Final Report for Period: 09/2008 - 08/2009   **Submitted on:** 11/29/2009

Principal Investigator: Ward, Matthew O.   **Award ID:** 0414380

Organization: Worcester Polytech Inst

Submitted By:

Ward, Matthew - Principal Investigator

Title:

Quality-Aware Visual Exploration Tools

Project Participants

Senior Personnel

> **Name:** Ward, Matthew
>
> **Worked for more than 160 Hours:** Yes
>
> **Contribution to Project:**
>
> Dr. Ward is primarily concerned with the visualization and user interface components of this project, including their design and evaluation.

> **Name:** Rundensteiner, Elke
>
> **Worked for more than 160 Hours:** Yes
>
> **Contribution to Project:**
>
> Dr. Rundensteiner is primarily involved with data management, analysis, and access issues in this project.

Post-doc

Graduate Student

> **Name:** Xie, Zaixian
>
> **Worked for more than 160 Hours:** Yes
>
> **Contribution to Project:**
>
> Zaixian worked on seamlessly incorporating quality of data (at record, tuple, and column granularity) into the four existing display types.

> **Name:** Cui, Qingguang
>
> **Worked for more than 160 Hours:** No
>
> **Contribution to Project:**
>
> Qingguang studied abstraction quality, focusing in particular on sampling and clustering as the two forms of abstraction.

> **Name:** Wad, Charudatta
>
> **Worked for more than 160 Hours:** Yes
>
> **Contribution to Project:**
>
> Charudatta investigates quality of clustering, and trade-off between quality of resolution and performance.

> **Name:** Yang, Di
>
> **Worked for more than 160 Hours:** Yes
>
> **Contribution to Project:**
>
> Di studied how we may exploit knowledge of prior data exploration in XMDV to facilitate future exploration.

> **Name:** Nguyen, Do Quyen
>
> **Worked for more than 160 Hours:** No
>
> **Contribution to Project:**
>
> Nguyen has focused on the measurement of visual clutter and the development of dimension reordering techniques to reduce visual clutter.

Undergraduate Student

Technician, Programmer

Other Participant

Research Experience for Undergraduates

    Name: Sader, Paul

    Worked for more than 160 Hours: Yes

    Contribution to Project:

Paul worked on optimization of the core XMDVTool system, both at the data access as well as the display feature level.

        Years of schooling completed: Junior

        Home Institution: Same as Research Site

        Home Institution if Other:

        Home Institution Highest Degree Granted(in fields supported by NSF): Doctoral Degree

        Fiscal year(s) REU Participant supported: 2007

        REU Funding: REU supplement

    Name: Chaplin, Jeremiah

    Worked for more than 160 Hours: Yes

    Contribution to Project:

Jeremiah worked on developing benchmark suites for performance evaluation and optimization of the system.

        Years of schooling completed: Junior

        Home Institution: Same as Research Site

        Home Institution if Other:

        Home Institution Highest Degree Granted(in fields supported by NSF): Doctoral Degree

        Fiscal year(s) REU Participant supported: 2007

        REU Funding: REU supplement

    Name: Wilson, Andrew

    Worked for more than 160 Hours: Yes

    Contribution to Project:

Andrew worked on enhancing the architecture of XmdvTool to facilitate porting between different platforms.

        Years of schooling completed: Junior

        Home Institution: Same as Research Site

        Home Institution if Other:

        Home Institution Highest Degree Granted(in fields supported by NSF): Doctoral Degree

        Fiscal year(s) REU Participant supported: 2007 2006

        REU Funding: REU supplement

    Name: Bryan, Christian

    Worked for more than 160 Hours: Yes

    Contribution to Project:

Christian has helped with the development of interaction tools that work between data and quality spaces.

        Years of schooling completed: Junior

        Home Institution: Same as Research Site

        Home Institution if Other:

        Home Institution Highest Degree Granted(in fields supported by NSF): Doctoral Degree

        Fiscal year(s) REU Participant supported: 2007 2006

        REU Funding: REU supplement

Organizational Partners

Other Collaborators or Contacts

Professor Jing Yang from UNCC is collaborating with our XMDV team on several aspects of the project, most notably, applications and extensions of the engine to other application domains such as scalable image exploration.

