ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 1, pp: (303-308), Month: January - March 2015, Available at: www.researchpublish.com

A Study on Visual Analytics

Shilpa M Patil[1], Dr. H S Guruprasad[2]

[1]PG Scholar, Dept. of CSE, BMSCE [Affiliated to VTU, Belgaum], Bangalore, India
[2]Professor and Head, Dept. of CSE, BMSCE [Affiliated to VTU, Belgaum], Bangalore, India

Abstract: **As we are living in a data-driven era with fast growing technologies, there exists a bulk of data generated, captured and also maintained for a variety of purposes. So there is a need of efficient, adequate tools to visualize the data and extract knowledge from them. This is possible with different visual analytic techniques. In this paper, we have presented a survey on different visual analytic techniques in various fields.**

Keywords: **Big Data, Visual Analytics, Visualization.**

I. INTRODUCTION

We are in a world which faces a fast rapidly increasing amount of data to be dealt with every day. Due to the very large amount of data, sometimes it is stored without filtering or refining. This increases the possibility of collecting and storing data at a faster rate than using it for making proper decisions. Always, we need to extract only knowledge from the raw data rather than complete raw data.

Visual analytics deals with capabilities of human visual information gathering and also abundant processing power of computers. Visual analytics are applied to the task of transforming data into valid information and finally to synthesize the knowledge. Many visual analytic systems combine both interactive visualization and also computational analysis. This helps for a complete understanding and decision making in very complex and huge data sets. Visual analytics contain various tools and techniques that help user to extract efficient knowledge from complex data and communicate effectively. Visual analytics is a multi-disciplinary area ranging from Bio Technology, Database research, Medicine, Business, Security and also Risk management.

II. LITERATURE SURVEY

Teng-Yok Lee et. al. [1] presents an integrated analysis and also visualization framework to clearly understand the Madden-Julian Oscillation [MJO] phenomenon from very large scale spatiotemporal climate simulation data. The paper discusses the tight integration of MJO domain knowledge, data analysis techniques such as feature tracking and also visualization methods such as virtual globe which can lead to powerful system for climate research. Gaither Kelly et. al. [7] describes how to employ a dataset of 4096 cells per time slice with 17 time steps, for a total 1 trillion cells to analyze Turbulent-flow simulations. The paper introduces a system for automatic feature detection and classification that is suited to infrequent data and also discuss about various ways to visualize data after analyzing the shapes and structures of entropy. The system identifies coherent paths and component birth, death and reconnection of cells. The paper implements the feature analysis on top of visualization, which is a popular large scale system that many use. Kendall Wesley et. al. [8] presents a visual analytics system for exploring feature characteristics in large scale flow data sets. The system uses their back end processing pipeline support and effective feature queries. One among them is OSUFLOW, flow tracing library that can generate millions of field lines in parallel using supercomputers. Other component is scalable query interface, which supports parallel querying. The paper demonstrates the systems utility using data generated from parallel ocean program and also provides detailed analysis of system performance. Gorg C et.al. [12] investigated how visual analysis supports the intelligence analysis which requires people and some organization to review and access the large set of information to understand situations. Jigsaw was designed to help investigators explore and understand collections of text documents and to follow trails of ideas embedded across the documents. The authors also formed few objectives for jigsaw and other visual analytics. The insights for visual analytics projects and hypothetical scenarios show the capacity of visual analytics to aid analysts who explore and evaluate volumes of data. Johanna Schmidt et. al. [16] worked towards new approach for comparative visualization for sets of images. In many tasks, there is a need to compare two sets of datasets and hence there is need of comparative visualization tool to find out the difference and also similarities among many datasets. The main challenge is in scaling and also parameters like color. The paper introduced new approaches to know differences and similarities which mainly preserve contextual information by allowing detailed analysis. The paper proves the flexibility of their approach by applying it to multiple different domains. Heer J et. al. [19] presented design considerations for asynchronous collaboration in visual analysis environments by highlighting issues of work parallelization, communication and also social organization. The paper attempted to identify methods which facilitates collaboration and also suggests mechanisms for achieving those. The focus is on the list of collaborative visualization tasks along with techniques to improve shared contexts and awareness and also to consider parallelized work that facilitate mutual understanding and also to reduce cost of various collaborative tasks. Stark Robert et. al. [27] attempted a design to convey geo spatial and more detailed information without cognitive overload and also to allow the user to quickly delve into details of anomalies, using an in house application frame work. To meet this, user needs to create system which interact “Big pictures” data visualization with detailed view, using clustering algorithm to promote geo spatial analysis, a unique display to charts to convey both health and anomalies. This software put in place the backbone of a powerful and multipurpose platform that intuitively ties together the visualization of geo spatial and quantitative data sets which allows user to easily analyze the overall status of network.

