Framework for Visual Analytics of Measurement Data

Paula Järvinen, Pekka Siltanen, Kari Rainio
VTT, PL 1000, 02044 VTT
Espoo, Finland
{paula.jarvinen, pekka.siltanen, kari.rainio}@vtt.fi

***Abstract*-Visual analytics is a recent research field for finding knowledge from data masses. It combines the strengths of automatic data processing, and the visual perception and analysis capabilities of the human user. This article represents a framework for analysis of measurement data with the help of visual analytics. The framework consists of a message service platform for collecting and unifying measurement data, and a visual analytics tool for analysis. The platform receives data from different types of measurement devices and transforms the data into unified format that applications can use through standard interfaces. With the visual analytics tool users can analyse the data coming from the platform. The tool shows data as interactive visualizations, both abstract and descriptive, and suggests analysis methods to the user. User performs reasoning by interpreting series of consecutive visualizations. The framework frees the analyst from the time demanding data pre-processing and makes analysis possible without special skills and expertise in data analysis. A prototype application to demonstrate the framework is presented. Examples with real data show how explanations for building energy consumption can be reasoned with the help of the visual analytics tool.**

Keywords- Visual analytics, measurement data

I. INTRODUCTION

Advances in sensor and data processing technology have extensively broadened the reach of measurement data. In industry, sensors monitor the state of machinery, environmental meters collect measurements from the environment, building automation systems store building information, and enterprises track the actions and behavior of consumers.

The accumulated data is expected to contain important knowledge that can be used to improve processes and support decision making. But the collected data coming from different sources and forms requires a considerable amount of pre-processing; the relevant data must be recognized and extracted from the data sources, and noisy data has to be cleaned. Data values can be in non-comparable units that have to be harmonized. Before analysis and visualization data needs to be transformed to specific data formats. The variety of data can be confusing to the user. Even a simple operational data base can contain dozens of data entities and each entity dozens of attributes. The analyst has to find the most meaningful properties and the right analysis methods.

Visual Analytics is a recent approach to finding knowledge from data masses [1–7]. It provides tools to support analytical reasoning. A visual analytic tool processes the raw data and shows the information in the form of abstract and interconnected visualizations. Users can then look for patterns, trends, anomalies, similarities and other relevant "nuggets of information" from the visualizations. Users can launch analysis, browse and navigate visualizations and highlight and select important areas for further study.

This article represents a framework for visual analytics of measurement data. It consists of a message service platform that collects and pre-processes the measurement data, and a visual analytics tool for analysis. The framework provides a streamlined environment for analysis of measurement data where users reason by interpreting series of consecutive visualizations. The tool combines both information visualization and descriptive visualizations of the monitored objects. A prototype is constructed to analyze energy consumption and indoor air quality measurements of office buildings in a research campus area in Finland.

.

II. BACKGROUND

Visual analytics is defined as "the science of analytical reasoning facilitated by interactive visual interfaces" [1, 2]. It provides visual tools for finding insight from complex, conflicting and dynamic data to support analytical reasoning and decision making. The basic idea of visual analytics is to unite the strengths of automatic data analysis and the visual perception and analysis capabilities of the human user. Humans can easily recognize patterns, colour, shape, orientation and spatial position, detect changes and movement, and identify specific areas and items from visual presentations. Humans are good at reasoning and generating problem-solving heuristics [8]. Showing data and analysis results in the form of abstract visualizations is an efficient way to find insight from data [9]. Computers, on the other hand, have superior working memory and unlimited information processing capacity without cognitive biases [10]. The goal is to create systems that utilize human strengths while providing external aids to compensate for human weaknesses [2]. Visual analytics is especially focused on situations where huge amounts of data and the complexity of the problem make automatic reasoning impossible without human interaction.

Visual analytics is a multi-disciplinary research area, combining information visualization science, data mining, mathematical and statistical methods, data management, user interface techniques, human perception and cognition research. The area has several challenges, listed in [7], including management and integration of very large, diverse and variable quality of datasets.