Dr. Bob Mann, University of Edinburgh, and his students are also working to integrate our XMDV-lite Java version into their exploration code base for their Virtual Observatory project, a multi-national research initiative between the U.K., the U.S., and Australia.

Activities and Findings

Research and Education Activities:

This project involved several key research and education activities :

1. Development of a holistic approach to incorporating the quality of data (at three levels of granularity) explicitly into the visualization to enable users to integrate knowledge of the quality of or confidence in the data into the decisions made during analysis.

2. Development of customized visualizations to convey the quality attributes of the data, with provisions for interactively selecting data based on its quality and linking these selections with the data visualizations.

3. Development of measures to compute the quality of data abstractions, in particular, sampling and clustering, to give the user an understanding of the loss of information they may experience in the abstracted and thus reduced data space.

4. Development of abstraction quality visualizations that differentiate the quality of selected subsets of the data and the unselected data. This allows users to maintain views that convey more information in the interesting portions of the data while reducing the volume of less interesting data in the display.

5. Development of interactive tools for controlling the amount of data being retrieved and visualized based on the needs of the user, trading-off between the computational performance needed to generate the abstraction versus the accuracy of the data in representing the entire dataset.

6. Design of new metrics for measuring clutter in the display space -- even if the data itself is perfect, and algorithms that target the reduction of such visual cluster in order to improve display quality. Measures specific to four distinct multivariate visualization techniques were developed and evaluated.

7. Development of a suite of benchmarks to assess the computational performance of our tool, along with a number of algorithm and data structure enhancements that resulted in a significant improvement in the speed of loading and rendering data.

8. Development of tools to interactively improve the quality of data and dimension clustering, making use of the analyst's knowledge of the data to create clusters that are more meaningful.

9. Introducing 4 undergraduate students to the research process via REU supplements. Two of the students were involved in performance benchmarking and analysis activities, while the other two focused on refinements to our software infrastructure, in particular the linkages between selection operations in different views.

Our solution techniques designed to tackle the above tasks have all been implemented and integrated into the XmdvTool suite. Further, we have conducted user studies with each of these features to assess their utility, as well as uncover possible limitations for practical exploration on real data sets.

Findings:

Our evaluations have resulted in a number of interesting observations, a few of which are highlighted below:

1. We found that displays that were reorganized (in our case via dimension reordering) and resulted in a lower clutter measure according to our metric were generally perceived by users to be easier to interpret compared to the original displays.

2. We also found that incorporation of data quality as additional attributes into standard display techniques via other visual attributes, such as thickness of lines, color, etc., can be achieved in ways that do not significantly impede the data analysis. This offers the advantage that our proposed strategy may be easily incorporated into other existing visualization tool suites.

3. Data is often abstracted to make it more manageable to users. However, it's typically impossible for the user to assess the relationship of this abstracted data with the original. Our measures of data abstraction quality are the first ones in the field aiming to tackle this issue -- that is, they give the user a handle on how representative of the actual base data the compact abstracted data may be.

4. For data views involving a large number of records, superimposing data quality attributes resulted in congested displays that were often difficult to interpret. As an alternative, we created a separate display of 'quality space', with interactive linkage to the various data visualizations. Thus users could make selections either via the data values or their underlying quality attributes. This proved to be a much more scalable solution.

Training and Development:

The project team is composed of students at all levels, from B.S. to M.S, to Ph.D.. These students work together closely in both solving their research challenges as well as developing a shared code base. As a consequence, this gives the students the opportunity to acquire many new skills, including:

1. communication: presenting their work to our group in regular group meetings and at conferences.
2. technical writing: helping to write technical reports and papers for submission to conferences and journals.
3. problem solving: addressing both research and coding issues on a daily basis.
4. research skills: identifying and formulating problems, and understanding the literature to be able to extract techniques and enhance them as needed so as to devise state-of-the-art techniques for this quality-aware data exploration project.
5. software engineering skills: designing, implementing, debugging, maintaining and releasing our shared code base XmdvTool, which has been under continuous development for over 13 years.
6. teamwork: working together with members with different skills sets and ambitions on shared goals.

Outreach Activities:

We make our research ideas available to the general public by regularly producing an easy-to-use release of our tool suite for the analysis of data sets. We include both source code and executables in these releases, as well as many test datasets.

