# Coordinated Multiple Views: a Critical View

Gennady Andrienko and Natalia Andrienko
*Fraunhofer Institute IAIS*
*Schloss Birlinghoven, 53754 Sankt Augustin, Germany*
*http://www.ais.fraunhofer.de/and*
*{Gennady|Natalia}.andrienko@iais.fraunhofer.de*

*Keywords* — **Coordinated and Multiple Views, Spatial and Temporal Data, Scalability, Exploratory Visualization, Decision Support, Visual Analytics**

Hundreds of papers have been presented at the five conferences on Coordinated and Multiple Views (CMV) in Exploratory Visualization and related conferences on information visualization and visual analytics [R07]. However, CMV are still rarely implemented in commercial systems and many users are even not aware that multiple views are useful and coordination can help them to solve their problems. Even toolkits especially designed for information visualization often do not offer any coordination mechanisms for multiple views. One of possible reasons for this may be that existing CMV tools and approaches are insufficiently suited to real-life problems.

Most of the CMV tools deal with single-table data sets of rather small sizes. A typical implementation allows one to create several displays (statistical graphics and/or geographical maps) representing individual entities and/or aggregates (e.g. histograms) and link them by brushing and selection. The displays are coordinated through exchanging references to selected entities. To our best knowledge, this approach works efficiently for tables with sizes up to about $10^5$ records. Linking displays that reflect larger tables causes significant delays.

There are practical applications that require analysis of much larger and complex data sets. Visual analytics [VA05], a science of analytical reasoning supported by visual interfaces, deals with problems that require making grounded decisions on the basis of voluminous, heterogeneous, and dynamic data sets. In our opinion, work with such data sets is not yet properly supported by existing CMV systems, which are not scalable.

The scalability problem has four major aspects: amount of data to be analyzed, dimensionality, complexity of information, and dynamics.

1. Amount of data. When a dataset is large, visualizations that represent individual entities suffer from overplotting. Even if just one pixel of the screen is used for representing one entity, as in pixel-oriented techniques [K00], it is impossible to present more than $10^7$ records on a screen. Traditional visualizations such as scatter-plots or parallel coordinates plots have much more restricted applicability.

Visualization methods that represent aggregates [UTH07] help to avoid overplotting. Examples of such methods are histograms, binned scatter-plots, and mosaic displays. However, any aggregation involves information loss, and therefore it is necessary to provide powerful tools for checking sensitivity to aggregation parameters.

Coordination between multiple displays for very large data tables requires significant computational power. For both individual and aggregated displays, it seems appropriate to use databases for storing and processing data. An open problem is the design of coordination mechanisms that are able to work with large data sets without loading them completely into RAM. Is it possible to implement dynamic querying or brushing in such a way that only a small part of the data is loaded in the memory and all processing is done by the database? What are practical limitations of such an approach?

One of the methods for improving the computational speed is the use of pre-defined aggregates in data warehouses and OLAP systems. However, this approach does not provide sufficient flexibility needed for data exploration [AA06].

2. Dimensionality of data. In practical applications it is often necessary to deal with multi-dimensional data. One aspect of this problem is the number of data dimensions. This aspect was in focus of the information visualization community for years, and we'll not discuss it here. The second aspect is that data may have such complex dimensions as geographical space and time [AA06] that require special attention. Geographical space is not limited to just two or three coordinates but includes the geographical context, which is difficult to formalize. Time has two models, linear and cyclical, and it is often necessary to consider simultaneously several temporal cycles (monthly, weekly, daily etc.; these cycles may overlap).

A combination of several ordinary dimensions with space and time dramatically increases the complexity and


0-7695-2903-8/07 $25.00 © 2007 IEEE
IEEE COMPUTER SOCIETY

is not properly supported by CMV yet. Let's consider an example of data about movement of entities [AA07]. The complexity of the problem is caused by the interplay of the properties of the geographical space, time, and moving entities. Dynamic query tools should process not only values of entity attributes but also dynamic characteristics of movement such as speed, direction, acceleration and turn, all in geographical context and temporal dynamics. Brushing between displays should be organized so as to allow seeing the histories of movement in the dynamically changing geospatial environment.

Design of interactive methods for analysing such data is a challenging task, and coordination of multiple views is essential for giving the user the necessary flexibility and power.

3. Complexity of data. Typically CMV tools are designed for dealing with a single table. Practical applications often require processing of several related tables. For example, in analysis of movement data it is necessary to work in parallel with tables describing moving entities, properties of their spatial positions, properties of various relevant objects located in geographical space, and properties of relevant events and processes occurring in the same time period.

A similar problem exists in data mining. Multi-relational data mining requires more sophisticated methods than those intended for single tables.

4. Dynamic data. There are practical applications that require simultaneous consideration of static data and continuous data streams. Imagine a group of coordinated displays that represent a situation at some moment. After new data come, these displays should show the changes in a right context. To our best knowledge, there are no CMV systems that address this complexity.