Jaegul Choo et. al. [2] discusses the importance of convergence and precision where these two aspects have not received the required consideration in visual analysis. The paper proposes various customizing computational methods which include low precision computational and also iteration level visualization to ensure real time visual analytics for big data. Khairi Reda et. al. [3] discusses how a new kind of visualization space called hybrid reality environments can achieve scalable visualization of various heterogeneous datasets. These environments synergize the capabilities of VR and high resolution tiled LCD walls, letting users unite 2D and 3D information spaces. The paper introduces two environments cyber commons and CAVE 2 along with some real world application. Fisher Danyel et. al. [9] describes on interactive exploration of very large datasets. The paper explores how user interacts with changing incremental visualization and reveals challenges unique to this visualization mode. The paper describes guide lines for visualizing changing, approximate results and apply these guidelines by adopting past visualization technique to convey uncertainty information. Interactive exploration continues to be a top challenge in extreme scale virtual analytics. Fekete J D et. al. [14] addresses various different nature of exploration. Visual analytic applications should rely on standard data management streams instead of building their own. And even exploration is the construction of a workflow as series of operations that filters, summarizes and analyze the data. Eventually visual analytics could become easier to implement and more widespread. Koch steffen et. al. [20] tried how relatively young field of visual analytics aims to combine automatic procedures with algorithms which are used to retrieve information along with various visual techniques. The paper presents an interactive method for dealing with exhaustive search and analyzing problems which indirectly facilitate the steering of machine learning technique. The paper also describes how a binary classifier can be created quickly by employing interactive visualization by providing detailed usage example.

Chinchor N A et.al. [10] put forward a step to address combining multimedia analysis and visual analytics to deal with information from different sources having different set of goals or even objectives which contains various media types and their combinations. Hence, multimedia analytics is a critical need for a broad range of applications which includes medicine, economics, social media and security. The paper also focuses on history and also interfaces that leverage metadata to deliver efficient, effective retrieval from multimedia. Tobias Schreck et. al. [11] tried various ways which are quite interesting to visually analyze different types of various social media data that includes community provided photo collections, streaming news data and also geo-referenced micro blog data. Visual representations can help to provide an initial over view of the data which enables the analysts to identify and navigate aspects of interest. Researchers have prototypes for numerous systems that extract the multifaceted information from social media streams and hence combine this information with other data and results in interactive visual displays. The paper also presents results which are mainly relied on image metadata. Jaiwan Zhang et. al. [17] proposed new visual analytics approach for analyzing and also visualizing the elegant wall painting degradation. The paper came up with multi scale visual analytics framework along with many hypothesis testing tools to develop assistance of work such as radial layout plus bar charts tool for overall visualization, aggregated visualization tool with line charts combined with radar maps to know overall degradation in different scales. Several case studies have also been carried out using real world painting data collected from world Heritage site as to verify the effectiveness of proposed method. Scharl A et. al. [18] presented the media watch on climate change which is a public web portal that captures and also aggregates the large achieves of digital contents from various multiple stakeholders groups. Every week it accesses the relevance of millions of documents and user comments from new media, blogs, web 2.0 platforms such as face book, twitter and even you tube along with other range of sources. The paper introduced interactive dash board with trend charts and complex map projections that not only shows how often and where it is published but also gives real time account of concepts that stake holders associate with climate change. Both sides are considered as to help gain a better understanding of other climate related issues. Horn Chris et. al. [30] attempts to develop a visual analytics tool which can be used to identify several areas of concern within provided data. This was presented by videos that depict the visual analytic interaction sequence of two to explore data sets using the tool. This paper mainly presents the design rationale behind the secure decisions team entry, aspects of tool and some technical details behind the process.