We have also invested time into developing a variety of import tools from common data formats into our tool, including excel, hdf, netcd, and so on, to make it more convenient for users to bring their data into our tool.

In addition, we have begun targeting specific application domains where we can demonstrate the software and solicit input on making the tool more useful across different user communities. In particular, we interacted with researchers, educators, and clinical staff at the University of Massachusetts Medical School to educate them on the power of interactive visual data exploration. We hope that such outreach activities will yield not only new research ideas but also help us find a large body of real users who would be able to assist in the assessment of our tool.

Journal Publications

Qingguang Cui, Matthew O. Ward, Elke A. Rundensteiner, "Enhancing Scatterplot Matrices for Data with Ordering or Spatial Attributes", Visualization and Data Analysis, Part of IS&T/SPIE Symposium on Electronic Imaging, p. 60600R, vol. , (2006). Published,

Shiping Huang, "Exploratory Visualization of Data with Variable Quality", MS Thesis, Worcester Polytechnic Institute Computer Science Department, p. 1, vol. , (2005). Published,

Zaixian Xie, Shiping Huang, Matthew O. Ward, Elke A. Rundensteiner, "Exploratory Visualization of Multivariate Data with Variable Quality", Proc. IEEE Symposium on Visual Analytics Science and Technology, p. , vol. , (2006). Accepted,

Qingguang Cui, Matthew Ward, and Elke A. Rundensteiner, "Measuring Data Abstraction Quality in Multiresolution Visualization", IEEE Transactions on Visualization and Computer Graphics, p. 709, vol. 12, (2006). Published,

Wei Peng, Matthew O. Ward and Elke A. Rundensteiner, "Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering", IEEE Symposium on Information Visualization (InfoVis 2004), p. 89, vol. , (2004). Published,

Jing Yang, "A General Framework for Multi-Resolution Visualization", dissertation, WPI, April 2005, etd-050505-113345, p. 1, vol. , (2005). Published,

Wei Peng, "Clutter-Based Dimension Reordering in Multi-Dimensional Data Visualization", M.S. Thesis, Worcester Polytechnic Institute Computer Science Department, p. 1, vol. , (2005). Published,

Zaixian Xie, Matthew Ward, Elke Rundensteiner, and Shiping Huang, "Integrating Data and Quality Space Interactions in Exploratory Visualizations", Proc. International Conference on Coordinated and Multiple Views in Exploratory Visualization, p. , vol. , (2007). Accepted,

Qingguang Cui, "Measuring Data Abstraction Quality in Multiresolution Visualizations", M.S. Thesis, Worcester Polytechnic Institute, Computer Science Department, p. 1, vol. , (2007). Published,

Elke Rundensteiner, Matthew Ward, Zaixian Xie, Qingguang Cui, Charudatta Wad, Di Yang, and Shiping Huang, "XmdvTool^Q: Quality-aware interactive data exploration", Proc. SIGMOD 2007 (Demo paper), p. 1, vol. , (2007). Published,

Wad, C, Rundensteiner, E., and Ward, M., "QoS: Quality-driven data abstraction generation for large databases", Proc. 5th Int. Workshop on Quality in Databases (QDB), p. , vol. , (2007). Published,

Yang, D., Rundensteiner, E., Ward, M., "Nugget Discovery in Visual Exploration Environments by Query Consolidation", ACM CIKM 2007, p. 603, vol. , (2007). Published,

Yang, Xie, Z., D., Rundensteiner, E., Ward, M., "Managing Discoveries in The Visual Analytics Process", ACM SIGKDD Explorations, p. , vol. , (2008). Accepted,

Yang, D., Rundensteiner, E., Ward, M., "Analysis Guided Visual Exploration to Multivariate Data", IEEE Symposium on Visual Analytics Science and Technology, p. 83, vol. , (2007). Published,

Ward, M., Xie, Z., Yang, D., Rundensteiner, E., "Quality-Aware Visual Data Analysis", Computational Statistics, p. , vol. , (2009). Accepted,

Books or Other One-time Publications

Wei Peng, "Clutter-Based Dimension Reordering in Multi-Dimensional Data Visualization", (2005). Thesis, Published
Bibliography: M.S. Thesis, Computer Science Department, Worcester Polytechnic Institute

Qingguang Cui, "Measuring Data Abstraction Quality in Multiresolution Visualizations", (2006). Thesis, Published
Bibliography: M.S. Thesis, Computer Science Department, Worcester Polytechnic Institute

Di Yang, "Analysis Guided Visual Exploration of Multivariate Data", (2007). Thesis, Published
Bibliography: M.S. Thesis. Computer Science Department, Worcester Polytechnic Institute

Web/Internet Site

URL(s):
http://davis.wpi.edu/~xmdv
Description:
Site contains source code, executables, data sets, instructional material, publications, and case studies.

