Available online at www.sciencedirect.com
ScienceDirect
Procedia Engineering 137 (2016) 709 – 718
Procedia Engineering
www.elsevier.com/locate/procedia

GITSS2015

Research on Visualization of Multi-Dimensional Real-Time Traffic Data Stream Based on Cloud Computing

Jia Chaolong[a,]*, WANG Hanning[b], WEI Lili[c]

[a]*School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China*
[b]*State Key Laboratory of Rail Traffic Control and Safety, Beijing JiaoTong University, Beijing, 100044, China*
[c]*Chongqing Municipal Public Security Bureau, Chongqing, 401147, China*

Abstract

Based on efficient continuous parallel query series algorithm supporting multi-objective optimization, by using visual graphics technology for traffic data streams for efficient real-time graphical visualization, it improve human-computer interaction, to realize real-time and visual data analysis and to improve efficiency and accuracy of the analysis. This paper employs data mining processing and statistical analysis on real-time traffic data stream, based on the parameters standards of various data mining algorithms, and by using computer graphics and image processing technology, converts graphics or images and make them displayed on the screen according to the system requirements, in order to track, forecast and maintain the operating condition of all traffic service systems effectively.

Peer-review under responsibility of the Department of Transportation Engineering, Beijing Institute of Technology

Keywords: Smart Transportation; Visualization; Cloud Computing; Parallel Query; Big Data

1. Introduction

IDC report shows that the total global data in 2020 will be more than 35.2ZB (the equivalent of 4 trillion GB, equivalent to 8 billion 4TB hard disk), which is 44 times of the data in 2011. In the past few years, the amount of data around the world grows with an annual growth rate of 58%, which will be faster in the future. If calculated in accordance with current storage capacity growth rate of 40% per year, the amount of data needs to be stored until 2017 will be greater even than the total capacity of the storage device. How to solve problems of science, health care,

* Corresponding author. Tel.: +86-023-62487732.
E-mail address: jiachaolong@bjtu.edu.cn

Peer-review under responsibility of the Department of Transportation Engineering, Beijing Institute of Technology
doi:10.1016/j.proeng.2016.01.308

energy, commerce, government, urban construction and other fields by using big data is the problem facing the whole world.

Big Data is permeating into each cell of social organization, generating almost subversive and revolutionary impact on all sectors. But only with large data is not enough, especially when data quality is generally not high. What comes with the vast amounts of data is vast amount of data noise. With no effective management and analysis on the big data, big data cannot play its role effectively. This is the problem facing the academia and industry, as well as the key issues in top building to root realizing.

Under the context of cloud computing and big data, visualization technology has been given importance as an effective data analysis means and has attracted the attentions of more and more scholars.

Kehrer J [1] studies existing methods for visualization and interactive visual analysis of multifaceted scientific data. Toker D [2] investigates the relationship between such characteristics and fine-grained user attention patterns. These results are discussed in view of our long-term goal of designing information visualization systems that can dynamically adapt to individual user characteristics. Chen X [3] employs the characteristic of QoS and achieves considerable improvement on the recommendation accuracy different from previous work. To help service users better understand the rationale of the recommendation and remove some of the mystery, they use a recommendation visualization technique to show how a recommendation is grouped with other choices. Ahn J [4] proposes a specific way to integrate interactive visualization and personalized search and introduces an adaptive visualization based search system Adaptive VIBE that implements it. Ben X [5] examined the contribution characteristics of developers in open source environment based on visual analysis, and presented approaches from three aspects-influencing factors, time characteristics and region characteristics. Corchado E [6] study introduces and describes a novel intrusion detection system (IDS) called MOVCIDS (mobile visualization connectionist IDS).By its advanced visualization facilities, the proposed IDS allows providing an overview of the network traffic as well as identifying anomalous situations tackled by computer networks, responding to the challenges presented by volume, dynamics and diversity of the traffic, including novel (0-day) attacks. Adeshina A M [7] attempts to adapt SurLens to possible visualization of abnormalities in human anatomical structures using CT and MR images and study shows SurLens' functionality as a 3-D Multimodal Visualization System. Gama S [8] used visualization method studied to which extent color blending provides users with the means to understand the provenience of data items by conducting a user study with 73 subjects using CIE-LCh blending to ascertain (i) to which extent people are able to, given a particular color, understand its provenience, and (ii) the color model in which to perform color blending so that users find blending intuitive. Aiming at the big data generated by war gaming. Xu X [9] proposed a visualization algorithm based on regular radius and constrained random direction is . Lam H [10] encapsulate the current practices in the information visualization research community and provide a different approach to reaching decisions about what might be the most effective evaluation of a given information visualization. Shahrestani A [11] proposes a proactive approach by adopting proper visualization techniques to increase the visibility of network traffic related to invariant bot behavior and botnet activities. Peck E M M [12] uses the classic comparison of bar graphs and pie charts to test the viability of fNIRS for measuring the impact of a visual design on the brain. Suzuki H [13] develops a Compton camera for quick visualization of the radioactive contamination. It features high detection efficiency by utilizing gamma ray detectors. Bertini E. [14] provides an overview of approaches that use quality metrics in high-dimensional data visualization and propose systematization based on a thorough literature review. Goyal N [15] tests the utility of a visualization of data links and a notepad for collecting and organizing annotations. The visualization significantly improved participants' ability to solve the crime whereas the notepad did not.