Johanna Beyer et. al. [4] describes a system for interactive exploration of Petascale volume data of neural tissues generated by high throughput electron microscopy imaging. This visualization driven systems let users handle multiple volumes of incomplete data, restricts most computations to a small subset of data, and achieves scalable computing with multi resolution virtual memory. Wood J et. al. [21] reported design decisions along with software to know about disease and its mutation by using visual analytics software. The study focuses on how the fictitious disease may have spread between various cities around the globe and how the genetic sequence taken from infected patients may be used to chart the evolution of the disease and its changes along with drug resistance and other characteristics. The paper uses lower level programming environment processing, a graphically oriented set of java libraries. Processing is engineered to ease the fast prototyping which is suitable for iterative tool development data analysis tasks. Carpendale S et. al. [31] focused on biological data visualization and the viewpoints presented are relevant to visualization and visual analytics in general. Ontology is essentially graphing based knowledge represented in which each node corresponds to concept and each edge specifies a relation between two concepts. Visualization plays a significant role in biology which has embraced ontologies purpose fully and energetically. Here, they have examined the potential impact of ontologies, increasing presence in research. The two goals of this work are: first, support ontology developers, data creators and data analysts with techniques. Second, stimulate new research programs for developing knowledge - assisted visualization by making use of ontologies as knowledge representations.

Rysavy S J et.al. [5] developed DIVE [Data Intensive Visualization Engine] as to fill the technological gap and also makes big data visual analytical approaches accessible. DIVE has some of the components like data pipeline that is extensible and also adoptable which transforms datasets at various speeds. DIVE is visual analytic software framework which is expressed ontologically and also sufficient to analyze the large datasets at various speeds. Here, authors present bioinformatics applications to prove that DIVE can handle data from many domains. Shaverdian Anna et. al. [6] discusses VA graph framework that will facilitate the design of techniques for scale graph analysis. Their algebraic based framework aggregates and selects attributes and also information from very large graphs. The paper focuses on the graph algebra's theoretical foundations and also present working implementation over cytoscope which is a popular biological network exploration tool. The implementation demonstrates the algebraic framework and scalable aggregation by making use of large social network datasets. Kwan-Liu Ma et. al. [13] presented survey of graph layout and simple methods that reveal considerable progress in various new directions. Graph visualization and analytics use advanced measurements that leads to rich interactive network view and also node grouping and filtering. Graph analysis is a key for research innovations to address the escalating and complexities of data. The paper discusses new methods that must address network representation, from the fundamental problems of laying out a large graph analytics and also simplifications for dynamic graphs. Jun Ma et. al. [15] introduced Flow Graph which is a novel compound graph representation which hierarchically arranges field line clusters and also spatiotemporal regions for occlusion-free and along with controllable visual exploration. The paper showed that Flow graph not only provides a visual mapping which abstracts field line clusters and region details but also serves as guiding tool for exploration and its understanding. Making use of brushing as well as linking in conjunction with standard field line view they demonstrated effectiveness of Flow graph. An expert evaluation is performed to confirm the usefulness of this technique. Shi Lei et. al. [26] introduces visualization along with data processing and visual analytics process. Visualization a graph with more than a hundred nodes faces two challenges. First, classical force directed methods in most cases fail to calculate an optimally graph layout in real time. Second, even though layout is computed it does not allow user from understanding graph in detail which is necessary. Hence in the proposed method, they used a combination of anomaly lists detected in the data pre-processing part to the graph links.

