iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

On the Design of an Immersive Environment for Visual Mining: Challenges and Results from an Early Study

E. Mozaffari, A. Seffah, A. Blakney and R. Naghshin
Environmental Software Technology Lab
Department of Computer Science and Software Engineering
Concordia University, Montreal, Canada

Abstract: Observational systems, with the complicity of Internet and information technology, have introduced new ways for collecting and disseminating environmental data. As a consequence, the analysis of increasingly large and complex environmental data sets has become a difficult and time consuming process which can hinder analytical comprehension and insight extraction for decision-making regarding humanity's environmental challenges. This paper introduces new concept of visual mining which presents the data in some visual form, allowing the 'human' to gain insight into the data, draw conclusions, and directly interact with the data. In addition, it discusses some principles toward VTab, a visualization tangible board to support visual mining as part of the whole environmental data lifecycle.

***Keywords**:* Visual Mining, Environmental data visualization, Immersive visualization.

1. INTRODUCTION

The visualization and analysis of large scientific data represents a very challenging task, especially in the earth, space and environmental sciences. The environmental data are complex, very large in volume, multidimensional, dynamic (time-varying in data and dimension) and consist of many parameters. It is important to take advantages of this flood of information, make sense of it and turn raw data into understandable information. Visualization provides an important means to handle the information for knowledge discovery processes. However, the visualization of data alone is not sufficient, human interaction with data should also be provided in a way that user can access and use the information.

Our goal is to move beyond "traditional" approaches to human-computer interaction to focus not only on how humans interact with machines, but on human access to and use of information. An important step toward meeting this goal is to assess existing technologies and develop new environmental data visualization methods and tools from a human-centered point of view that puts the emphasis on user needs. We believe that the most efficient way to achieve this is through a visual computing process that allows users to transform the raw data into useful knowledge and patterns via the use of visual representations and immersive interaction techniques to facilitate insight into that data.

This paper proposes some principles towards an immersive tangible visualization toolkit for the integration, interaction and visualization of large collections of environmental data. In addition to basic interaction and visualization tools, our goal is to provide data comprehension and visual mining features to support the whole environmental data engineering lifecycle. This includes: (1) the analysis of data, (2) complex phenomena and problems modeling, (3) design of solutions to overcome the problems or to understand the phenomena, and (4) decision making in crisis situations. We intend to develop and validate new types of interactions that are appropriate to interact with different kind of environmental data.

2. FROM VISUALIZATION TO VISUAL MINING

Data mining has proven its capabilities to outperform traditional spatial analysis and statistical methods with better scalability, strong fault tolerance, and more importantly, the capability of inference/reasoning tasks (Miller et al., 2001). As a result, over the last decade, increasing efforts have been put into applying data mining techniques to handle the complexity of geo-spatial problems. Early studies (Koperski et al., 1996; Tung et al., 2000) show the capability of data mining to discover patterns in data. These efforts covered a range of spatial classification, association rule extraction, clustering analysis, outlier analysis and trend detection.

These efforts opened new possibilities for inductive spatial reasoning in the process of exploring both spatial and non-spatial attribute spaces. Geographers also realized potential advantages resulting from integrating data mining methods with geographic visualization tools to support human pattern-identifying abilities (Miller et al., 2001).

Visual data exploration aims at integrating the 'human' into the data exploration process by applying 'human' perceptual abilities to the large data sets available in modern computer systems. The basic idea of visual data exploration is to present the data in some new or more useful visual form, which allows the 'human' to get insight into the data, draw conclusions, and directly interact with the data. Visual data mining techniques have proven to be of high value in exploratory data analysis and they also have a high potential for exploring large databases (Keim, 2002).

Visual mining (Mozaffari et al. 2008) combines visualizations techniques and data mining. The goal of this area is usually to improve understanding of the data being presented. As 3D visualization techniques, such as Virtual Reality (VR), become more available and more commonly used, new and powerful visualization technologies, such as interactive and immersive visualization, are becoming an important part of the scientific process. Many tools and techniques are currently available to transform environmental data to its 3D representation and synthesize the data.