Morris, B.T [16] combines high-resolution real-time traffic data with instantaneous emission models to estimate these environmental measures in real time and presents a system that estimates average traffic fuel economy, CO_2, CO, HC, and NO_x emissions using a computer-vision-based methodology in combination with vehicle-specific power-based energy and emission models. Anwar, A [17] propose Traffic Origins, a simple method to visualize the impact road incidents have on congestion to aid Traffic management controllers decision making and help them understand how past incidents affected traffic. Zhang [18] proposes a visualization method to represent traffic flow as a texture image. O'Brien W.J. [19] discusses the process of development of an integrated visualization methodology for viewing traffic and geometry information and demonstrates its benefits and challenges using case studies. Song [20] describes an exploratory visualization toolkit for large traffic flow databases. It is based on the concept of the traffic cube: an extension of the data cube in data mining. Zahran [21] contributes to the knowledge by devising a new three-dimensional (3D) visualization approach for modeled air quality before and after the implementation of potential urban transport schemes. Sewall J [22] presents a novel concept, Virtualized Traffic, to reconstruct and visualize continuous traffic flows from discrete spatiotemporal data provided by traffic sensors or generated artificially to enhance a sense of immersion in a dynamic virtual world.

2. Visualization Theory and Implementation

Visualization or scientific visualization, analyses data and calculation results in depth in order to obtain the understanding and insight on data, converts data information involved and generated in the process of computing to intuitive, represented with image or graphic, physical phenomena or physical quality changing with time and space, and display them in front of professionals, so that they will be able to observe the simulation and calculation process, which is invisible in traditional sense. Also, it provides means of visual interaction with simulation and calculation. The purpose of visualization is to promote a deeper understanding of the observed data, and to develop a new insight into the underlying processes relying on the powerful human visual capacity. Visualization technology is a combination of scientific computing and graphics technology, which involves multiple disciplines and technologies in the field of science and engineering computing, computer graphics, image processing, human-machine interface. As an emerging technology, it gained rapid development in the various disciplines and has been applied widely since it's born. With the advent of the era of big data, visualization technology has produced more and more research, and the most representative research fields include scientific visualization, data visualization, information visualization, knowledge visualization. Figure 1 shows the information contained in various research fields.

