Visual Exploratory Data Analysis of Traffic Volume

Weiguo Han[1], Jinfeng Wang[1], and Shih-Lung Shaw[2]

[1] Institute of Geographic Sciences & Natural Resources Research, CAS,
No. 11A Datun Road, Beijing 100101, China
[2] Department of Geography, University of Tennessee,
Knoxville, TN 37996-0925, USA
hanweiguo@263.net, wangjf@lreis.ac.cn, sshaw@utk.edu

Abstract. Beijing has deployed Intelligent Transportation System (ITS) monitoring devices along selected major roads in the core urban area in order to help relieve traffic congestion and improve traffic conditions. The huge amount of traffic data from ITS originally collected for the control of traffic signals can be a useful source to assist in transportation designing, planning, managing, and research by identifying major traffic patterns from the ITS data. The importance of data visualization as one of the useful data mining methods for reflecting the potential patterns of large sets of data has long been recognized in many disciplines. This paper will discuss several comprehensible and appropriate data visualization techniques, including line chart, bi-directional bar chart, rose diagram, and data image, as exploratory data analysis tools to explore traffic volume data intuitively and to discover the implicit and valuable traffic patterns. These methods could be applied at the same time to gain better and more comprehensive insights of traffic patterns and data relationships hidden in the massive data set. The visual exploratory analysis results could help transportation managers, engineers, and planners make more efficient and effective decisions on the design of traffic operation strategies and future transportation planning scientifically.

1 Introduction

Conventional approaches to tackling transportation congestion problems attempt to increase transportation supply by widening existing roads and building new highways. However, traffic congestion often occurs shortly after, if not before, a transportation improvement project is completed [1]. Intelligent transportation systems (ITS), which aim at improving efficiency of existing transportation systems through the use of advanced computing, real-time data sensors and communication technologies, have been suggested as an alternative approach of tackling transportation congestion problems. With the increasing deployment of ITS services, it appears that they tend to focus on using real-time data to improve traffic operations. The large amount of traffic data collected from ITS can be a useful source to assist in transportation planning and modeling by identifying major traffic patterns from the ITS data. Unfortunately, most ITS data are underutilized for planning and modeling purposes. This paper

A. Gelbukh and C.A. Reyes-Garcia (Eds.): MICAI 2006, LNAI 4293, pp. 695 – 703, 2006.
© Springer-Verlag Berlin Heidelberg 2006

examines several visual exploratory data analysis methods for identification of traffic patterns based on ITS data collected in Beijing, China.

Due to its rapid economic growth, China has seen a fast increase of private automobiles and worsening traffic congestion problems. Beijing, the capital city of China, has experienced an annual growth of 150,000 automobiles in the past four years. According to Beijing Municipal Traffic Management Bureau, 40 percent of wage earners in Beijing spend at least one hour for one-way commute between their homes and workplaces on each workday. Beijing municipal government has realized that, if it does not address the worsening traffic congestion situation, it will become a major problem to the city's future development and the 2008 Olympic Games. Many traffic regulation approaches that have been successfully implemented in other countries have been adopted in Beijing. In addition, Beijing has deployed ITS monitoring devices along selected major roads in the core urban area of Beijing in order to help relieve the city’s traffic congestion and improve the city’s traffic conditions. Currently, the system collects real-time data, such as travel speed and traffic volume, and transmits the data to a database server at the traffic control center. The data are mainly used to assist in real-time control of traffic signals. It has been realized that the data should be utilized to extract hidden and valuable traffic flow patterns to support other functions such as performance monitoring, operations evaluation, transportation planning and transportation policy making.

