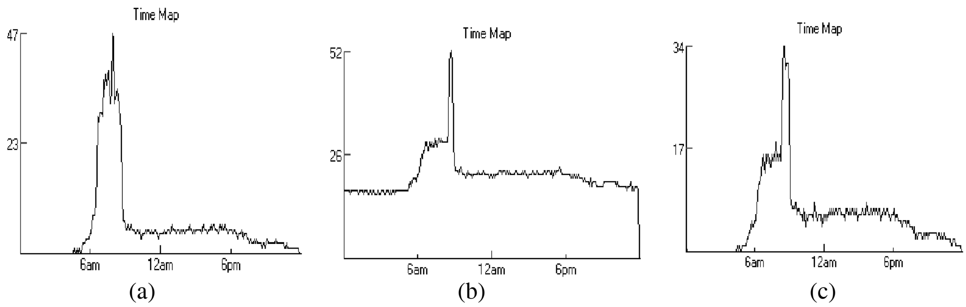


4.4 *Station–time of the day comparison*

Another very useful view to analyse the traffic patterns on a particular highway very closely is the station-time of the day comparison. This view shows the traffic patterns on a particular highway for each of the stations on the highway on a particular day. The view provides three different pictures. First it provides a map of all the highways highlighting the highway in question. Secondly, it provides a visual representation of the comparison of the traffic pattern for each station on the highway for different times during the day. Thirdly, on clicking on each row on this representation (each row represents a particular station), it shows the corresponding point on the highway map and at the same time shows the graphical representation of the traffic flow for that station for that day in a graph below. The X-axis on the graph represents different times during the day and the Y-axis shows the measure that the user chooses such as total volume, occupancy or average volume. This is an extremely useful analysis to study the traffic patterns on a specific highway and on particular days, which can be helpful in eliminating the congestion points during emergencies. Figure 6 shows the traffic pattern on I-35W for 01/09/97. The figure also shows the outliers (shown as red bars on volume map) on clicking the Outlier button. Outliers are an exception to the normal flow of traffic. An outlier in a set of data is an observation or a point that is considerably dissimilar to or inconsistent with the remainder of the data (Ramaswamy et al., 2000).

Detection of outliers can help in identification of abnormal traffic patterns. These abnormalities can be mainly seen during emergency situations when congestion is the main problem. Outlier detection techniques can immediately identify the event points in such cases. Spatial outliers detection is one of the important techniques of spatial data mining and very useful in spatial data analysis. A spatial outlier has been defined as a spatially referenced object whose non-spatial attribute values are significantly different from those of other spatially referenced objects in its spatial neighbourhood (Shekhar et al., 2002). Since traffic data is also spatial data, application of spatial outlier detection techniques can be very useful in case of contingencies. A closer look at the outliers can show how they can be helpful in identifying congestion spots during emergency. Figure 10 shows an example of traffic flow outliers. The Figures 10 (a)–(c) show traffic on three stations 581,582,583 on 17 January.1997 on I-35W. In the traffic data set, each station is a spatially referenced object with spatial attributes (e.g. location) and non-spatial attributes (e.g., measurements). The X-axis is the time in a day, and the Y-axis is the volume of a station. As shown in the figure, the abnormal station (Station 582) had the volume values significantly inconsistent with the volume values of its neighbouring stations 581 and 583.

Figure 10 (a) Station 581 (b) Station 582 (c) Station 583



(a) (b) (c)