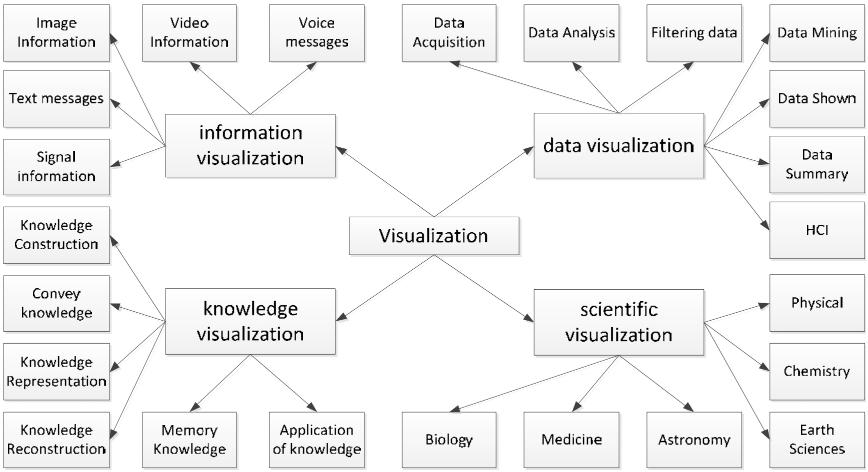

Fig. 1 Content of the Visualization study

2.1. Data Visualization

A new research field rose in early 1990s, known as the "information visualization", which provides support for analysis work of abstract heterogeneous data sets among the many applications. This freshman term "data visualization" was gradually accepted which covered scientific visualization and information visualization.

Data visualization mainly targets data in large databases, and it expresses the relationship between the data and the data visually by using parallel coordinating method, surface-pixel method and graphical method, to obtain inherent information in data and do in-depth observation and analysis. Data visualization mainly includes seven steps: data acquisition, data analysis, data filtering, data mining, data display, and data summary and human-computer interaction.

Data visualization can be divided into six sub-areas from the perspective of computer science, which is shown in Figure 2:

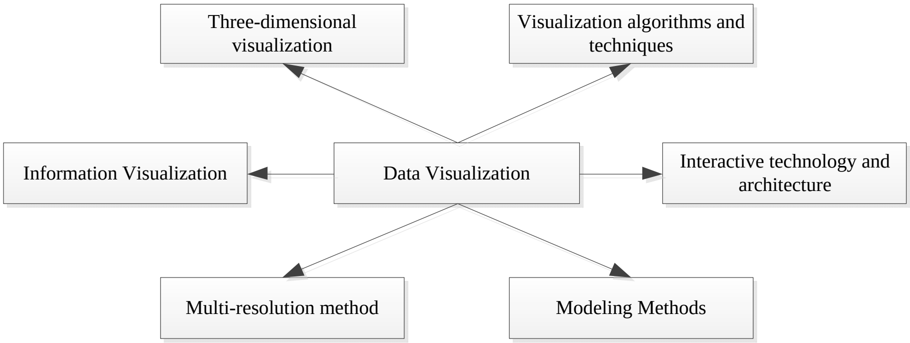

Fig. 2 research field of data visualization

2.2. Scientific Visualization

Scientific visualization is the theory, method and technology that employs computer graphics and image processing technology, converts the output data of scientific computing and data generated by observation in other fields to graphics and images, and ultimately displays them on the screen and achieves interactive processing.

Scientific data visualization is a complex process from scientific data unrelated to graphics to graph represented by a pixel finally, it includes: Generating, pretreatment, mapping, rendering and display, the conversion step and the logic flow, which is shown in Figure 3.

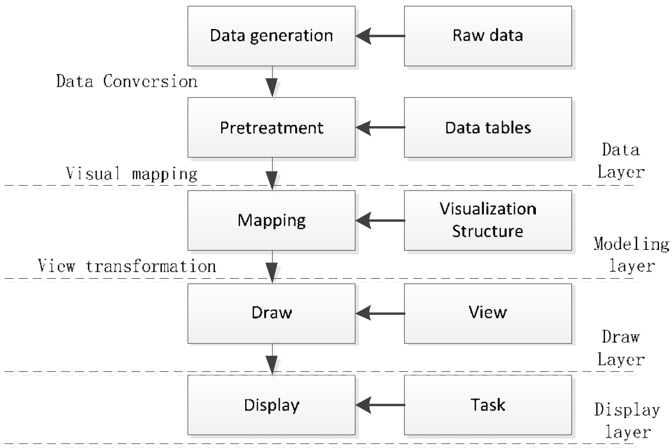

Fig. 3 conversion step and the logic flow of scientific data visualization

2.3. Information Visualization

Information Visualization is the technology supported by computer, represents abstract information in a visual form by using interactive tools, to reveal the relationship between the abstract and the characteristics of information. Information visualization has been independent from the data visualization as a branch field from the early 1990s. Information visualization is generally applied to visualization for large-scale non-digital information resources. Information visualization is committed to creating those means and methods to convey abstract information visually. Information visualization summarizes the visualization process as adjustable mapping form data to visualization, and then to human perceptual system. Data retrieval is the core of information visualization. Cognitive psychology and graphic design are two basis of information visualization. Currently information visualization research focuses on the visualization of hierarchical information, multidimensional information visualization, text information visualization, and web visualization aspects.