The markets and research have produced a variety of visual analytics tools, including tools by major software houses, analysis and visualization environments, libraries, application-oriented tools, and technology-oriented tools. But, the use of visual analytic is not yet widespread. Market survey companies Frost & Sullivan, ReportLinker and Gartner Group have recently recognized the new field and made references to recommended application areas. Integrating a visual analytics tool to a measurement data platform and using both abstract and descriptive visualizations, as in this work, is a novel approach.

III. RESEARCH APPROACH

The research was implemented as constructive research, by

1. identifying the problems faced when using large amounts of measurement data to support reasoning based on the measurements,
2. creating a reasoning framework, combining the structural data model of the object monitored and the measurements,
3. building a prototype application to prove realistic implementation possibilities of the framework, and
4. testing the framework with real life reasoning examples.

A. *Identifying the problem*

Data quality is a well-known problem in data mining research [11]. The problem is partly caused by missing or erroneous measurements, as well as disparate data formats when combing data from several sources. In practical implementations, for example when comparing and estimating measurements (for example indoor temperatures, CO_2 values) from different buildings and from the surrounding environment, the data quality problem is faced immediately. Buildings have their own building automation systems, each storing the monitored data in different format. Things get even more complicated when the building monitoring data is combined with environmental measurements, such as outdoor temperatures or aerosol emissions that are again represented in different formats.

More problems are faced when trying to use the measurement data to support reasoning. Often the data is handled by facility management systems that reduce the data into simple performance indices and trend graphs. Even though the tools used by the data reasoning experts, such as R Statistical computing package [12] contain functions to import data from different sources, the task of transforming and importing the data is often the most time consuming task in the reasoning process. Also, the data analysis tools often have poor visualization capabilities, producing just static pictures for human viewing.

B. *Framework*

Figure 1 shows the framework concept. Measurement data is collected from sensors and databases and delivered to applications through a message service platform in unified format. Our application, the visual analytics tool receives data from the platform and stores it into its own database. The database consists of time stamped measurement values, together with background data, properties of the monitored object and metadata of measurements. Data is stored in such a form that visualizations and analysis can be easily applied. Performance indices and other quantities are calculated from the data. In our experiments, building sensor measurements were used, but the data can be as well environmental data or human behavior data.

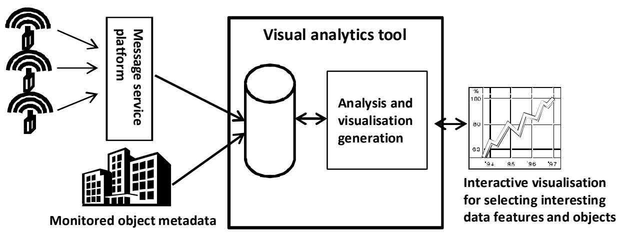

Figure 1. Framework concept

A set of pre-defined data analysis and visualization methods is coupled to the database. In our example a set of analysis methods that is relevant for measurement data, such as finding correlations, trends, outliers, interesting subsets and clusters from data, is selected. When a new data set requires analysis, the data from the platform is loaded to the visual analytics tool database. The predefined analysis and visualization methods would be automatically available.

All analysis show results in visual form and the visualizations are interactive. Visual reasoning is intensified by combining descriptive models of the monitored object (e.g. maps, 3D models or product structures) with abstract visualizations. This approach is expected to strengthen the interpretation of the results and help users set the results into right context. MMEA prototype

During the reasoning process user selects variables from the database for analysis. The tool guides the user to choose the most suitable analysis and visualizations based on user's selections. For example, if user has selected power and water consumption measurements for analysis of a whole week, user is suggested calculation of correlations and scatterplots with regression lines and time series of correlations, to start with. Users can select data sets from the visualization for further analysis.

A web application was constructed using the framework. The monitored objects consist of office buildings of a research campus area in Finland. Measurements of power, reactive power, district heat and water used of each building, and environmental measurements including outdoor temperature and relative humidity were available, with measurement frequency measurement/hour. Indoor temperature, occupancy of people and CO_2 measurements were available of one office building. In addition, the building gross volume, gross floor area and building age were available. A map of the campus area was used as descriptive visualization.