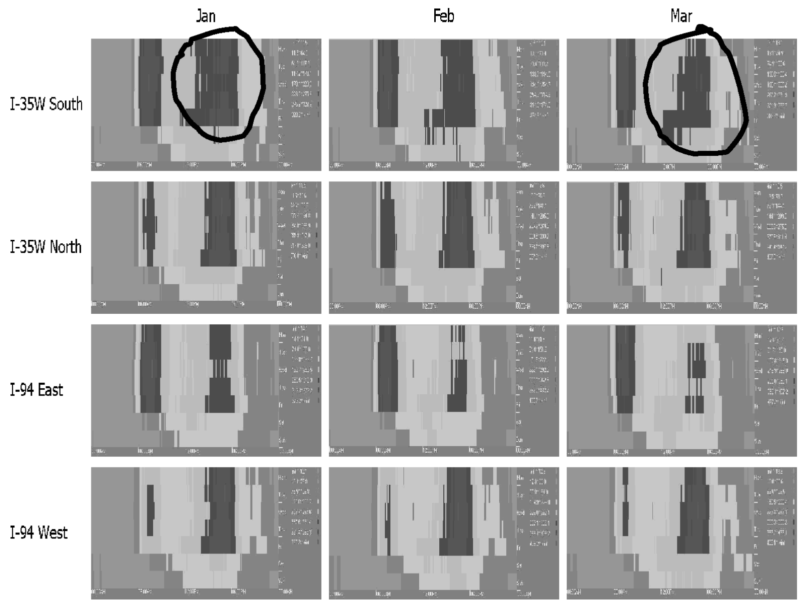
The identification of spatial outliers can help traffic managers quickly respond to faulty detectors or abnormal situations. This can be of particular help while processing current data. Users can immediately identify the trouble station and respond promptly thus reducing the response time in cases of emergency. With the help of visualisation, identification of outliers thus becomes easier.

4.5 Time of day, day of week comparison for different routes and months

This node on the dimension lattice shows the traffic on a particular day of week, at a particular time for various months and for the various routes. This view presents a summary of all the information at one glance. It provides a complete picture of the traffic situation in the past. Figure 11 shows a comparison of the traffic patterns for four different highways- I-35W South, I-35W North, I-94 East and I-94 West for the months of January, February and March. Each picture in the view shows the traffic pattern for a particular month and on various days of the month for various times during the day. A comparison of traffic patterns for different months and highways can be made. The outer X-axis represents the different months. The outer Y-axis represents different highways. The inner X-axis in each of the small figures represents the different times during the day and the inner Y-axis represents the days in the week.

Apart from the analysis of traffic patterns on each highway, it is also important and useful if the traffic patterns for various highways can be compared to each other. Also it can be helpful if the view can provide comparison for traffic patterns on the weekdays for various months. This view provides such summary. It helps to put all the information together for comparison. On comparing the figures horizontally, we can see the traffic patterns for a particular highway for various months on day-to-day basis. On comparing the figures in vertical rows, we can see the traffic patterns for the same month on different highways. For instance, for I-35W, the number of stations congested during the month of January is more than that for the month of March (shown in circled areas). This shows the typical behaviour of the motorists and their preferences and how the patterns are not constant from month to month. This view thus provides a very useful representation of the total analysis.

Figure 11 View comparing the traffic patterns on different highways for different months. (Dimension T_{TD} T_{DW} T_{MY} S)

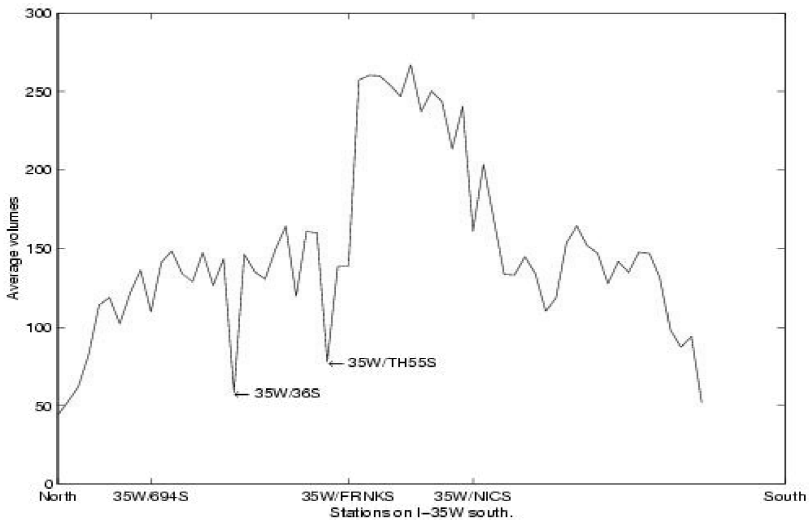


This analysis of the traffic patterns can help the traffic engineers identify the regions that are usually crowded at particular times. This would help them in identifying the risks that might accompany the emergencies that might occur on these roadways.