The integration and analysis of diverse data about various phenomena, in current Geographic Information Systems (GIS) and statistical tools, are usually represented as sampled or simulated data (e.g. on a regular or non-regular grid, or as point data). While GIS and existing tools are often used for integrating heterogeneous data, they still lack the capabilities to visually comprehend and link multivariate data while identifying relationships.

Building on data mining, we define visual mining as a process that:

- Aims at integrating the 'human' in the data exploration process with the goal to enhance their perceptual abilities to explore large data.
- Presents the data in some visual form, allowing the 'human' to gain insight into the data, draw conclusions, and directly interact with the data.
- Maps and transforms computer data into perceptual and visual representations that link factual (facts which can be reasonably assumed to contribute to the investigation, understanding, and solution of the problem) and singular (the available data) data while facilitating the overall comprehension of a large set of information.
- Lets the 'human' interact with and engineer data by manipulating visual representations, either as data objects or as objects representing relationships between data that are not necessarily visual themselves.
- Combines visualization techniques and the iterative cognitive process of data mining: (1) Filtering and analysis of data, (2) Modeling of complex phenomena and problems, (3) Design of solutions to overcome the problems or to understand the phenomena, and (4) Decision making.

As a cognitive information processing process, three major milestones characterize visual mining:

1) Gaining insight into singular data. Using different visualizations, the 'human' can build up an understanding of their data sets.
2) Relating factual and singular data. Besides visualization techniques, the 'human' exploits interaction techniques to directly interact and link the visualized data and navigate among links while relating data and facts.
3) Identifying patterns. Advanced data mining and visualizations techniques may be used for uncovering unknown patterns and relationships in large data sets.

Visual mining concept is similar to visual analytics (Keim et al., 2007) but it uses different visualization techniques simultaneously enhance users perceptual abilities to explore large data set and help them to relate actual and singular data in order to discover patterns in the large datasets. In addition, user can engineer data by manipulating visual representations, either as data objects or as objects representing relationships between data that are not necessarily visual themselves. In addition,

3. AN OVERVIEW OF THE PROPOSED INFRASTRUCTURE

Our objective is to provide an immersive platform for the integration, interaction and visualization of large collections of environmental data. A specific goal is the development and validation of a toolkit to support the visual mining process. The toolkit would offer interaction and visualization services to help users get insight into the data, relate factual and singular data, and discover patterns.

Figure 2 provides an overview of the proposed infrastructure. *Architecture of the toolkit* summarizes some of the services included in the toolkit. Table 1 contains a summery of descriptions of these services.

The *physical structure* portrays the proposed technological setup. The proposed visualization immersive tangible board (VTab) is built around a physical table that embeds a large interactive whiteboard and a second monitor to create a tangible workspace. Environmental data are represented by physical forms directly manipulable and perceivable by human sensors. Sensing based interaction mechanisms are used, allowing users to interact directly with the tangible forms on the large whiteboard. While exploring data, users will also be able to annotate data using, for example, real-time audio/video.

As it is shown in the *physical structure,* beside the interactive table, the environment consists of a ceiling-mounted video projector projecting computer generated world onto a white table around which the users sit. The users are equipped with an HMD system and electronic data gloves. They can be immersed into the virtual environment and they will be able to interacts with and engineer environmental data by manipulating visual representations. Sensing based interaction allows users to take advantage of their many senses and provides interactions that are closer to human cognitive processes. Also, this system includes a vertical display that shows information in more detail. It will be focused on the same position in the 3D model that user touched on the table.

We believe that VTab enhances comprehension and visual mining capabilities to support the whole environmental data engineering lifecycle. The *Visual mining process* in figure 2 shows visual mining milestones (described in section 2) which can be achieved using proposed approach. In order to validate proposed approach empirical user study should be conducted.