The user gets a visual user interface (Figure 2), where the user can select monitored buildings and their properties for analysis. The buildings can be classified and visualized using different symbols or colors to help selecting.

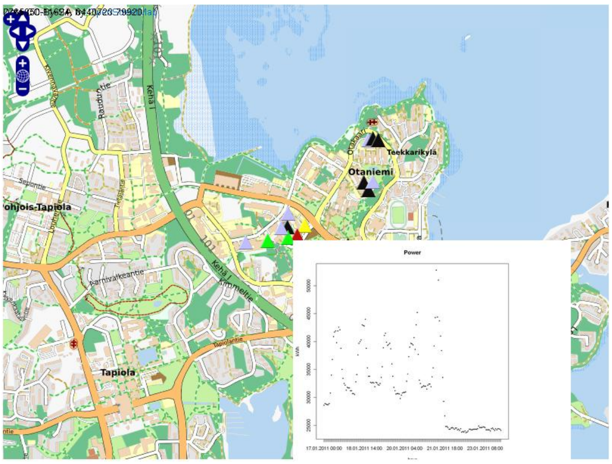

Figure 2: User interface

An index for energy performance for each building is calculated. The set of visualizations and analysis methods include basic statistics (min, max, mean, standard deviation), histograms, time series with different time granularities (hour/day/week), clustering, correlations, scatterplots, linear regression, time series of correlations, autocorrelations and cross correlations.

C. *Implementation details*

The current prototype application uses a set of open source technologies for implementing the required features (Figure 3). The user interface is implemented as a browser interface, showing the monitored buildings on Open Street Map environment. The user interactions and the visual markers representing the monitored building on the map are implemented using Openlayers JavaScript library. The building 3D geometry details can also be shown in the browser interface, by using X3DOM technology.

The MMEA message service platform takes care of collecting the data from different sensors and other measurement sources, as well as transforming the measurements to unified data representation. The measurements, together with measurement metadata and building properties from the Real estate information register maintained by Finnish Population Register Centre, are stored in the MySQL data base.

Statistical computing is performed using R statistical library [12]. In the current prototype version, the graphics for example shown in Figure 2 are generated using R, too. However, in the future the R graphics will be replaced with more visually pleasing and interactive JavaScript visualizations, using JavaScript visualization packages such as Google Visualisation API.

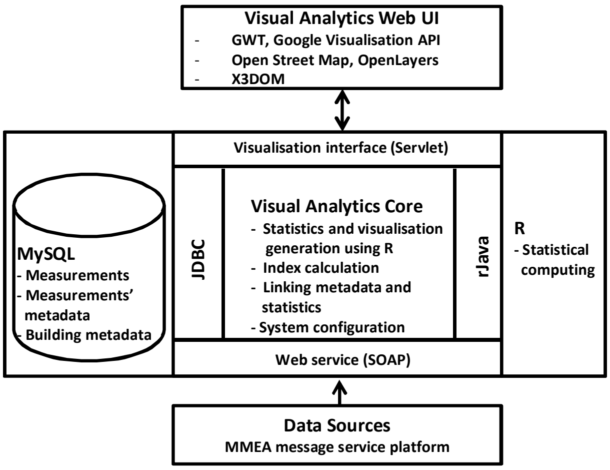

Figure 3. Prototype architecture

Interfaces between MMEA Visual Analytics tool core and other components are implemented using standard technologies: browser server communication is implemented using REST servlets, database is connected to the core via JDBC, R statistical library is integrated to the core using rJava package, and data is received from message service platform using Web Service interface.

IV. EXAMPLE OF USING THE FRAMEWORK

The goal of the visual analytics tool is to help users find interesting phenomena and explanations to support decision making. It is performed by interpreting series of consecutive visualizations. The principle is presented here by simple examples.

Let us assume that a user is interested in the energy efficiency of the campus area buildings. First, the user gets an overview of the energy efficiency of the buildings on the map (Figure 4). Different colours indicate the energy performance index values of each building, showing a rough estimate of energy performance of each building. User can click the buildings and get more details.