Campbell C et. al. [22] in response to a challenge, developed an analytical pipeline which combines manual analysis with automated entity extraction and visualization tool to gain insight into source document. Visualization is also key component of this challenge as tools were used throughout the analysis starting from hypothesis to illustrating conclusions. This approach makes the tool agnostic and also other few tools could be added or subtracted based on the future analysis, on data types and subject areas. Pascale Proulx et. al. [23] focuses on the analytics using the n Space/ Geo time tools which helps to investigate visual analytic application or that deal with text records. Geo time and n Space are two innovative visual analytical applications which support investigation of large and complex multi-dimensional data sets. These tools provided an open and integrated analytical environment that acted as visual extension of their cognitive capabilities. Barcelos Yussif et. al. [24] briefly describes the tool which was monitored to know the health status of network of fictitious bank and mainly focusing on machines policy status and activity flags. This tool provides traditional and high density visualizations to evidence trends, patterns and also used to spot various exceptions and casual relationship. Visual analysis is the science of analytical reasoning facilitated by interactive visual interfaces. The challenge was to create own tools to facilitate the management of network of fictitious bank of money. The proposed tool was developed which help users to explore and have insights about network data. Cao Yong et. al. [25] introduced GPU - Accelerated visual analysis tool. The main design objective of the tool is to identify temporal patterns from huge and complex data. In order to meet the objectives, the paper introduced three unique features automatic animation, disjunctive data filters and also time synced visualization by multiple data sets. The tool supports various correlated data views such as histogram view, parallel coordinated view and also dynamic view. This AVIST tool is used to identify three noteworthy events which include data sets of firewall log and Intrusion detection system. Dudas L et. al. [28] proposed a visual analytics tool that allows the user to browse and drill down the various multi-dimensional data on line with the possibility to export results into zooming representation frame work. And this is addressed by aiding the cognitively hard tasks of understanding attributes. The objective is to handle big size of data efficiently. Since OWLAP has server client architecture, heavy computation can be performed while the geographical clients run on simple personal computers. So server generates OLAP cubes using parallel threads, dividing into smaller parts and each handles its own fragment. Victor Yingjie Chen et. al. [29] presented visual analytics system semantic Prism, which aims to analyze large scale high dimensional data sets containing logs of million computers and this system visualization the data from different perspective: geo temporal, time series curve and also pixel visualization. Mainly system is developed using adobe flash, PHP and MySQL. It is a web application that analyst can run using most modern web browsers and along with this the client server structure is naturally suitable for such a problem by keeping large scale data in central location. On the whole, this analyst explores the data from different aspects.

III. CONCLUSION

As the relevance and importance of extracting knowledge from raw data is continuously increasing, the importance of Visualization and Visual Analytics is increasing parallely. This paper discusses the various techniques of visual analytics for different applications which is the cutting edge of research today.

ACKNOWLEDGEMENT

The work reported in this paper is supported by the college [BMSCE, Bangalore] through the TECHNICAL EDUCATION QUALITY IMPROVEMENT PROGRAMME [TEQIP-II] of the MHRD, Government of India.

REFERENCES

[1] Teng-Yok Lee, Xin Tong, Han-Wei Shen, Pak Chung Wong, Hagos S, Leung, L.R, "Feature Tracking and visualization of the Madden-Julian oscillation in climate simulation", IEEE Journal on Computer Graphics and Applications, Volume 33, Issue 4, April 2013, pp 29-37, DOI: 10.1109/MCG.2013.36.

[2] Jaegul Choo, Haesun Park, "Customizing Computational Methods for Visual Analytics with Big Data", IEEE Journal on Computer Graphics and Applications, Volume 33, Issue 4, April 2013, pp 22-28, DOI: 10.1109/MCG.2013.39

[3] Reda, K, Febretti A, Knoll A, Aurisano J, Leigh J, Johnson A, Papka M.E, Hereld M, "Visualizing Large, Heterogeneous Data in Hybrid-Reality Environments", IEEE Journal on Computer Graphics and Applications, Volume 33, Issue 4, April 2013, pp 38-48, DOI: 10.1109/MCG.2013.37.

[4] Beyer J, Hadwiger M, Al-Awami A, Won-Ki Jeong, Kasthuri N, Lichtman J.W, Pfister H, "Exploring the Connectome: Petascale Volume Visualization of Microscopy Data Streams", IEEE Journal on Computer Graphics and Applications, Volume 33, Issue 4, July 2013, pp 50-61, DOI: 10.1109/MCG.2013.55.

[5] Rysavy S.J, Bromley, D, Daggett V, "DIVE : A Graph Based visual analytics Frame work for Big data", IEEE Journal on Computer Graphics and Applications, Volume 34, Issue 2, Mar 2014, pp 26-37, DOI: 10.1109/MCG.2014.27.

[6] Shaverdian Anna A, Zhou Hao, Michailidis George, Jagadish Hosagrahar V, "A Graph Algebra for scalable visual analytics", IEEE Journal on Computer Graphics and Applications, Volume 34, Issue 4, May 2012, pp 26-33, DOI: 10.1109/MCG.2012.62.