2.4. Knowledge Visualization

Since the beginning of the 21st century, knowledge visualization has been developing on the basis of scientific visualization, data visualization and information visualization, which is an integrated complex emerging research field involving multi-discipline. Knowledge Visualization mainly focuses the role of visual expression means between more than two persons in improving knowledge creation and transfer. When we get useful knowledge about the data, by using computer graphics technology and image processing technology, we express, construct or convey complex knowledge to other people, to facilitate their understanding and using, and to facilitate the spread of knowledge. This process converts implicit knowledge to explicit knowledge, and creates new knowledge to promote learning, understanding, and collaboration.

The worlds of information include three basic elements: data, information and knowledge. Correspondingly, three kinds of visual forms relative to them are data visualization, information visualization and knowledge visualization. Knowledge visualization can be considered as the highest stage of visualization development. Meanwhile, generation and development of scientific visualization and data visualization also belong to the same period.

3. Traffic flow information visualization

Analysis and processing technology of traffic flow data is one of core content of information processing in Smart Transportation, supporting the collaborative operation between the various traffic systems, and the development of the technical level determines the service level of Smart Transportation system. Take traffic management and planning as the example, the Smart Transportation in this area mainly includes three aspects: Advanced Traffic Management Systems, Transportation infrastructure intelligent monitoring system and Transportation planning decision supporting system. Among them, multi-means and full-scale system of traffic information collection and the condition monitoring of the Road Network, automatic traffic checkpoint monitoring system and all kinds of advanced electronic police monitoring system have return massive dynamic traffic flow data (Beijing, more than 2000 surveillance cameras with 7.2PB every year, about 70000 taxi GPS with 5TB every year, geographic information of 20000 kilometers streets with 15TB a year). Another example, in travelling service field with the rich service content, real-time traffic data stream processing technology is the basis of real-time traffic notification and reception of intelligent traffic guidance system and vehicle navigation system, traffic control, the optimal and dynamic route guidance.

Traditional intelligent traffic focuses on the Automatic function to replace the artificial features such as Automatic toll collection of vehicles, license plate recognition, and image contrast. In the background of the emergence of new technology, it apparently has been unable to meet the needs of the traffic management department. Modern Smart Transportation is based on real-time traffic data, provides the real-time traffic information of traffic data, combines the internet of things, cloud computing and other high-tech IT technology to collect traffic information, uses a lot of data processing technique such as data model and data mining.

Visualization is an important feature of Smart Transportation. Research contents of the modern Smart Transportation and traditional intelligent transportation system are shown in figure 4.

The rough information release and simple traffic flow prediction in traditional traffic information management has been unable to meet the needs of modern Smart Transportation. It needs a more advanced technology. Visualization technology, with its superiority of intuitive and clear graphics has been paid more and more attention. Traffic visualization is an important application field of visualization.

The traffic information data have four basic aspects of big data: volume, variety, velocity and value. Volume means of the traffic information data is large; Variety means data are provided in a large amount by a lot of traffic information detection equipment and means, and data is generated from every corner throughout the city; Velocity means of the acquisition of traffic information data is very fast thanks to the computer information system; value refers to the traffic information data value is very high, from which we can explore the information that urban traffic management requires.