The scalability problem is in the focus of the emerging discipline of visual analytics [VA05]. In addition to the coordination of visualizations, it is necessary to coordinate visualization and computations. D.Keim [K05] proposed to replace B.Shnederman's Information Seeking Mantra "Overview, zoom & filter, details-on-demand" [SH96] by the Visual Analytics Mantra: "Analyse First - Show the Important - Zoom, Filter and Analyse Further - Details on Demand". The Visual Analytics Mantra stresses the fact that fully visual and interactive methods do not properly work with big datasets. It is necessary to start with database operations and computations ("Analyse First") and apply visualisation to the results obtained ("Show the Important"). The user may interact with the visualisation and the secondary data it represents (i.e. the outcomes of the analysis but not the original data), in particular, zoom and filter, and trigger further analysis, which, again, requires visualisation of the results. In this way, visual analytics is an iterative process involving three major steps, computational analysis, visualisation of the results of the computational analysis, and interactive visual analysis of these results. A detailed consideration ("Details on Demand") is possible for small data portions when they require, for some reason, a special attention of the analyst. This does not necessarily happen at the end of the process.

The challenging task is to design computational data analysis methods so that they integrate visualizations, thus making computations more intelligent due to involvement of users [CRC03]. There are only few papers that describe coordinated use of multiple visualizations and computational methods. Visualization systems usually consider computational modules as black boxes, only preparing inputs for them and presenting results.

Another dimension of the CMV world is the purpose of visualization. One of the possible purposes is exploratory data analysis and sense making. Another purpose is support of informed decision making, which requires comprehensive analysis of situation. Let's consider these two purposes.

1. Exploratory data analysis. A challenging problem of visual data analysis is the difficulty to externalize findings and synthesize knowledge from them. Several prototypes have been developed recently for collecting traces of user's actions during data exploration and replaying these traces. A complementary direction is to support users in annotating the process of discovery and findings being made. However, these approaches are limited to static data sets and, in most cases, to static visualizations. For example, how can one annotate the dynamics of a spatial pattern represented on an animated map?

2. Decision support. Evaluation of decision alternatives on the basis of their properties that have spatial and/or temporal characteristics is the problem that essentially requires using multiple coordinated displays. Making grounded justifiable decisions based on a comprehensive analysis requires special coordination mechanisms that reflect the nature the decision problem and the alternatives in their geospatial and temporal context [VA07]. There are only few works that address this problem, e.g. [AA03] for spatial multi-criteria site selection and [KM07] for evaluation of simulation scenarios on the basis of their dynamic properties.

One more problem of CMV is caused by the sheer number of the existing visualization methods. How can one find out which combination of methods is suitable and sufficient for analyzing specific data and for solving specific problems? We have recently developed a general framework [AA06] and demonstrated how it can be applied for the systematic design of a suite of CMV tools for solving a complex spatio-temporal problem [AA07]. However, such designs require significant efforts and are difficult to automate.

Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)
0-7695-2903-8/07 $25.00 © 2007 IEEE

Visualization is attracting more and more interest nowadays. The progress of the Internet and Web 2.0 technologies demonstrates exciting examples of analytical visualizations, which are actively used and developed by large distributed communities; see, for example, such systems as GapMinder (http://www.gapminder.org) and ManyEyes (http://services.alphaworks.ibm.com/manyeyes/home). The CMV community should not miss its chance to contribute to these processes.

**References**


[1] [AA03] N.Andrienko and G.Andrienko, "Informed Spatial Decisions through Coordinated Views", *Information Visualization*, 2 (4), 2003, pp.270-285

[2] [AA06] N.Andrienko and G.Andrienko, *Exploratory Analysis of Spatial and Temporal Data: a Systematic Approach*, Springer, 2006

[3] [AA07] N.Andrienko and G.Andrienko, "Designing visual analytics methods for massive collections of movement data", *Cartographica*, 42 (2), 2007, pp.117-138

[4] [VA07] G.Andrienko, N.Andrienko, P.Jankowski, D.Keim, M.-J.Kraak, A.MacEachren, S.Wrobel, "GeoVisual Analytics for Spatial Decision Support: Setting the Research Agenda", *International Journal Geographical Information Science*, 21, in press

[5] [CRC03] A.Ceglar, J.F.Roddick, and P.Calder, "Guiding knowledge discovery through interactive data mining", In P.C. Pendharkar (Ed.): *Managing data mining technologies in organizations: techniques and applications*, Hershey, PA: Idea Group Publishing, 2003, pp. 45-87

[6] [K00] D.Keim, "Designing Pixel-Oriented Visualization Techniques: Theory and Applications", *IEEE Transactions on Visualization and Computer Graphics*, 6 (1), 2000, pp.59-78

[7] [K07] D.A.Keim, "Scaling visual analytics to very large data sets", *presentation at the Workshop on Visual Analytics*, June 4th, 2005, Darmstadt, Germany, http://infovis.uni-konstanz.de/index.php?region=events&event=VisAnalyticsWs05

[8] [KM07] Z.Konyha, K.Matković, D.Gračanin, M.Duras, J.Jurić, H.Hauser, "Interactive Visual Analysis of a Timing Chain Drive Using Segmented Curve View and other Coordinated Views" (*this volume)*

[9] [P04] C.Plaisant, "The challenge of information visualization evaluation", in *Proceedings AVI 2004*, ACM Press, pp.109-116

[10] [R07] J.C.Roberts, "Where are we with Multiple Coordinated Views in Exploratory Visualization" (*this volume*)

[11] [SH96] B. Shneiderman. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations", In *Proc. of the IEEE Symposium on Visual Languages*, pages 336–343. IEEE Press, 1996

[12] [UTH07] A.Unwin, M.Theus, H.Hoffmann, *Graphics of Large Datasets: Visualizing a Million*, Springer, 2007

[13] [VA05] J.J.Thomas and K.A.Cook, editors, *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, IEEE Computer Society Press, 2005



0-7695-2903-8/07 $25.00 © 2007 IEEE