Data mining is an approach of discovering useful information, knowledge, and rules hidden in large data sets [2]. As an important tool for data mining, data visualization displays multi-dimensional data in the forms that reflect information patterns, data relationships and trends in order to help users observe and analyze the data more intuitively. Data visualization also allows users to control and steer the data mining process based on the given visual feedback. Users therefore can take advantage of their experience and knowledge to discover implicit and valuable patterns and data relationships hidden in large data sets [3]. This paper examines several data visualization techniques, including line chart, bi-directional bar chart, rose diagram, and data image, as exploratory data analysis tools to identify hidden traffic flow patterns from the ITS data collected in Beijing. The visual exploratory analysis covers different geographic scales from street intersections to main highway arterials, which require different visualization methods to discover the hidden traffic patterns. The remaining parts of this paper are organized as follows. Section 2 is a brief review of data visualization and exploratory data analysis. Section 3 presents the visualization techniques used to explore and analyze traffic volume data. The final section offers concluding remarks and future research directions.

2 Data Visualization

Data visualization is a process of transforming information into a visual form, enabling users to observe the information. The resulting visual display enables scientists or engineers to perceive visually the features that are hidden in the data but nevertheless are needed for data exploration and analysis [5]. It is often easier to detect a pattern from a picture than from a numeric output. Graphical illustrations such as plots and graphs therefore play an essential role as tools for analysis and exploration of inherent structure in data [6]. As one of the important tools in data mining, data visualization not only assists knowledge discovery but also controls the process of data analysis. There is no single general visualization method suitable for all problems. It is important to choose appropriate visualization methods according to the task and the property of data in order to provide critical and comprehensive appreciation of the data that will benefit subsequent analysis, processing, and decision-making.

Transportation practitioners at Beijing transportation departments now use traditional online transaction processing (OLTP) of database in tabular format to evaluate, summarize, and report the current traffic status. This conventional approach makes it difficult for them to discover hidden traffic patterns in the data and to provide more specific analysis and future plans of the existing system to help relieve the worsening traffic congestion problems. The goal of this study is to provide useful visual data analysis methods for transportation managers, engineers, and planners to explore traffic volume data intuitively and to discover the hidden traffic patterns. The visual analysis results in turn could help them make more effective decisions on the design of traffic operation strategies and future transportation planning. Choosing appropriate visualization methods that are suitable for traffic volume data and can effectively convey the information to transportation managers and engineers is not a trivial task. For example, if a visualization technique such as the parallel coordinate is used to represent the traffic volume data, it may be too complex for transportation practitioners to easily interpret and compare the data.

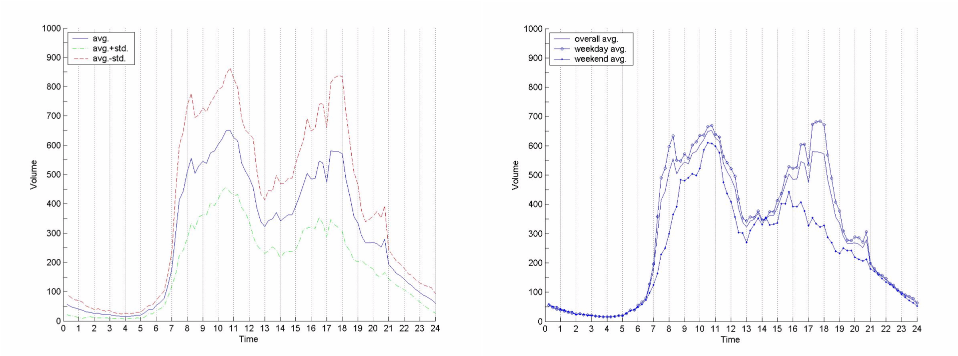
Catarci et al. provide a set of logic rules to select effective visual representation and graphic design for visualizing the facts and data relationships [7]. Bertin also offers guidelines on how to choose the suitable visual methods to reflect data attributes [8]. Based on these guidelines reported in the literature, this study presents several visualization methods that are appropriate for representing traffic volume data and are comprehensible for transportation managers and engineers to perform effective data exploration and analysis. These methods include line chart, bi-directional bar chart, rose diagram, and data image. They could be used at the same time to gain better and more comprehensive insights of traffic patterns and data relationships hidden in the massive data set.