4.6 Station (dimension S)

This view shows the comparison of traffic volume for various stations on a highway. The X-axis represents the various stations on the highway and the Y-axis represents the average volume at one particular time for those stations. A comparison of traffic volume for various stations at one point of time can help identify the trouble spots for that highway. When used while processing current data, the view can signal congestion in advance by detecting outliers. For instance in Figure 12, a comparison of traffic at various stations on 35W, indicates that there are two problem spots, where the average volume increases suddenly – 35W/36S and 35W/TH55S. This analysis can be very helpful while monitoring the highways for abnormal incidents. Traffic personnel can identify the trouble spot very easily from the representation. The view will prove to be a continuous close-monitoring tool while processing current data. In case of emergencies, this would provide a close-view of the flow of traffic on the highway.

Figure 12 View showing the traffic flow at different stations on a highway at one time



4.7 CubeView and emergency control

CubeView traffic visualisation software can prove to be a useful tool for studying the traffic patterns during emergency situations. In emergency situations, evacuation and the leading traffic congestion are the two major issues to be resolved. Road traffic data analysis tools can help in the traffic management aspect of emergency situations. Computer simulation models of traffic flow are used to estimate the time it takes to evacuate or 'clear' an at-risk region by means of vehicular evacuation (Rathi and Solanki, 1993). CubeView can be used to simulate the traffic patterns during such situations.

The Office of Emergency Preparedness, states a few important things related to emergency control and preparedness, which apply to any emergency response information systems (Turoff, 2002). It states that an emergency system not used on a regular basis before an emergency will never be of use in an actual emergency. This highlights the importance of the archived data and the use of this data for emergency preparedness. Unlike some of the other transportation information systems, CubeView can be used for the analysis of both past and current data. It can be a useful tool for the analysis of traffic patterns in the past. And this analysis can be used to simulate the control actions to be taken during emergency situations. Learning and understanding what actually happened before, during and after a crisis is extremely important for improvement of the response process (Turoff, 2002).

CubeView provides a visual representation of the traffic data. Visual representation provides more insight into the data than the actual data itself. It helps to find patterns and relationships between the elements of data that would otherwise not have been evident enough. It gives a summary of the traffic patterns based on the highways, days of the week, time of day. These views help in the study of traffic flows on the roads. This approach can be helpful in finding out the alternate routes, avoid congestion and help in finding better solutions for dealing with emergencies. As mentioned before, CubeView can play an important role in detection, preparedness, response and recovery from emergency situations. The system can be used to provide a visual representation of current data as well. CubeView uses colour-coding to differentiate the different levels of traffic flows. It can be used to represent three different aspects of traffic flow – total volume, average volume, occupancy. Various shades of green represent a lower than average level of traffic flow. Various shades of yellow show a medium or average traffic flow, while the shades of red represent high levels of traffic flow. This colour-coding can be helpful in determining the specific points on the highways that are near saturation point. In such cases, the congestion can be detected earlier and action can be taken immediately. CubeView also plays an important role in training the traffic personnel to prepare for taking appropriate actions in case real emergencies arise. Since CubeView can be used to analyse the archived data, it can be used to study the traffic patterns in the past. Traffic personnel can be trained through the different views that are created in CubeView. Being a visualisation tool, it helps in providing them with a pictorial representation of the traffic flow on a particular day, or particular highway or a particular month for example. It can thus be a helpful learning tool. At the same time, traffic personnel would learn how to use the tool and get familiarised with the various aspects of the system, which is an important step in preparation for emergencies.

In case of emergencies, CubeView would present a live picture of current traffic flows, using the current data fed into the system. All the traffic personnel are presented with a uniform view of the traffic patterns. Based on the study of the previous traffic patterns, they can determine the routes to which the traffic has to redirected and avoid congestion on particular highways. CubeView can also help in the recovery from the emergency situations. The data collected during the emergency situation can be helpful in finding out the weaknesses of the highway system. New highways or access points can then be proposed and the new model can be simulated in the CubeView. This new model can then be tested using the data collected from the emergency situation, to test how well the new proposal for highways works.

