Challenging Problems of Geospatial Visual Analytics

Gennady Andrienko¹, Natalia Andrienko¹, Daniel Keim², Alan M. MacEachren³, and Stefan Wrobel¹,⁴

¹ Fraunhofer Institute IAIS (Intelligent Analysis and Information Systems), Sankt Augustin, D-53754, Germany
² University of Konstanz, Germany
³ Penn State University, USA
⁴ University of Bonn, Germany

Introduction
Visual analytics aims at combining the strengths of human and electronic data processing. This is achieved by means of visualization and interactive visual interfaces, which allow humans and computers to converse and cooperate (Keim et al. 2008). Visual analytics is conceived as a multidisciplinary research field in which scientists specializing in information visualization, scientific visualization, and geographic visualization closely cooperate with researchers from analytical disciplines, such as statistical analysis and modeling, machine learning and data mining, and geographical analysis and modeling, on developing new approaches to solving complex problems faced by the modern society. Geovisual analytics (or geospatial visual analytics) deals with problems involving geographical space and various objects, events, phenomena, and processes populating it. Since most of the things populating space occur or change in time, geovisual analytics must give proper attention to time and relationships between space and time.

This special issue on challenging problems in geovisual analytics has two major aims. First, the research papers included present a sample of problems that are dealt with and approaches that are developed in the area of geovisual analytics. This gives the reader an idea about the scope of the geovisual analytics research and about the diversity of approaches. Second, the group of guest editors has used this opportunity to reflect more generally on the mission of geovisual analytics, the ways in which the field is developing, and the issues that require attention of the community.

The guest editors’ thoughts are summarized in this editorial article. Alan MacEachren ponders on the variety of interpretations of the term “geovisual analytics” and the more general term “visual analytics” and thereby outlines the scope of the research area and the existing trends. Daniel Keim expresses his concerns about the deficiency of inter-disciplinary cooperation between the scientists specializing in visual and interactive techniques and the researchers focused on computational analysis, which may hinder the progress of visual analytics. D.Keim also notes a trend that visual analytics approaches are now developed within different application areas, which raises thoughts about the possible fate of the visual analytics as an independent research field. Stefan Wrobel presents a viewpoint of a computer scientist specializing in computational analysis methods. He believes in a great potential offered by combining these methods with visual and interactive approaches but notes that, unlike other areas of computer science, visual analytics has not yet established clear standards for evaluating results of research and development. Developing such standards is essential for the progress in the field. Gennady and Natalia Andrienko strive to attract
the attention of the visual analytics research community to the problem of preserving personal privacy, which may be endangered by the powerful and flexible tools for data analysis that visual analytics develops. Currently, the visual analytics researchers have directed little attention to privacy issues even though many of the tools developed have privacy implications. This situation needs to be changed. Close cooperation between visual analytics and privacy protection research is required.

In addition to these notes, the article includes a synopsis of the research agenda for visual analytics in relation to spatial and temporal data and problems that has been recently developed within the European project VisMaster. It persuades the geovisual analytics community to give proper attention to time and to users. This is followed by a brief overview of the research papers included in this special issue.

**What is geovisual analytics**

The term geovisual analytics has been interpreted to mean quite different things by different research groups. Most researchers agree that the “geo” component suggests that geographical data are an important focus, with one challenge being to develop methods that recognize and leverage the unique characteristics inherent in geographic data. In addition, sometimes explicitly and sometimes implicitly, there is a recognition that a focus of the analytical methods is on place and space, thus on understanding the complexity of specific localities and/or understanding patterns and relationships in geographic space and perhaps their change over time or processes that are responsible for them. Those researchers with a geographic background tend to view geovisual analytics as an approach to deal with the complexity of the world and often attempt to ground their geovisual analytics research in geographic concepts and theory (e.g., as in development of O-D maps by Wood et al., 2010, with their grounding in past geographic research on human movement, commodity flow, and related theory of spatial dynamics). Thus, the “geo” in geovisual analytics is interpreted along a continuum from meaning only that some of the data analyzed are geographical to meaning that the analysis is “geographical”, thus grounded in an appropriate geographical framework or theory that attempts to explain patterns, relationships, processes, change, etc. within and among places.