3 Visual Analysis of Traffic Volume Data

Traffic volume data possess several characteristics that require different visualization methods to clearly illustrate and communicate these characteristics. It is more effective for transportation managers and engineers to analyze the figures generated from these methods to convey the traffic characteristics efficiently and concisely than to read through pages or tables of data describing the traffic status. This section presents a select set of visualization methods for exploration and analysis of traffic volume data at different levels of spatial granularity (i.e., at street intersections and along main arterials) to analyze existing traffic demand, and identify ways to improve traffic flow.

3.1 Line Chart

Line chart is a simple and easy-to-understand method to show trends or changes of traffic volume over a period time or over a distance range. It is easy to identify traffic peak periods from a line chart.



(a) (b)

Fig. 1. Line chart of the westbound average traffic volume

Using the westbound traffic volume at Xidan Intersection, which is located in the center of Xidan Culture and Shopping Area and near many central and municipal government agencies, as an example, Fig. 1(a) shows the average traffic volume curve, along with the standard deviation curves above and below average curve, and Fig. 1(b) presents the overall, weekday, weekend average traffic curve of these months respectively. These charts indicate that the location has a large daily variation of traffic volume, with two daytime peak periods occurring during working hours between 8 am and 12 am and between 2 pm and 6 pm, and one nighttime peak periods during between 8 pm and 10 pm, and the volume trend of weekends is similar to that of weekdays, but the weekend average does not demonstrate the pronounced morning peak at 8 am that is common on weekdays, the morning peak hour on weekends exists between 11 am and 12 am. Transportation engineers can use the information to assist them in evaluating the road capacity, adjusting the traffic signal timing. Transportation planners, on the other hand, can use the chart to figure out "time in a day" traffic distribution pattern for travel demand modeling.

3.2 Bi-directional Bar Chart

When traffic volumes in both directions are important to control traffic signals or to plan the number of traffic lanes, it is better to represent the data with a bi-directional bar chart. Fig. 2 shows the eastbound and the westbound average traffic volumes at Jingxi Hotel Intersection. It clearly illustrates the directional difference of traffic flows. The eastbound overall average traffic volume is much higher during the morning peak hours than the westbound overall average volume, while a reversed pattern occurs during the afternoon peak hours, so does the weekday one. The workday and

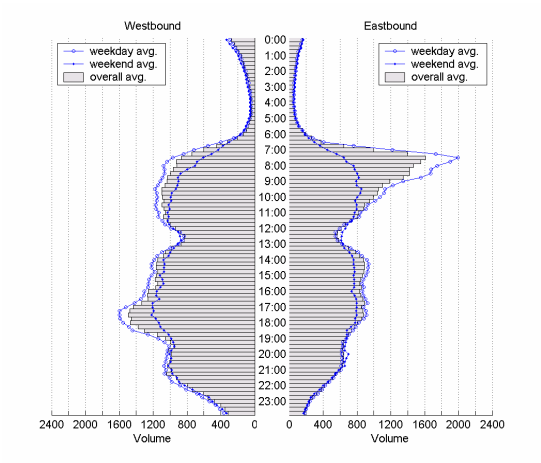


Fig. 2. Bi-directional bar chart of eastbound and westbound traffic volume

weekend average volume curves in the plot also demonstrate varying patterns of traffic, such as a significant difference of eastbound traffic volume on the morning between weekdays and weekends, and a bi-directional same trend on weekends. In addition, this chart allows an easy comparison of traffic volumes in the two opposite directions during any selected time period. These patterns reflect truly the spatial distribution patterns of workplaces and residences of citizens along the West Chang'an Street in Beijing, and more people in Beijing prefer to choose West Chang'an Street as their westbound road for business or home. Information derived from this chart can help transportation engineers to set different lengths of signal cycles at the intersection for different time periods in a day. It is also useful for the transportation department to consider the creation of reversible lanes along this street.