5 Discussion and future directions

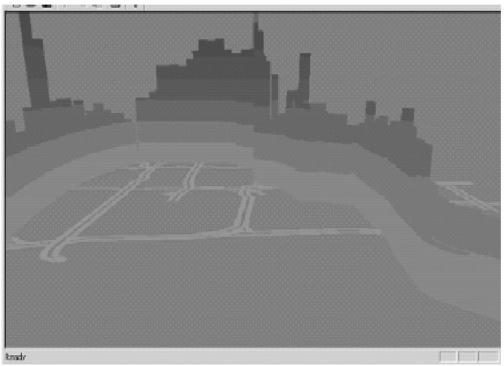
Real-time data processing system should provide and analyse information in very short periods of time. Time is a critical factor in such systems. However, accessing the database over the network and the processing of information does require some time depending on the connection speed and the efficiency of the system in processing the data. An information management system should provide for a robust infrastructure, information search and retrieval, and provide for compatibility of formats among various systems (Information technology Trends Relevant to Crisis Management Summary of a Workshop on Information Technology Research for Crisis Management, 1999; Information Technology Office, 2004). The information management system should be robust enough to operate even in the conditions threatening the system itself. The performance of CubeView system is dependent on the proper functioning of the sensors and the sensor network. If the crisis involves damage or destruction of a part of this network, it is difficult to rely on the real-time data analysis provided by CubeView. In those cases, CubeView might not be able to present a correct real-time picture of the emergency situation.

The system should be able to provide for easy information search and retrieval. In our visualisation software system, there are two possible bottlenecks. One involves the network and the other is the database. Depending on the performance of the underlying database, the information search process results can vary.

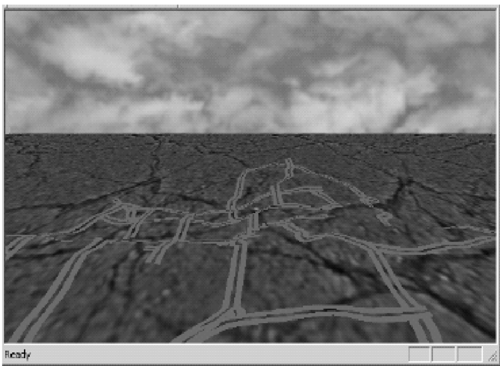
Another important feature of information management systems, as mentioned above, is the compatibility of formats among different information systems. Presently information has to be cleaned and formatted in order to be used in the CubeView system. It cannot be used as is. The amount of useful information in times of crisis, is typically too large for managers and incident commanders to process and utilise to good effect (Jenvald et al., 2001). It is therefore important to simplify and consolidate information so that it is really useful in managing resources and effectively communicating among agencies. CubeView does help in providing a simplified version of the information collected in the database. However, the use of this information in a fruitful manner requires proper training of the people involved in crisis management. In case of emergencies, it is always easier to deal with a known system, instead of relying on a completely new system. Proper training of users would help them to use the vast amount of incoming information in an intelligent manner.

Many new visualisation techniques have been developed in the recent times and are gaining popularity due to much more clearer presentation of data. 3D-visualisation of data is one such presentation. Computer-aided design, video, computer graphics, animation and virtual reality are some of the emerging concepts in the field of data visualisation. Three-dimensional representation of data provides a greater appeal to the eye and is closer to reality than 2D maps. We are currently developing a 3D version of the CubeView. The objectives of developing a 3D system were to create an interactive traffic visualisation system for analysis of loop-detector traffic data collected at the Traffic Management Center. This implementation is proposed to be done in geographic markup language (GML), the language used for virtual environments. The basic map data required for the prototype has been stored in a text file. In the 3D implementation, each station polygon for the highway map is drawn. The coordinates of vertices in each polygon are converted to the world coordinate. The user can easily navigate and control the different views of the system. Figure 13(a) shows one of the features and representation of traffic data in 3D. The average volume has been represented as the height of a polyhedron. Station polygons form the base of the polyhedron. The maximum volume during the day is divided into ten levels. Different colour is then used to represent different levels. A more realistic feeling of the map has been given by using texture-mapping techniques as shown in Figure 13(b). Adding the texture mapping and various real-life features to the view can help in more realistic visualisation of the map. Users' experience would be better than the 2D representation as it creates a feeling of power to control to navigate through the system.