The “visual analytics” component of the term also is interpreted in different ways by different authors. The 2005 visual analytics research agenda report, Illuminating the Path (Thomas and Cook 2005), put specific emphasis on support for “analytical reasoning” as the primary distinction between previous work in information and scientific visualization and what was proposed as the challenging problems for visual analytics. One goal in doing so was to make a clear distinction between previous interactive visual and visual-computational methods that focused mostly on data exploration and generation of hypotheses and methods that supported a more comprehensive reasoning (or sensemaking) and decision-making process. The term analytical reasoning, of course, also has multiple interpretations. The most important distinction is whether the reasoning is done primarily by the human analyst (with the support of interactive visual interfaces of the kind characterized by the Jigsaw system for investigative analysis, Stasko et al., 2008), or the reasoning is primarily computational (with the interactive visual interfaces enabling control and interpretation of the computational methods, as characterized by Schreck et al., 2009, approach to analysis of trajectory data with interactive Kohonen maps). The latter perspective has precursors in earlier geovisualization research; see, for example, MacEachren et al. (1999), Andrienko and Andrienko
(1999), and Gahegan et al. (2001), as well as in earlier research strategies to enhance data mining through visualization; sees for example, Keim (1996) and Shneiderman (2001). As the field has evolved over the past five years, there has been a focus on both perspectives by different subsets of authors and the occasional effort to integrate them, e.g., the IEEE VAST Best Paper by Koch et al. (2009) focused on enabling patent search and analysis through a combination of computational reasoning and visual support for human reasoning.

The focus of the present special issue is primarily on the computational side of geovisual analytics. Papers selected for the issue (submitted in response to the call for papers that put emphasis on scaling geovisual analysis through close coupling with computational methods) address an array of complex analytical challenges with geographic data and about problems for which patterns and relations across geographic scale space are paramount.

**Improve inter-disciplinary cooperation**

Visual analytics techniques are essential for dealing with today's data sets that are growing fast in size and complexity to gain understanding, to discover patterns, and to optimize and steer complicated processes. Approaches that work either on a purely analytical or on a purely visual level cannot help due to the dynamics and complexity of the underlying processes.

In the future visual analysis techniques must satisfy a multiplicity of new requirements, which result from the rapid technological development in the hardware, software and network infrastructure. Beside high-dimensional data, continuous data streams, whose evaluation must take place immediately or in the context of given time frames, are becoming common. These developments pose high requirements on the data analysis and visualization. The graphical representation of relevant information from a large and fast growing total data volume of rising complexity makes new demands particularly on the scalability of the techniques. Scalable visual analytics systems should tightly integrate state-of-the-art automatic data analysis methods with interactive visualization techniques and be integrated smoothly into custom-designed processes for the exploration and analysis of complex information spaces.

One of the current problems of the field of visual analytics is that the automatic data analysis community (machine learning, data analysis, statistics, etc.) is not represented sufficiently in the visual analytics research community to enable a fast progress of the field. While fully automatic techniques only work if the problem is clearly specified, they are essential for the success of the field of visual analytics. The visual analytics community has to make sure that more researchers from the above domains join the community and help to develop visual analytics systems. While a number of successful applications of visual analytics have been developed over the last five years, the development of tightly integrated data analysis and visualization methods is still in the beginning and more research is needed to make progress in this respect.

Due to the high application demand, visual analytics approaches are also developed within the application areas, sometimes without involvement of our visual analytics research community. In such cases, visual analytics is embraced and considered as a part of the application area, as can be seen from the recent geovisual analytics workshops and the respective special issues of geographic research journals (see Workshop 2006, 2008, 2010, 2011). It is an open question what impact this
development will have and whether visual analytics will ultimately become part of the application domains. Details have been discussed at a VisWeek 2010 panel (May et al., 2010).