The followings are steps that should be taken in order to perform the empirical studies:

1) Observe and describe a ***phenomenon*** which is an interaction between a human and a computer.
2) Formulate a ***hypothesis*** (research question) to explain it.
3) Use the hypothesis to ***predict*** (predictive model) or ***describe (***descriptive model***) other phenomena.***
4) Perform ***user study*** to test hypothesis.

Then the user study should be perfumed in a sequence shown in user study protocol (more details are given in section 4).

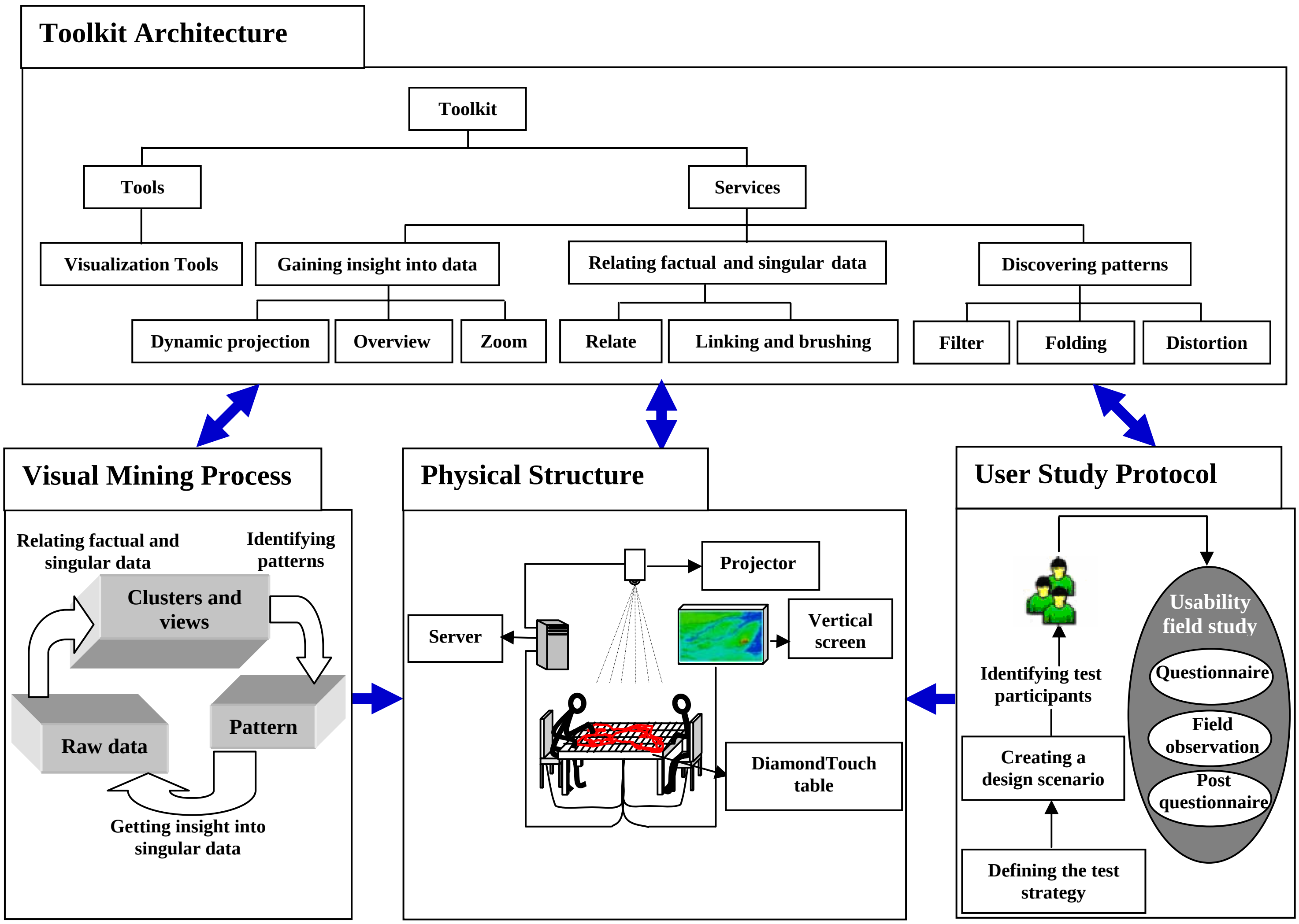

Figure 1. The proposed infrastructure

Table 1. Examples of services provided

Visual Mining Milestones	Tool	Description
Gaining insight into data	Zoom	Look at specific items of interest
	Overview	Gain an idea of the entire collection
	Dynamic projection	Dynamically change the projections
Relating factual and singular data	Relate	View relationship among items
	Linking and brushing	Multiple views dynamically linked, a subset of the data can be interactively selected and selected data portion will be highlighted in related views
Discovering patterns	Filter	Eliminate uninteresting items
	Folding	Pull a portion of the dataset: distant portions of the dataset can be positioned side by side for easier comparison
	Distortion	Show portion of data with a high level of detail while others shown in lower level of detail

4. RESULTS OF AN EARLY STUDY

For a better understanding of interaction techniques for visual mining, we conducted a field observation (Figure 2). Originally in this method the investigators observe the users in their natural environment while they are doing the related tasks (Nielsen 1993, Blomberg 1993, and Plowman 1995). In our study, we simulated the situation by creating a design solving problem