[7] Gaither Kelly P, Childs Hank, Schulz Karl W, Harrison Cyrus, Barth William, Donzis Diego, Yeung Pui-Kuen, "Visual Analytics for finding critical Structures in Massive time varying Turbulent Flow simulations", IEEE Journal on Computer Graphics and Applications, Volume 32, Issue 4, May 2012, pp 34-45, DOI: 10.1109/MCG.2012.63.

[8] Kendall Wesley, Huang Jian, Peterka Tom, "Geometric Quantification of Features in Large Flow Fields", IEEE Journal on Computer Graphics and Applications, Volume 32, Issue 4, April 2012, pp 46-54, DOI: 10.1109/MCG.2012.49.

[9] Fisher Danyel, Drucker Steven M, Konig A Christian, "Exploratory Visualization Involving Incremental, Approximate Database Queries and Uncertainty", IEEE Journal on Computer Graphics and Applications, Volume 32, Issue 4, April 2012, pp 55-62, DOI: 10.1109/MCG.2012.48.

[10] Chinchor N.A, Thomas, J.J, Wong P.C, Christel M.G, Ribarsky W, "Multimedia Analysis + Visual Analytics = Multimedia Analytics", IEEE Journal on Computer Graphics and Applications, Volume 30, Issue 5, Aug 2010, pp 52-60, DOI: 10.1109/MCG.2010.92.

[11] Tobias Schreck, Daniel Keim, "Visual Analysis of Social Media Data", IEEE Computer Society, Computer, Volume No. 46, Issue 5, May 2013, pp 68-75, DOI: 10.1109/MC.2012.430.

[12] Gorg C, Youn-ah Kang, Zhicheng Liu, Stasko J, "Visual Analytics Support for Intelligence Analysis", Computer, Volume 46, Issue 7, March 2013, pp 30-38, DOI: 10.1109/MC.2013.76.

[13] Kwan-Liu Ma, Muelder, C.W, "Large-Scale Graph Visualization and Analytics", Computer, Volume No. 46, Issue 7, pp 39 - 46, DOI: 10.1109/MC.2013.242.

[14] Fekete J D, "Visual Analytics Infrastructures: From Data Management to Exploration", Computer, Volume 46, Issue 7, March 2013, pp 22-29, DOI: 10.1109/MC.2013.120.

[15] Jun Ma, Chaoli Wang, Ching-Kuang Shene, Jingfeng Jiang, "A Graph-Based Interface for VisualAnalytics of 3D Streamlines and Pathlines", IEEE transactions on visualization and computer graphics, Volume 20, Issue 8, Sept 2013, pp 1127-1140, DOI: 10.1109/TVCG.2013.236.

[16] Johanna Schmidt, M. Eduard Groller, Stefan Bruckner, "VAICo: Visual Analysis for Image Comparison", IEEE Transactions on Visualization and Computer Graphics, Volume 19, Issue 12, December 2013, pp 2090-2099, DOI: 10.1109/TVCG.2013.213.

[17] Jiawan Zhang, Kai Kang, Dajian Liu, Ye Yuan, Yanli E, "Vis4Heritage: Visual Analytics Approach on Grotto Wall Painting Degradations", IEEE Transactions on Visualization and Computer Graphics, Volume 19, Issue 12, 2013, pp 1982-1991.

[18] Scharl A, Hubmann-Haidvogel A, Weichselbraun A, Lang H P, Sabou M, "Media Watch on Climate Change - Visual Analytics for Aggregating and Managing Environmental Knowledge from Online Sources", 46th

International Conference on System Sciences (HICSS), 2013, Wailea, Maui, Hawaii, 07-10 Jan 2013, pp 955-964, DOI: 10.1109/HICSS.2013.398.

[19] Heer J, Agrawala M, "Design Considerations for Collaborative Visual Analytics", IEEE Symposium on Visual Analytics Science and Technology, Sacramento, CA, Oct. 30 2007-Nov. 1 2007, pp 171-178, DOI: 10.1109/VAST.2007.4389011.