3.3 Rose Diagram

Transportation analysts also need to examine traffic flow data of straight flows and turning movements at each intersection. In this case, rose diagrams can be used to show percentages of each directional flow at an intersection. Transportation engineers then can use the information to adjust traffic signal phases to facilitate traffic flows. Fig. 3 shows an example of directional average traffic flows at 8 AM at the Mingguangcun Intersection (e.g., "S" for southbound straight flows and "SE" for southbound-to-eastbound turning flows, so do the remainder figures in this paper), which is a very congested intersection near the Xizhimen Subway Station. The rose diagram clearly shows that eastbound and left-turn traffic account for most traffic flows at this intersection at 8 AM. Note that right-turn traffic flow data are not included on this rose diagram since they are not currently recorded by the traffic control center in Beijing. Right-turn traffic flows at many intersections in Beijing however do cause significant interferences to other traffic flows (e.g., left-turn traffic, bicycle and pedestrian flows). Beijing transportation department should consider to record right-turn traffic volumes and include them in future traffic analysis.

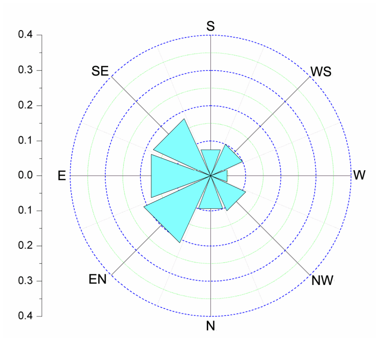


Fig. 3. Rose diagram of traffic volume of eight directions

3.4 Data Image

As traffic flows change over time, especially by time-of-day, transportation engineers need to develop multiple signal timing plans to accommodate these changes. Clearly, the rose diagram does not show the actual traffic volume value variations of each direction. Data image has been suggested as an approach of mapping data attributes to color features for visualization and exploration of higher dimensional data [9]. Marchette et al. suggest that data image method can be used to detect data outliers [10]. Healey further indicates that data image can quickly differentiate elements based on their colors for exploratory data analysis of identifying clusters or performing classification [11].

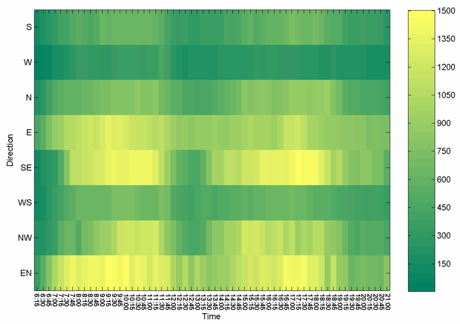


Fig. 4. Data image of average traffic volume of eight directions

Fig. 4 displays average traffic volumes from 6:00 to 21:00 for eight directional flows at the Mingguangcun Intersection in a data image. In this figure, the horizontal axis represents 15-minute time intervals and the vertical axis shows the eight directions. Again, right-turn flows are not included in this figure. Light yellow color indicates the heaviest traffic condition, and deep green color is for the lightest traffic.

In order to further help transportation engineers obtain a clear picture of the traffic flow patterns among different directions over various time period, a cluster analysis is performed to group together similar traffic patterns among different time intervals.

A hierarchical clustering algorithm based on the complete linkage clustering scheme is used to identify the clusters [9] considering its advantages over than the other clustering algorithms, such as less sensitivity to the input parameters and ease of handling of any forms of similarity (i.e. nested partitions) [12]. The leaf nodes are the directional volumes at each individual time interval, while intermediate ones indicate larger groups of homogeneous volume at several time intervals. Fig. 5 shows the hierarchical structure and the same traffic flow data with four clusters identified as bands marked on the data image. Each band suggests a particular signal timing plan that is more appropriate for specific time intervals. This enables transportation engineers to get a better idea of the time-of-day variations to help them adjust and fine-tune signal timings at each intersection. This method offers an inexpensive way to evaluate the existing operational strategies and could be used to automate the design of signal timing plans.