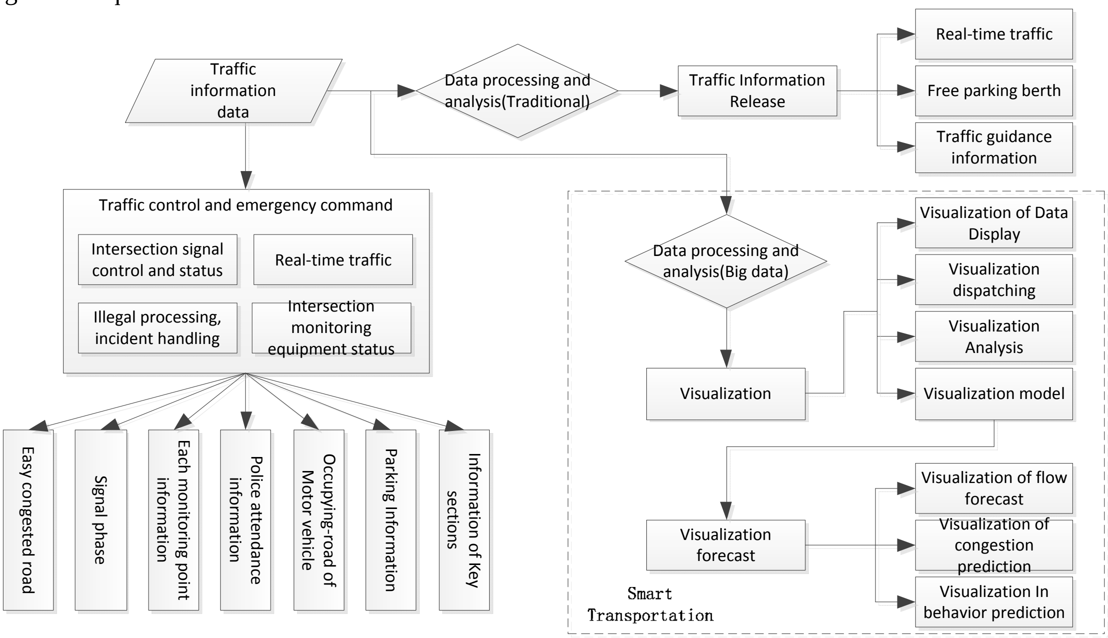

Fig. 4 Research contents of the modern Smart Transportation and traditional intelligent transportation system

With the use of visualization method, a large number of traffic data can form visual traffic information with the human-computer interaction, accuracy, reliable and high efficiency. We can understand its internal law through the transformation of graphics and image. Traffic flow visualization simulation can display the traffic simulation data directly by the computer system and graphics system. Thus, traffic management and control personnel can conveniently master city traffic status and issue traffic control instructions timely and accurately, and to ensure the traffic safety and smooth of the city. The overall framework of traffic flow data visualization system is shown in figure 5.

Data visualization based on dimensionality reduction mapping is an effective multidimensional data visualization technology. It considers multidimensional data set as a whole, and do visual display to the inner structure and topological relationship of data set in a low dimensional space by using optimization method [23], such as, the method of principal component analysis (PCA) [24] which combines the dimension in linear, and extracts the "dominant dimension" which contains most and independent information (with the maximum dispersion) to represent multidimensional data; multidimensional scaling (MDS) [25] uses the dissimilarity of data between points with good distance and approximation high dimension to reconstruct multidimensional data in the low dimensional space; self-organizing map [26] uses neural network neighborhood learning method to recombine to produce a "new dimension", in order to express the original multidimensional data; isometric mapping (ISOMAP) and locally linear embedding method(2009) (LLE) (2008) uses manifold learning algorithm, with the shortest field path length approaching the general geodesic distance as the input of MDS algorithm. Although this kind of technology can effectively solve the problem of The Curse of dimensionality [29], and has been applied successfully in many practical problems, though, because of its high computational complexity, it is not suitable for the visualization of the massive data flow. And since the technology tries to lower dimensions based on the protection of the main features of data sets, and then shows the multidimensional data in 2, 3 dimensional visual spaces, so it cannot directly reflect the distribution of multidimensional objects data stream in each dimension.