[20] Koch Steffen, Heimerl Florian, Ertl Thomas, "Visual Document Retrieval: Supporting Text Search and Analysis with Visual Analytics" Computing in Science & Engineering: Volume 15, No. 4, 2013, pp 66-74.

[21] Wood J, Slingsby A, Dykes J, "Designing visual analytics systems for disease spread and evolution", IEEE Symposium on Visual Analytics Science and Technology, VAST 2010, 25-26 Oct 2010 , Salt Lake City, Utah, US, DOI: http://dx.doi.org/10.1109/VAST.2010.5652689.

[22] Campbell C, Blanchard S, Chin S, Henderson C, Holland M, Jennings K, Kuehl P, Lucey D, McCoy M, McCracken J, Pecheux B, Pietrzak M, Roberts D, Sanders M, Taylor K, Whitford D, DiMassimo R P, "Multi-Viz data fusion", IEEE Symposium on Visual Analytics Science and Technology (VAST), Salt Lake City, UT, 25-26 Oct. 2010, pp 265-266, DOI: 10.1109/VAST.2010.5653004.

[23] Pascale Proulx, Adeel Khamisa, Robert Harper, "Integrated visual analytics workflow with GeoTime and nSpace VAST 2010 mini challenge 1 award: Outstanding Analysis and Accuracy", IEEE Conference on Visual Analytics Science and Technology, IEEE VAST 2010, Salt Lake City, Utah, USA, 24-29 October 2010, pp 273-274.

[24] Barcelos Yussif, Aburjaile Flavia, Leite Laura R, Oliveira Solange T, de Melo-MinarcTi Raquel C, "Combining traditional and high-density visualizations in a dashboard to network health monitoring", IEEE Conference on Visual Analytics Science and Technology (VAST), Seattle, WA, USA, 14-19 Oct. 2012, pp 295-296, DOI: 10.1109/VAST.2012.6400509.

[25] Cao Yong, Moore Reese, Mi Peng, Endert Alex, North Chris, Marchany Randy, "Dynamic analysis of large datasets with animated and correlated views", IEEE Conference on Visual Analytics Science and Technology (VAST), Seattle, WA, 14-19 Oct. 2012, pp 283-284, DOI: 10.1109/VAST.2012.6400515.

[26] Shi Lei, Liao Qi, Yang Chunxin, "Investigating network traffic through compressed graph visualization", IEEE Conference on Visual Analytics Science and Technology (VAST), Seattle, WA, 14-19 Oct. 2012, pp 279-280, DOI: 10.1109/VAST.2012.6400517.

[27] Stark Robert F, Wollocko Arthur, Borys Michael, Kierstead Megan, Farry Michael, "Visualizing large scale patterns and anomalies in geospatial data", IEEE Conference on Visual Analytics Science and Technology (VAST), Seattle, WA, 14-19 Oct. 2012, pp 271-272, DOI: 10.1109/VAST.2012.6400521.

[28] Dudas L, Fekete Zs, Gobolos-Szabo J, Radnai A, Salanki A, Szabo A, Szucs G, "OWLAP - using OLAP approach in anomaly detection", ", IEEE Conference on Visual Analytics Science and Technology (VAST), Seattle, WA, 14-19 Oct. 2012, pp 267-268, DOI: 10.1109/VAST.2012.6400523.

[29] Victor Yingjie Chen, Ahmad M Razip, Sungahn Ko, Cheryl Zhenyu Qian, David S Ebert, "Semantic Prism: a multi Aspect view of large high dimensional data", IEEE Conference on Visual Analytics Science and Technology (VAST), Seattle, WA, 14-19 Oct. 2012, pp 159-160, DOI: 10.13140/2.1.1849.9204.

[30] Horn Chris, Ellsworth Chris, "Visual analytics for situation awareness of a large-scale network", IEEE Conference on Visual Analytics Science and Technology (VAST), Seattle, WA, 14-19 Oct. 2012, pp 263-264, DOI: 10.1109/VAST.2012.6400525.

[31] Carpendale S, Min Chen, Evanko D, Gehlenborg N, Gorg C, Hunter L, Rowland F, Storey M A, Strobelt H, "Ontologies in Biological Data Visualization", IEEE Journal on Computer Graphics and Applications, Volume 34, Issue 2, March 2014, pp 8-15, DOI: 10.1109/MCG.2014.33.