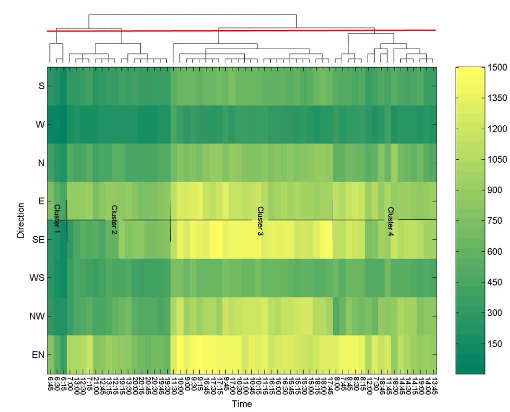


Fig. 5. Data image of average traffic volume sorted using clustering algorithm

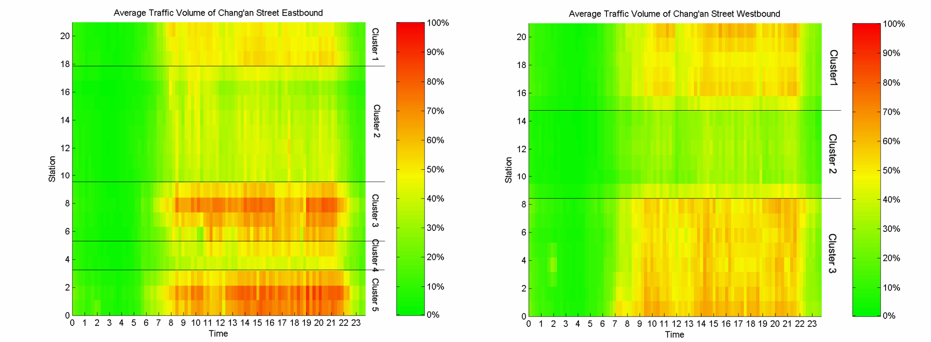
Data images also can be used to display traffic volumes along a particular street. Fig. 6 shows examples of using data images to explore the average flow patterns for westbound and eastbound traffic along Chang'an Street, respectively. The horizontal axis again represents 15-minute time intervals. The vertical axis displays sequentially the data collected at all traffic detectors along Chang'an Street starting from Jianguomen on the east to Fuxingmen on the west. The color index is computed by

$$Colorindex(i, j) = \left[\frac{Volume(i, j)}{MaxTraffic} * Colornums \right] \quad (1)$$

where,

[…]: operator of rounding to the nearest integer
Volume(i, j): traffic volume at detector i for time interval j
MaxTraffic: maximum capacity of the road segment
Colornums: color numbers used in the color map, here is 100.

Considered that the detectors are spatially contiguous, the classical K-means partitioning clustering algorithm is used to identify the clusters in Fig. 6, it aims to divide the data set into several homogeneous clusters, which may not overlap with each



(a) (b)

Fig. 6. Data image of the average traffic volume of Chang'an Street

other but together cover the whole data space [13]. Another reason of applying this approach is that given the number *k* of partitions to be found, it is very often the case that the *k* clusters found by a partitioning method are of higher quality (i.e., more similar) than the *k* clusters produced by a hierarchical method [14]. The conceivable number, which allows the algorithm to identify the clustering structure of traffic volume datasets, could be easily discerned from the figures.