**Develop evaluation standards**

Visual analytics is a subfield of information science that draws both upon previous research on visualization and on algorithmic research in data mining and machine learning. It has the potential to truly revolutionize the way we analyze data, gain insight, draw conclusions and make decisions. The human visual system and the possibilities arising from interaction have long been known to be an extremely powerful way to deal with complex phenomena. At the same time, data analysis techniques have shown their potential of finding patterns and models in data almost fully automatically. Combining the two leads into a class of approaches that employ the best from both worlds can potentially enhance the power of each approach exponentially.

This is particularly evident whenever geographical data are to be analyzed, since these data have a natural mapping to a two-dimensional representation. Using this natural mapping as a basis, and extending it by new forms of interaction with geographical data, provides a special potential for visual analytics techniques. Geographical data are also special with respect to the multitude of changes that have affected the collection of geographical data in recent years. Increasingly, manually collected and carefully checked high quality geographical data are being augmented with automatically collected data where completeness and correctness can no longer be taken for granted, such as those collected by GPS devices or extracted from collected imagery. Here again, visual analytics offers the potential to bring the human into the loop and to use the human’s judgment capability to separate the meaningful from the insignificant, while at the same time freeing the human from tedious processing of mass data by blending in automated analysis algorithms.

As existing research, including the papers in this special issue, shows, impressive systems can be constructed based on this paradigm, and great advances have already been achieved. Nonetheless, for the future development of the discipline, the beginning efforts at establishing a paradigmatic basis need to be continued and intensified. As the basis for a discipline or subdiscipline, it is not only important to know what the goals are and upon which array of methods one wants to rely, but it is similarly important to arrive at clear standards that allow us to recognize when a certain piece of research or a certain piece of software has actually reached the goals that we are setting forward for ourselves.

Here again, of course, we can look at the evaluation standards that have been established in the two contributing domains of visualization and algorithmic data analysis. There are standards in the visualization field as to when a new visualization is to be considered a worthwhile research contribution, and when it is to be considered a practically useful tool. In algorithmic research, there are well defined and statistically well founded evaluation procedures for demonstrating that a new algorithm actually constitutes an improvement of the existing work. Visual analytics now needs to select and combine the proper elements from these two disciplines, arriving at evaluation standards that are both empirically well founded while at the same time not being so demanding that further research is stifled. Thus, in the future, visual analytics will need to produce more than technical
advances and system development; appropriately designed evaluation at the proper scale is essential to advancing visual analytics as a scientific endeavor.

**Beware privacy**

Collection and analysis of data about individuals is vital for progress in many areas such as public health, transportation, security, to name a few. Technologies enabling collection and analysis of various kinds of personal data have developed rapidly. A negative side of these developments is the growing threat to personal privacy. This particularly applies to data containing locations of people. Analysis of such data may conflict with the individual rights to prevent disclosure of the location of one’s home, workplace, activities, or trips. A number of geoinformation scientists have been working on the privacy issues associated with the use of geospatial technologies (e.g. Armstrong & Ruggles 2005, Kwan et al. 2004, Duckham et al. 2006, Cho 2008).

Intensive research on protecting personal privacy in data publishing and analysis is done in the areas of statistics and data mining, which address, among others, the problems of preserving geographical privacy. The recently completed European research project GeoPKDD (Geographic Privacy-aware Knowledge Discovery and Delivery; http://www.geopkdd.eu/) had a particular focus on data about mobility (Giannotti & Pedreschi 2007) and resulted in creation of new methods for anonymization and privacy-preserving analysis of such data. The ongoing European project MODAP (Mobility, Data Mining, and Privacy; http://www.modap.org/) is a coordination action that continues the efforts of GeoPKDD by coordinating and boosting the research activities in the intersection of mobility, data mining, and privacy preservation.