Figure 13 (a) Dynamic representation of the average volume from 0.00 am to 12.00 pm on Jan 1st, 1997 (b) CubeView with texture mapping



(a)



(b)

Presently CubeView processes historical data stored in databases. However, in order for CubeView to be useful during real-time analysis, it is important that it should efficiently process current stream data. There are many factors that make the processing of stream data entirely different from processing historical data. A data stream is a real-time, continuous, ordered sequence of items. It is impossible to control the incoming data and it cannot be stored locally. Also the queries on such data run continuously over a period of time and incrementally return new results as new data arrive (Golab and Ozsu, 2003). Thus time plays an important role in processing such information and the results have to be displayed instantly. CubeView system has to be adopted accordingly to achieve better performance results for processing stream data.

6 Conclusion

The concepts of data mining and visualisation have proven to be important techniques for analysing information in large data sets. CubeView is an attempt to use these techniques for the analysis of traffic data sets. Traffic data visualisation proves to be an extremely useful analysis tool. The techniques of visualisation make the knowledge discovery process much less burdensome and thus can be useful in cases of crisis when time is the limiting factor. Besides visualisation, the techniques used for spatial data mining have been used for the analysis and filtration of traffic databases. One such techniques is the identification of outliers, which plays an important role in identifying abnormal situations and emergency congestions. Although the prototype developed for the CubeView system gives an analysis in 2D, proposals have been made to develop CubeView to provide 3D representation. Such representation will provide users with immersive experience. Future directions of the work involve researching the issues in processing online stream data. Also many different views can be further incorporated in CubeView for efficient emergency analysis. Extensions to the system would also involve designing the system in such a way that user can choose the various dimensions in which data is to be represented and developing outlier detection algorithms in various circumstances.

References

Arizona Department of Transportation and Transportation Technology group (2002) *Freeway Management System, Infrastructure Design Guidelines*, Kimley-Horn and Associates, Inc, September.

Arizona Department of Transportation, http://tpd.az.gov.

Baldonado, M., Woodruff, A. and Kuchinsky, A. (2000) 'Guidelines for using multiple views in information visualization', *Proceedings of the Working Conference on Advanced Visual Interfaces*, pp.110–119.

Chawla, S., Shekhar, S., Wu, W.L. and Ozesmi, U. (2001) *Modeling Spatial Dependencies for Mining Geospatial Data: An Introduction*, in Miller, H.J. and Han, J. (Eds.), Taylor and Francis.

Fayyad, U.M., Grinstein, G.G. and Wierse, A. (2001) 'Information visualization in data mining and knowledge discovery', *Morgan Kaufmann*.

Fayyad, U.M., Piatetsky-Shapiro, G. and Smyth, P. (1996) 'From data mining to knowledge discovery: an overview', *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, pp.1–34.

Fayyad, U.M., Piatetsky-Shapiro, G. and Smyth, P. (1996) 'The KDD process for extracting useful knowledge from volumes of data', *Communications of the ACM*, p.39.

Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (1996) *Advances in Knowledge Discovery and Data Mining*, MIT Press.

Georgia Department of Transportation, http://www.georgia-navigator.com/index.shtml.

Golab, L. and Ozsu, M.T. (2003) 'Issues in data stream management', *ACM SIGMOD Record*, Vol. 32, pp.5–14.

Grinstein, G., Trutschi, M. and Cvek, U. (2001) 'High-dimensional visualizations', *Workshop on Visual Data Mining, Knowledge Discovery and Data Mining (KDD) – 2001.*

Information Technology Office (2004) 'Information management program goals', *Defense Advanced Research Projects Agency (DARPA).* Retrieved from http://www.darpa.mil/ito/research/im/goals.html in January.