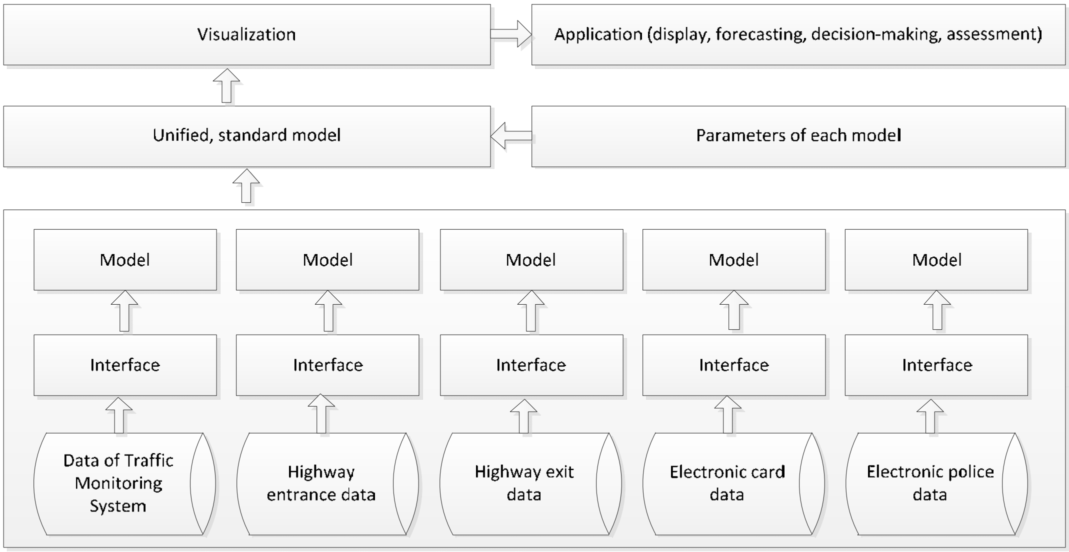

Fig. 5 overall framework of traffic flow data visualization system

4. Multidimensional uncertain traffic flow data visualization

In this paper, traffic data of Chongqing in 2014 is used for the study, by the end of 2014-11, the amount of traffic checkpoint data is 770 million, the amount of highway entrance data is 540 million, the amount of highway exit data is 530 million and the amount of e-cards data is 2.8 billion, and all the data goes up with a rate of 40% each year.

Part of the statistical results of Chongqing traffic data in 2014 is shown in Table 1 and traffic flow of eight traffic collecting point one day is shown in Figure 6.

Due to the transportation system belongs to the discrete event system, the internal state of system change is random, the same internal state can be changed to a variety of state, it is difficult to describe the change state in the internal of the system used function, only can grasp the statistical regularity of the internal state of the system changes.

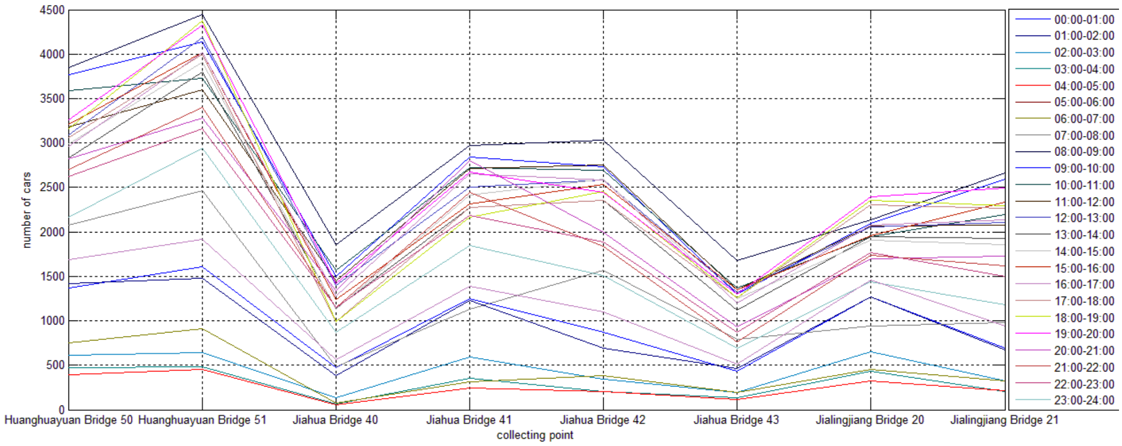

Fig. 6 traffic flow of eight traffic collecting point in 2015-04-27

Table 1 Part of the statistical results of Chongqing traffic data in 2014

data	by the end of 2014-08	by the end of 2014-09	by the end of 2014-10	by the end of 2014-11
traffic checkpoint data	696253489	728307619	743345445	770015777
Highway entrance data	521749409	523940794	527444742	540978983
Highway exit data	513829670	515973868	519450339	533026804
Electronic card data	2477198796	2569583864	2719830079	2797430922

In order to improve the real-time and accuracy in traffic decision ultimately, Traffic management department need fully understand the needs of users in real time decision for massive data flow based on the background knowledge of city traffic field; analysis overall structure of multidimensional query result set, information of distribution of each dimension and clustering, optimize the corresponding relationship between the original data space dimension and initial surface of the dimensions, improve the order of dimensions shaft; strengthen ability to deal with categorical data, combined with the method of dimensionality reduction of multidimensional complex data and design of optimized hierarchical relationships, and to enhance the user's cognitive ability to analyze data with dynamic interactive manner.