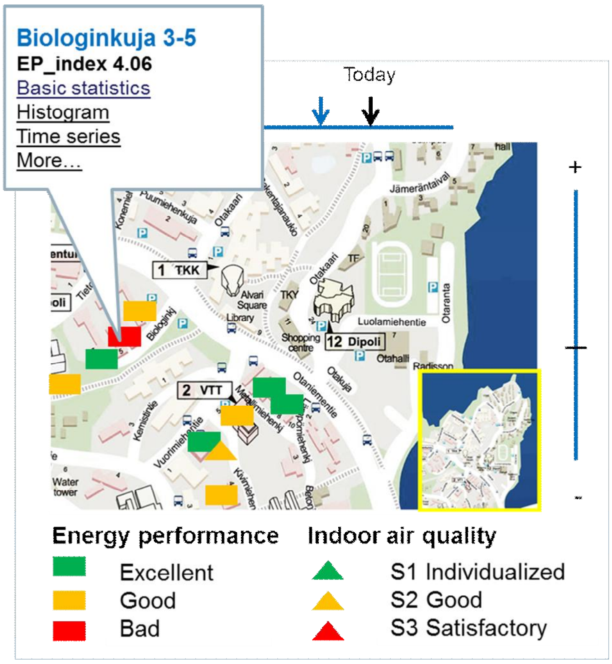

Figure 4. Overview.

To get a better overview, the user wants to see the basic statistics and time series of all the buildings (Figure 5). They show two interesting things: Most buildings have low energy performance index (i.e. low energy usage) but some have very high values. The time series show that energy usage on weekend, and on day and night differ a lot.

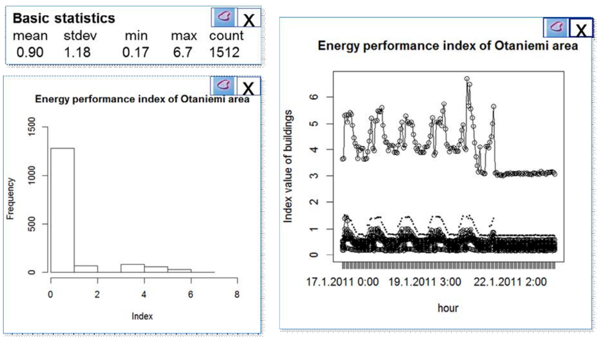

Figure 5. Basic statistics

Next, user wants to find reasons for the big variances in the energy performance index values. User considers if the building age is behind the high energy performance values. A scatterplot with correlation coefficient value 0.43 and an ascending regression line give hint that there is some correlation between age and energy performance index (Figure 6). The user can select an object from the scatterplot and see the corresponding buildings on the map. In the scatterplot user has paid attention to plots that indicate high index value and building age. They prove out to come from one building that is highlighted on the map. Similarly, user can study other background variables.

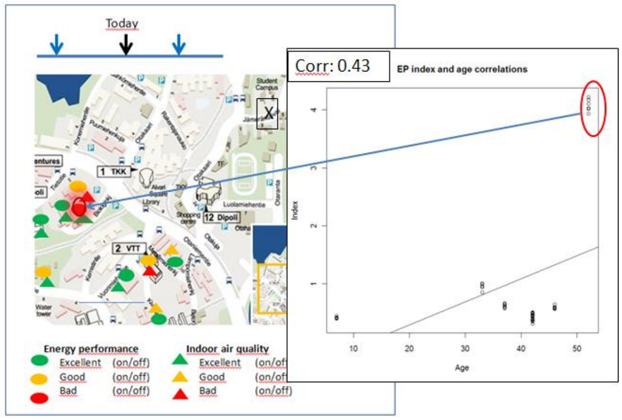

Figure 6. Correlation with age

User can also select data subsets of visualizations and perform analysis with subsets. In the example in Figure 7 three clusters are formed based on the indices, and the statistics and details of one cluster are shown.