and inviting users to participate in two three-hour tests sessions. After the field observation, we also asked users for their opinion about the tools they used. Observing the participants and analyzing the data gathered from the questionnaires and the interviews, we slightly modified our original physical setting.

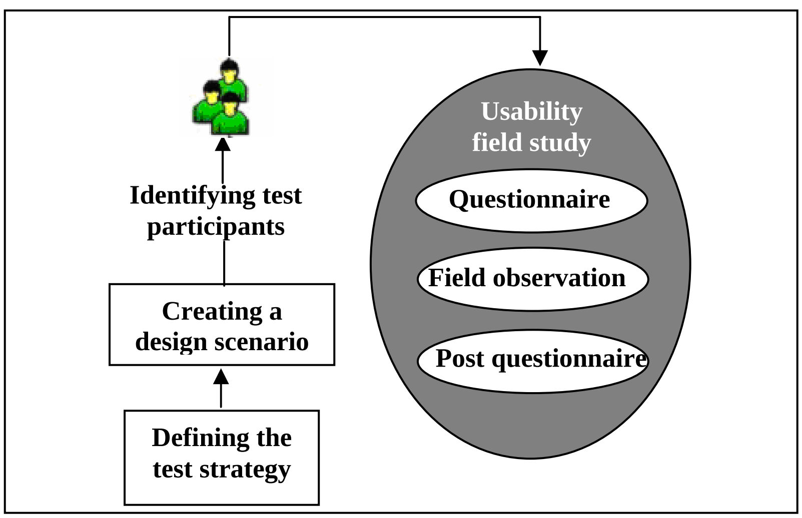

Figure 2. Protocol used in our field study

First, we used two computers (one user station one observer station) a digital camcorder, wireless microphone and the software Morae. Morae software was recording the participants' interactions with the designers' station by recording the computer screen output, as well as mouse clicks and keystrokes. It was also recording the output from the digital camcorder and the microphone. It then combined everything into one editable data file. It also provided us with real-time remote viewing and annotation (across a computer network) of a recording in progress. We realized that one camera was not enough for recording all interactions. Therefore we modified our setting (Figure 3).

1. Digital camera capturing the user, using the computer which recorded via Morae
2. Overhead camera capturing the table from the top recorded via Adobe premier
3. Corner camera capturing the participants and the white board
4. Projector for projecting the computer screen on the white board
5. Design station consisting software design tools (the users' interaction with this station was recorded via Morae)
6. Design station consisting multimedia software
7. Observer station recording overhead camera and Morae
8. Observer station recoding corner camera
9. Microphone capturing participants' conversations
10. Observers
11. Participants
12. White Board
13. Dividing panel between monitor station and the design environment