The traffic volume data collected at each detector is modeled as an *n*-dimensional vector with the form of $Vol_i = <v_1, v_2, \ldots v_j, \ldots, v_n>$, where v_j denotes the average traffic volume at detector *i* for time interval *j* and *n* is the total number of time intervals. Euclidean distance is used to measure the distance between the vectors. Five clusters are identified for the westbound traffic along Chang'an Street (see Fig. 6 (a)), and three clusters are found for the eastbound traffic (see Fig. 6(b)). Results of the cluster analysis can help transportation engineers determine better ways of synchronizing traffic signals located along a major street. The techniques of data image also could be used to evaluate traffic situation to identify the irregular patterns or locate the faulty detectors.

4 Conclusion

In this paper, several visualization methods, which are appropriate and comprehensible for visual exploratory analysis of the ITS Data, are applied to discover traffic patterns and data relationships hidden in the massive data sets. Transportation practitioners can take advantage of these techniques to extract hidden and valuable traffic flow patterns to help monitor the system performances, evaluate traffic situations, adjust the traffic signal timing, make transportation plan or policy, and so on.

Nevertheless, the functions provided by these solutions are not comprehensive in terms of the analysis and visualization of urban traffic data, and data visualization is not a substitute for data analysis, instead it complements data analysis to yield greater insights into the traffic data. The visualization tools should be utilized and developed to improve understanding of behaviors in time and space [15]. In the next future, we will program to implement an integrated framework based on data visualization, data mining, Web and GIS to provide a powerful and real-time on-line tool for transportation managers and researchers to analyze traffic situation, improve traffic condition, etc, for transportation engineers and planners to evaluate traffic capacity, design traffic signal plans, etc, and for drivers or travelers to select their routes, etc.

References

1. Miller, H.J., Shaw, S.L.. Geographic Information Systems for Transportation: Principles and Applications. Oxford University Press, New York, 2001.
2. Han, J., Kamber, M.. Data Mining: Concepts and Technologies. Morgan Kaufmann Publisher, San Francisco, 2001.
3. Miller, H.J., Han, J.. Geographic Data Mining and Knowledge Discovery. Taylor and Francis, London, 2001.
4. Gershon, N.. From perception to visualization. In: L. Rosenblum et al., (Eds.), Scientific Visualization 1994: Advances and Challenges, Academic Press, New York, 129-139, 1994.
5. Cristina, M., Oliveira, F.D., Levkowitz, H.. From visual data exploration to visual data mining: a survey. IEEE Transactions on Visualization and Computer Graphics, 9(3), 378-394 (2003).
6. Edsall, R.M.. The parallel coordinate plot in action: design and use for geographic visualization. Computational Statistics & Data Analysis, 43(4): 605-619 (2003).
7. Catarci, T., Santucci, G., Costabile, M.F., Cruz, I.. Foundations of the DARE system for drawing adequate representations. In: Proceedings of International Symposium on Database Applications in Non-Traditional Environments, IEEE Computer Society Press, 461-470, 1999.
8. Bertin, J.. Semiology of Graphics. University Wisconsin Press, Wisconsin, 1983.
9. Minnotte, M., West, W.. The data image: a tool for exploring high dimensional data sets. In: Proceedings of the ASA Section on Statistical Graphics, Dallas, Texas, 25-33, 1998.
10. Marchette, D.J., Solka J.L.. Using data images for outlier detection. Computational Statistics & Data Analysis, 43(4), 541-552 (2003).
11. Healey, C.G.. Choosing effective colors for data visualization. In: Proceedings of IEEE Visualization 1996, IEEE Computer Society Press, 263-270, 1996.
12. Kaufman, L., Rousseeuw, P.J.. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York, 1990.
13. Jain, A.K., Murty, M.N., Flynn, P.J.. Data clustering: a review. ACM Computing Surveys, 31(3), 264-323 (1999).
14. Raymond, T. Ng, Han, J.. CLARANS: a method for clustering objects for spatial data mining. IEEE Transactions on Knowledge and Data Engineering, 14(5), 1003-1016 (2002).
15. Dykes, J.A., Mountain D.M.. Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications. Computational Statistics & Data Analysis, 43(4), 581-603 (2003).