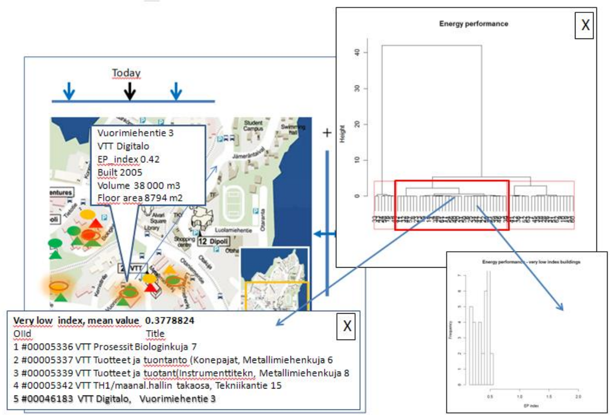

Figure 7. Analysis of subsets.

The second example is about multivariate analysis. It studies correlations between the energy performance index, district heat and water consumption (Figure 8).

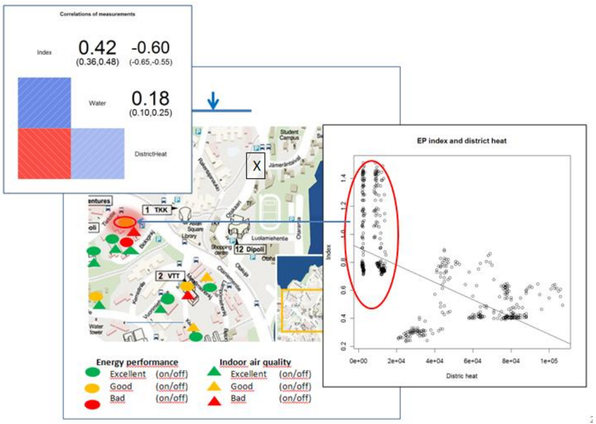

Figure 8. Multivariate analysis example

It shows that there is a strong correlation between district heat and energy performance. User wants to study more this correlation and produces a scatterplot. It shows an interesting pattern that the user wants to study more. It proves out to represent one building that is highlighted on the map.

V. Conclusion and Future Research

A novel visual analytics based approach to tackle the information overload problem was introduced. It consists of a framework including a message platform and a visual analytics tool for experts and researchers to analyze measurement data. The approach was evaluated with identification and diagnostics of problems from building energy consumption.

The trials suggest that the framework provides solutions to the introduced data analysis problems. It provides a streamlined environment for analysis of measurement data. Data is available from the message service platform in a standard form, avoiding the time demanding pre-processing. The analyst gets high quality data from the platform and can focus on the analysis task. Also the same pre-processed data is available to other applications, thus avoiding overlapping work.

Data analysis has been considered a job that requires special skills and expertise. With the suggested visual analytics tool the user needs no beforehand knowledge of data analysis or statistical methods. The complex analysis algorithms are hidden behind the user interface. The user only needs to select data, launch analysis from a pre-defined set of suitable analysis and interpret the results. Using abstract visualizations together with descriptive visualizations of the monitored objects intensifies the interpretation of the results. Thereby the data analysis can be performed by the domain expert, instead of a data analysis expert, as before.

Taking new datasets for analysis is straightforward. The data model of the tool, the user interface and the analysis methods are general purpose and can be applied as well to environmental data or human behavior. Data coming from the message platform is ready for analysis after simple specifications. If new analysis methods are required, they can be easily taken into use from the R statistical package resources. The only time demanding tasks are adding the application specific calculations, such as performance indices, and adding the descriptive visualizations.

The current prototype does not contain a full set of interactive features. A real visual analytics application would enable making most of the interactions needed in reasoning by interacting with the visualizations generated by the system. Creating truly interactive visualizations in the browser environment is a challenge that will be tackled next. Another objective would be to develop further the user support in reasoning. At the moment the tool suggest analysis methods that are based on the qualities of the selected data. In addition to them knowledge of user actions and semantics could be utilized. The user could also discard obvious relationships to diminish the problem space. The third goal would be to study the usability of the solution: does the framework help users to utilize measurement data, find interesting phenomena and explanations, and improve decision making