Visualization analyses combined with static graphs and Dynamic interaction can assist transportation business decisions by stimulating the user's visual thinking. For example, by multi-view display to the same result set, allowing users to have a synchronized view of more than a single result set, and allow it to select a particular data point take the form of a click or circle to, and do visualization analysis.

It can presenting the connotation from traffic data stream query result set through massive traffic data flow visualization analysis, and hold the relationship of data in "global" and "local", division the importance of the data rational, taking into account the dynamic display of time dimension and interactive layering, and then mining out the inside law hidden in large data, providing real-time decision making for urban transport services.

5. Conclusions

In this paper, the real-time graphic visualization technology is employed to improve the human-computer interaction, improve the efficiency and accuracy of the visual analysis of massive traffic data. Visualization analysis of massive traffic data stream presents connotation of traffic data stream query result set, grasps the relationship between data "global" and "local", makes a reasonable division of the degree of importance of data, taking into account the dynamic display of time dimension and the level sense of interaction, and then digs out the law hidden within big data, to provide real time decision for city traffic. It can bring the following values to the Smart Transportation system:

Ⅰ. Reduce the deaths of malignant traffic accident, collect vehicle information and run real-time analysis through the monitoring system, monitor behavior of high-accidental vehicles (such as engineering truck), and reduce the accident rate.

II. The road congestion rate will be decreased. It can collect traffic information and run real-time processing through traffic monitoring equipment, which can accurately draw line graph of road congestion, and provide traffic control department for dealing with traffic accidents, and provide the public for reference to divert traffic.

In short, the traffic is the artery of the economic operation in the process of industrialization. Intelligent transportation is an important part of intelligent city. Through networking front-end data acquisition, and video surveillance, collecting traffic video camera information, comparing to the historical record, analyzing the flow of real-time traffic and people, we can calculate and predict the road traffic conditions in the current and future. It also can dynamically adjust traffic conditions and issue real-time warning, and truly reflect the potential applications and value of the big data in intelligent traffic.

6. Acknowledgements

This study was supported by the Fundamental and Advanced Research Projects of Chongqing CSTC(Grant No.:cstc2015jcyjA30001),Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant No. KJ1500421,KJ120522) and Natural Science Foundation of Chongqing University of Posts and Telecommunications (Grant No.: A2014-97), the Doctoral Scientific Research Foundation of Chongqing University of Posts and Telecommunications (Grant No.: A2014-08)

References

[1]Kehrer, J., Hauser, H., 2013. Visualization and visual analysis of multifaceted scientific data: A survey. Visualization and Computer Graphics, IEEE Transactions on, 19(3), 495-513.

[2]Toker, D., Conati, C., Steichen, B., et al., 2013. Individual user characteristics and information visualization: connecting the dots through eye tracking. proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 295-304.

[3]Chen, X., Zheng, Z., Liu, X., et al., 2013. Personalized qos-aware web service recommendation and visualization. Services Computing, IEEE Transactions on, 6(1), 35-47.

[4]Ahn, J., Brusilovsky, P., 2013. Adaptive visualization for exploratory information retrieval. Information Processing & Management, 49(5), 1139-1164.

[5]Ben, X., Beijun, S., Weicheng, Y., 2013. Mining developer contribution in open source software using visualization techniques. Proceedings of the 2013 Third International Conference on Intelligent System Design and Engineering Applications. IEEE Computer Society, 934-937.

[6]Corchado, E., Herrero, Á., 2011. Neural visualization of network traffic data for intrusion detection. Applied Soft Computing, 11(2), 2042-2056.

[7]Adeshina, A. M., Hashim, R., Khalid, N. E. A., et al., 2013. Multimodal 3-D reconstruction of human anatomical structures using surlens visualization system. Interdisciplinary Sciences: Computational Life Sciences, 5(1), 23-36.

[8]Gama, S., Gonçalves, D., 2014.Studying color blending perception for data visualization.