Commission on Physical Sciences, Mathematics, and Applications (1999) *Information Technology Trends Relevant to Crisis Management Summary of a Workshop on Information Technology Research for Crisis Management,* The National Academic Press, pp.13–24.

Jenvald, J., Morin, M. and Kincaid, J.P. (2001) 'A framework for web-based dissemination of models and lessons learned from emergency-response exercises and operations', *International Journal of Emergency Management*, Vol. 1, pp.82–94.

Keim, D.A. (2001) 'Visual exploration of large data sets', *Communications of the ACM*, Vol. 44, pp.38–44.

Keim, D.A. and Kriegel, H-P. (1995) 'VisDB: a system for visualizing large databases', *International Conference on Management of Data and Symposium on Principles of Database Systems. Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data*, p.482.

Knorr, E. and Ng, R. (1998) 'Algorithms for mining distance-based outliers in large datasets', *Proc. 24th Very Large Data Bases Conference*.

Koperski, K. and Han, J. (1995) 'Discovery of spatial association rules in geographic information databases', in *Advances in Spatial Databases, Proc of 4th International Symposium, SSD'95*, pp.47–66.

Koperski, K., Adhikary, J. and Han, J. (1996) 'Spatial data mining: progress and challenges', in *Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKS '96)*, pp.1–10.

Los Angeles Department of Transportation, http://trafficinfo.lacity.org/.

Pearce, V. (2002) *Improving Surface Transportation Operations in Emergency Situations*, Federal Highway Administration, Intelligent Transport Systems and Services (ITS) World Congress.

Ramaswamy, S., Rastongi, R. and Shim, K. (2000) 'Efficient algorithms for mining outliers from large data sets', *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, pp.427–438.

Rathi, A.K. and Solanki, R.S. (1993) 'Simulation of traffic flow during emergency situations: a microcomputer based modeling system', *Proceedings of the 1993 Winter Simulation Conference*.

Rundensteiner, E.A., Ward, M.O., Yang, J. and Doshi, P.R. (2002) 'XmdvTool: visual interactive data exploration and trend discovery of high-dimensional data sets', *International Conference on Management of Data and Symposium on Principles of Database Systems. Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data*, p.631.

San Antonio District of the Texas Department of Transportation, *TransGuide, An Intelligent Transportation System*, http://www.transguide.dot.state.tx.us/index.php.

San Diego District 11 Traffic Management Centre, http://www.dot.ca.gov/dist11/d11tmc/sdmap/tmc_main.html.

Shekhar, S. and Huang, Y. (2001) 'Co-location rules mining: a summary of results', *Proc. Spatio- temporal Symposium on Databases*.

Shekhar, S., Lu, C.T. and Zhang, P. (2002) 'Detecting graph – based spatial outlier', *International Journal of Intelligent Data Analysis (IDA)*, Vol. 6, pp.451–468.

Shekhar, S., Lu, C.T., Zhang, P. and Liu, R. (2002) 'Data mining for selective visualization of large spatial datasets', *Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence*.

Swayne, D.F., Cook, D. and Buja, A. (1998) 'XGobi: interactive dynamic data visualization in the X window system', *Journal of Computational and Graphical Statistics*, Vol. 7.

The Minnesota Department of Transportation's Traffic Management Center, http://www.dot.state.mn.us/.

Thearling, K. *An Introduction to Data Mining*, http://www.thearling.com/text/dmwhite/dmwhite.htm.

Turoff, M. (2002) 'Past and future emergency response information systems', *Communications of the ACM Supporting Community and Building Social Capital*, Vol. 45, April, pp.29–32.

US Department of Transportation Research and Special Programs Administration and VOLPE National Transportation Systems Center (1997) 'Surface transportation research and development plan', *Transportation Strategic Planning and Analysis Office*.

US Department of Transportation Research and Special Programs Administration and VOLPE National Transportation Systems Center (1999) 'Surface transportation research and technology assessment', *Transportation Strategic Planning and Analysis Office*, May 1999.