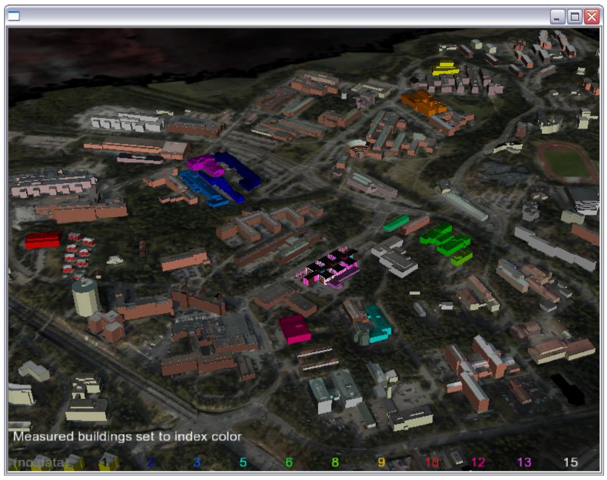

Figure 9. 3D model of the campus area with monitored buildings highlighted

The framework described above enables linking the measurements into the object that is monitored: in examples above map based visualization of a set of buildings was used as a user interface to access the reasoning algorithms. Currently the prototype supports map visualization and experimental 3D model visualization showing the monitored objects in the browser. In the future the visualizations of the monitored object will be further developed. Also, the prototype architecture enables embedding the visualizations to other user interfaces, such as aerial 3D model, allowing the user navigate in the 3D model instead of using the map user interface (Figure 9).

Acknowledgement

This work has been partly funded by the CLEEN Cluster for energy and environment, a company managing the Strategic Centre for Science, Technology and Innovation in the theme energy and environment, Tekes, the Finnish Funding Agency for Technology and Innovation, and VTT Technical Research Centre of Finland.

References

1. Wong, P.C. & Thomas, J. Guest Editors' Introduction--Visual Analytics. IEEE Computer Graphics and Applications, 24 (5): 20-21 2004, Vol. 24, No. PNNL-SA-41935.

2. Thomas, J.J. & Cook, K.A. Illuminating the path: The research and development agenda for visual analytics. 1st ed. Los Alamitos, CA: IEEE Computer Society, 2005. ISBN 0769523234.

3. Keim, D.A., Mansmann, F., Schneidewind, J. & Ziegler, H. Challenges in visual data analysis. Tenth International Conference on Information Visualization. July 5-7. London, UK: IEEE, 2006. Pp. 9.

4. Cook, K., Earnshaw, R. & Stasko, J. Guest editors' introduction: discovering the unexpected. Computer Graphics and Applications, IEEE 2007, Vol. 27, No. 5, pp. 15-19.

5. Wong, P.C. Visual analytics science and technology. Information Visualization 2007, Vol. 6, No. 1, pp. 1-2. ISSN 1473-8716. doi: 10.1057/palgrave.ivs.9500149.

6. Keim, D., Mansmann, F., Schneidewind, J., Thomas, J. & Ziegler, H. Visual analytics: Scope and challenges. In: Anonymous Visual Data Mining. First edition ed. Heidelberg: Springer, 2008. Pp. 76-90.

7. Keim, D., Kohlhammer, J., Ellis, G. (eds.). . Mastering the information age: solving problems with visual analytics. First edition ed. Goslar, Germany: Eurographics assiociation, 2010.

8. Card, S.K., Mackinlay, J.D. & Shneiderman, B. Readings in information visualization: using vision to think. 1st ed. Morgan Kaufmann, 1999.

9. Ware, C. Information visualization: perception for design. Vol. 22. 1st ed. Morgan Kaufmann, 2004.

10. Green, T.M., Ribarsky, W. & Fisher, B. Visual analytics for complex concepts using a human cognition model. IEEE Symposium on Visual Analytics Science and Technology VAST'08. 19-24 Oct. Columbus, Ohio, USA: IEEE Computer Society, 2008. Pp. 91.

11. Rahm, E. & Do, H.H. Data cleaning: Problems and current approaches. IEEE Data Engineering Bulletin 2000, Vol. 23, No. 4, pp. 3-13.

12. Venables, W.N. & Smith, D.M. An introduction to R, 2012, (October 26-last update). Available: http://cran.r-project.org/doc/manuals/R-intro.pdf [2012, October 12] .