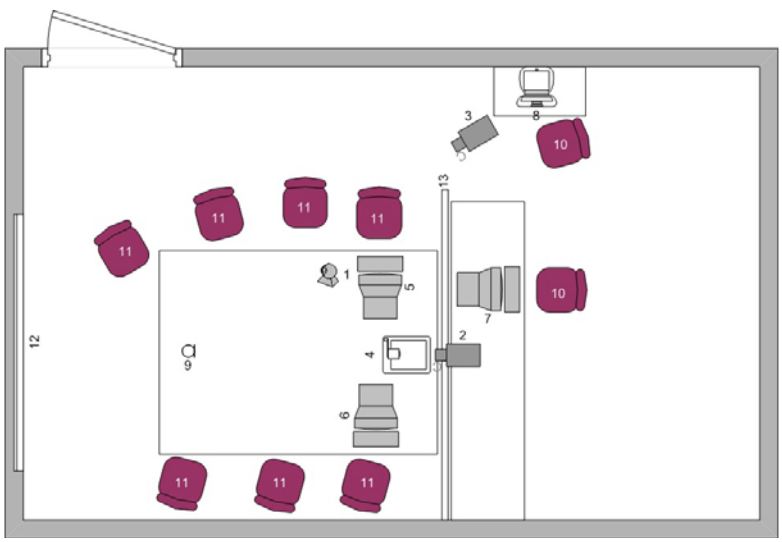

Figure 3. Study infrastructure

The tools that were provided for the designers in the first session included:

- Design stationeries (different pen, pencil, papers and markers),
- One computer including software such as Microsoft word, Microsoft Visio, and Photoshop.
- A white board with different color of markers for brainstorming,
- A projector was connected to the design station. The designers could project the monitor on the whiteboard.

Getting the feedback from the designers in the first session, we added more tools including:

- One computer as design station,
- Adobe Illustrators and Macromedia Flash for graphic designers,
- Rational Rose for software designers,
- Internet access.

The findings from this study are as fallows:

Software versus tangible tools. The most popular tools in both sessions were pen and papers and the whiteboard. Designers attempted several time to use computer software such as Photoshop for creating the interface and game characters or Flash for making a prototype. However, they found these software tools time consuming and impractical so they went back to pen and paper. One of the software designers, after producing his class diagram on the paper, asked monitors if this was enough or he had to redraw them with Rational Rose in order to make it more presentable. The only effective use of software was during the last hour of the first session when they were documenting their design decisions using Microsoft Word. They projected the computer screen on the whiteboard, and one designer was typing others' suggestions.

Whiteboard versus pen and paper. Both tools were used extensively in both design sessions and were rated as the most usable tools. As we noticed, pen and paper are used mostly for individual works or when a small group of designers work together, whereas the whiteboard is used during the design brainstorming. The large size of a white board allows a larger group of designers to present and explore the design ideas rapidly. However, a horizontal interface where the designers can sit and work for a longer period of time will be more preferable during the detail design.

Some of the requirements, we gathered from this field study are:

- The new environment should accommodate multidisciplinary activities using a large vertical interface as well as team activities using tabletop.
- Each user, representing a certain design discipline, would have their own horizontal table customized according to their design perspective.
- The tables would be connected to the hub which is connected to a vertical smart board, and the information should be shared among the disciplines by transferring any design artifact to the smart board. The information on the smart board could be modified during brainstorming by the whole team (Figure 4).
- The software used in this setting should help coordinate multiple points of view. Design experts from different disciplines bring different point of view, which should be converged in the final product. Early design tools must present and visualize design knowledge in a way that accommodates multiple design perspectives and goals.
- Using the existing CAD tools based on WIMP interaction for sketching, storyboarding, and prototyping is extremely expensive and time consuming. However, the disadvantage of making sketches using pen and paper is that these sketches are hard to modify and evaluate as the design evolves. The new setting should be based on tangible and sensor interaction. It will support natural interaction used for sketching and rapid prototyping, making the design iterations faster.