[9]Xu, X., Hu, X., He, X., 2014. A Wargame Data Visualization Algorithm Based on Regular Radius and Constrained Random Direction. Proceedings of the 3rd International Conference on Multimedia Technology (ICMT 2013). Springer Berlin Heidelberg, 175-182.

[10]Lam, H., Bertini, E., Isenberg, P., et al., 2012. Empirical studies in information visualization: Seven scenarios. Visualization and Computer Graphics, IEEE Transactions on, 18(9): 1520-1536.

[11]Shahrestani, A., Feily, M., Masood, M., et al., 2012. Visualization of invariant bot behavior for effective botnet traffic detection. Telecommunication Technologies (ISTT), 2012 International Symposium on. IEEE, 325-330.

[12]Peck E M M, Yuksel B F, Ottley A, et al. 2013. Using fNIRS brain sensing to evaluate information visualization interfaces. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 473-482.

[13]Suzuki, H., Nakamura, S., Hirayanagi, M., et al., 2013. Gamma-ray visualization module. Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2013 IEEE. IEEE, 1-3.

[14]Bertini, E., Tatu, A., Keim, D., 2011. Quality metrics in high-dimensional data visualization: an overview and systematization. Visualization and Computer Graphics, IEEE Transactions on, 17(12), 2203-2212.

[15]Goyal, N., Leshed, G., Fussell, S. R., 2013. Effects of visualization and note-taking on sensemaking and analysis. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2721-2724.

[16]Morris, B.T.; Cuong Tran; Scora, G.;Trivedi, M.M. 2012. Real-Time Video-Based Traffic Measurement and Visualization System for Energy/Emissions. Intelligent Transportation Systems, IEEE Transactions on, 13(4), 1667-1678.

[17]Anwar,A.;Nagel,T.;Ratti,C., 2014. A Simple Visualization Technique to Support Traffic Incident Analysis, Pacific Visualization Symposium (Pacific Vis), Yokohama, IEEE, 316-319.

[18]Zhang, X., Li, N., Cao, C., et al., 2014. Traffic flow visualization based on line integral convolution. International Symposium on Optoelectronic Technology and Application 2014. International Society for Optics and Photonics, 930133-930133-6.

[19]O'Brien, W. J., Mandali, Y., Goyat, J., et al., 2015. An Integrated Visualization Technique for Transportation Management Planning in Highway Infrastructure Projects. Transportation Research Board 94th Annual Meeting, 15-3148.

[20]Song, Y., Miller, H. J., 2012. Exploring traffic flow databases using space-time plots and data cubes. Transportation, 39(2),215-234.

[21]Zahran, E. S. M., Smith, M. J., Bennett, L. D., 2012. 3D Visualization of Traffic-Induced Air Pollution Impacts of Urban Transport Schemes. Journal of Computing in Civil Engineering, 27(5), 452-465.

[22]Sewall J, Van Den Berg J, Lin M C, et al. 2011.Virtualized traffic: Reconstructing traffic flows from discrete spatiotemporal data. Visualization and Computer Graphics, IEEE Transactions on, 17(1), 26-37.

[23]Shao Chao, Huang Houkuan. 2006. A new data visualization algorithm based on SOM, Journal of Computer Research and Development, 43(3), 429-435.

[24]Ghodsi A, Huang J, Schuurmans D. 2010.Local tangent space embedding, Technical Report,CS-2003-32, School of Computer Science, University of Waterloo,Waterloo,On-tario,Canada.

[25]Quist M,Yona G. 2004. Distributional scaling: an algorithm for structure-preserving embedding of metric and nonmetric spaces, Journal of Machine Learning Research, 5,399-420.

[26]Ressom H,Wang D,Natarajan P. 2011. Adaptive double self-Organizing maps for clustering gene expression profiles, Neural Networks, 16(5/6), 633-640
[27]Lei Yingke, Xu Yangrning, Zhang Shanwen, et a1. 2010. Fast ISO-MAP based on minimum set coverage, Lecture Notes in Computer Science, 6216,173-179.
[28]Pan Yaozhang, Ge S S, Mamtm. 2009. A Weighted locally linear embedding for dimension reduction. Science, 42(5), 798-811.
[29]Lavergne P,Patile V. 2008. Breaking the curse of dimensionality in nonparametric testing, Journal of Econometrics, 143(1), 103-122.