Other Specific Products

Contributions

Contributions within Discipline:
Our project has resulted in several key contributions to the visual analytics field:

1. new metrics for measuring clutter in the display space, and algorithms for several visualization mappings that target the reduction of such visual cluster in order to improve display quality.

2. a holistic approach at incorporating the quality of data
(at three levels of granularity) explicitly into the visualization to enable users to conduct quality-aware exploration.

3. a new visualization that conveys only the quality attributes of data, with interactive linkages to the corresponding data visualizations.

4. methods for measuring and displaying the quality of data abstraction, in particular, sampling and clustering, to give the user an understanding of the loss of information they may experience in the abstracted and thus compressed data space.

5. interactive tools to enable users to control the trade-offs between abstraction quality and data retrieval time, so analysts know when they have sufficient information to make decisions.

6. interactive tools to refine clustering of data and dimensions to improve the quality of the abstraction using the analyst's domain knowledge and visual perception.

Each of these contributions have been the focus of evaluation via both user studies and computational performance measurement. These evaluations resulted in several enhancements to our algorithms, visualizations, and interaction techniques.

These results have the potential to influence all development of future visualization tools by providing means to evaluate visualizations and the processes that generate them in terms of the quality of the data, transformations on the data, and the visual mapping being used, and thus the confidence that can be associated with any conclusions drawn from the visualization.

Contributions to Other Disciplines:
Our findings are effectively contributing to any discipline that has the requirement for interactive exploration and mining of complex data repositories, including biology, chemistry, earth and space science, homeland security, and so on. While much work has been done on developing novel methods for displaying data and information, little has been done on evaluating and optimizing the quality of the resulting visualization in terms of information content and interpretability. Similar criticisms can be directed at both the user interaction and data management components of the visualization process. We have taken the first steps towards addressing the shortcomings of existing tools, and thus better supporting analysts in their exploration and decision-making. That is, adding quality-awareness explicitly into the exploration tools will give power into the hands of data explorers to better understand the data they are dealing with and draw conclusions in a more accurate, confident, and timely fashion.

Contributions to Human Resource Development:
This project has involved students at all levels, including B.S. students in the form of REUs, Master's students, and Ph.D. students. One student on the XMDV project, Jing Yang, is currently a tenure-track faculty member at UNCC -- thus contributing in turn to the education and the competitiveness of our workforce. Current and past supported students on the project include multiple females as well as a Hispanic student, thus contributing to the diversity of students in science and engineering.

Contributions to Resources for Research and Education:
We release our Xmdvtool suite (including source code) as freeware on a regular basis, which provides a valuable resource to others around the world both in support of education (for courses at several universities) and in support of research (for a variety of application domains using this tool for their data analysis as well as a foundation for other research in visualization).

A new architecture for Xmdvtool, based loosely on the visualization reference model of Ed Chi, is under development. By basing the architecture on an operator-centric, rather than data-centric view, we anticipate a reduction of the code complexity as well as a facilitation of future extensions. It is being developed in Eclipse using Qt as the under interface library, which will enhance its portability. This is scheduled for release in the Fall of 2009.

Contributions Beyond Science and Engineering:

One of the PIs (Ward) has been involved in the development of the National Visualization and Analytics Center, which is being funded by the Department of Homeland Security. Some of the concepts developed in this research project, especially the integration of data with variable quality/confidence, play an important role in the analysis of intelligence data. He was also a participant on the NSF/NIH panel that developed the visualization research challenges report, which will likely be a key guiding force in the long-term development of the field of visualization and its application to many societal problems.

<u>**Conference Proceedings**</u>

<u>**Categories for which nothing is reported:**</u>

Organizational Partners

Any Product

Any Conference