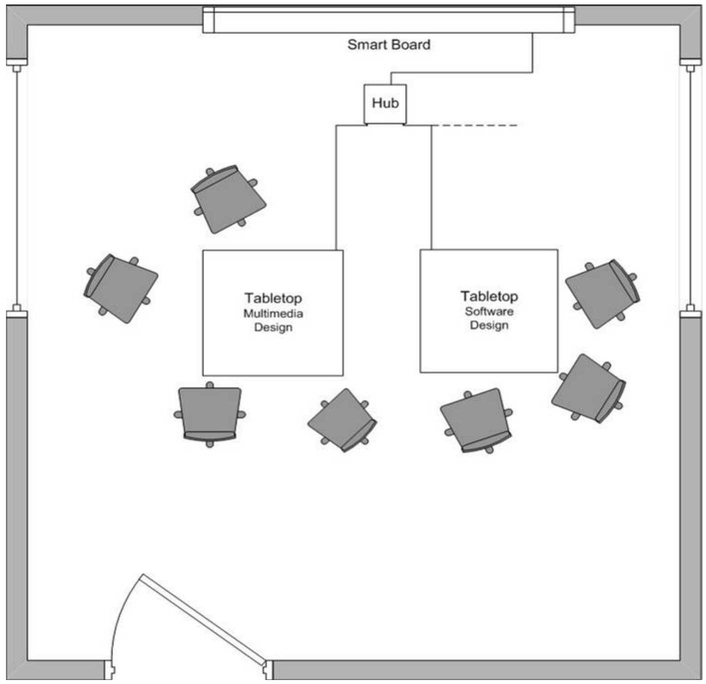

Figure 4. Layout

5. **CONCLUSION** and **CHALLENGES AWAITING FOR FURTHER INVESTIGATIONS**

This paper introduces an innovative visualization tangible board for enhancing comprehension and visual mining capabilities to support the whole environmental data engineering lifecycle. The proposed system includes visual paradigms to enable the human mind to process vast quantities of information, interaction paradigms to allow a higher order of interaction that is closer to the human cognitive processes, and physical devices to allow humans to take advantage of their sensory perception. In addition we described proposed framework. In the future, we intend to develop new types of interactive methods which are more appropriate for interactions with different kinds of environmental data, and later we will develop formal methods to assess the usability of these interaction techniques.
There are many research challenges that are associated with visual mining process. The following is a list of challenges in our scope of work:

1) Develop new types of interactions that are appropriate to interact with different kind of environmental data.
2) Perform empirical studies to test the effectiveness of these interactions. There are three themes:

- Raise the research questions: Formulate a hypothesis to explain the human computer interaction.
- Observe and measure: Gathering observation either manually (human observers) or automatically (computers, software, sensors, etc.). In addition the measures should be defined in a way that they fulfill the requirements.
- User studies: Perform user study to test hypothesis. The data should be analyzed to answer the research question.

6. REFRENCES

Keim, D. A., Schneidewind, J., 2007: Introduction to the Special Issue on Visual Analytics, *SIGKDD Explorations Special Issue on Visual Analytics*, 9, 2007.

Keim, D.A., Information visualization and visual data mining. *IEEE Transaction on Visualization and Computer Graphics*, 8, 1-8, 2002.

Koperski, K. and Han, J., Discovery of spatial association rules in geographic information databases. *In Advances in Spatial Databases, Proceedings of 4th Symposium, SSD'95*, Springer-Verlag, 47-66, 1996.

Miller, H. J. and Han, J., Geographic data mining and knowledge discovery. *Taylor and Francis,* London, 3–32, 2001.

Mozaffari, E. and Seffah, A., From visualization to visual mining: application to environmental data, *The First International Conference on Advances in Computer-Human Interaction,* 2008.

Nielsen, J., Usability engineering. London: Academic Press, 1993.

Plowman, L., Rogers, Y. and Ramage, M., What are workplace studies for?". *Proceedings of ECSCW 95*, 309-324, 1995.

Tung, A., Hou J. and Han, J., COE: clustering with obstacles entities, a preliminary study. *In Proceedings of 2000 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'00)*, 3, 165-168, 2000.