Privacy issues have not so far received close attention in the area of visual analytics research (the paper by Weaver and Gahegan 2007 could be mentioned as one of a few exceptions). While privacy is acknowledged as an issue in some papers, many researchers may be still unaware of its particular relevance to visual analytics. The following argument is meant to explain why we deem it relevant and why privacy protection needs to be considered from the visual analytics perspective in addition to the research that is done in other disciplines.

Visual analytics is about creating such working conditions in which humans and computers can utilize their inherent capabilities in the best possible ways while complementing and amplifying the capabilities of the other side. The unique capabilities of humans include

- the capability to flexibly employ previous knowledge and experience, not only those related to special education and to professional activities but also those related to the everyday life and common sense intelligence, and
- the capability to establish various associations among pieces of information.

Since these qualities are precious for analysis, visual analytics aims at enabling humans to utilize them in the most effective ways. However, the utilization of these capabilities in data analysis has the potential of increasing the threats to the privacy of individuals whose characteristics or activities are reflected in the data. In particular, this is true for data containing people’s locations.

Researchers on privacy protection in data analysis are typically concerned with the possible threats to privacy arising from computational data processing and from integration of two or more datasets.
They do not study the privacy issues arising from the involvement of human analysts empowered with interactive visual tools. It appears logical that these issues should be studied in the field of visual analytics. In particular, geovisual analytics should focus on the problems involved in analyzing georeferenced data about people.

Visual analytics can contribute to privacy protection in two ways. First, visual analytics researchers can identify what kinds of information can be extracted from various types of data by means of visually supported analysis and consider potential implications to personal privacy. These findings can be communicated to privacy protection researchers for developing methods to remove or decrease the detected privacy threats. Second, to allow humans to deal with large datasets, visual analytics researchers often employ techniques for data generalization and abstraction. Some of the techniques that are devised for the purposes of visualization can be adapted for protecting personal privacy (e.g., Maciejewski et al. 2008). Both work directions require close inter-disciplinary collaboration, as exemplified by Monreale et al. (2010).

**Give more attention to time and to users**

VisMaster (http://www.vismaster.eu) was a European coordination action with the aim to define a roadmap for the future of visual analytics research. Space and time were conceived as key topics to consider. Accordingly, a working group on space and time issues in visual analytics was formed. The VisMaster consortium collectively analyzed the state of the art, identified the key issues, challenges, and opportunities, and outlined the most important directions for the development of the visual analytics research field. The results of the work are published in the book “Mastering the Information Age. Solving Problems with Visual Analytics” (Keim et al. 2010) oriented to a broad audience. One of the book chapters is devoted to space and time. Simultaneously, the “space and time” working group has published a journal paper (Andrienko et al. 2010) addressing the research communities that deal with analysis and visualization of geographical data and phenomena: geographic information science, cartography, and geovisualization.

The book chapter and the paper draw attention to the core and unique characteristics of phenomena in space and time and the data and methods that people use to understand them. The authors argue that a synergy of visual and computational approaches is essential for spatio-temporal analysis and that there is a high demand for effective and usable geovisual analytics tools. Such tools could be beneficial not only for a small community of professional analysts but also for many citizens who need or would be interested in undertaking analysis of information in time and space. The authors also admit that time has not yet received proper attention in the research dealing with space. Therefore, they suggest that the slogan “Think temporally!” should be propagandized among researchers in geovisualization and geovisual analytics.

The challenges and opportunities identified by the “space and time” working group of VisMaster include:

- dealing with diverse spatio-temporal data (e.g. spatial time series, events, movement data, sequences of satellite images) and integrated analysis of different types of data;
- supporting analysis at multiple spatial and temporal scales and verification of discovered patterns and relationships across scales;
• understanding and adequately supporting diverse users by improving the understanding of human perceptual and cognitive processes and developing effective solutions for different user categories;
• reaching the broad community of potential users by developing lightweight, easily deployable and usable software that allows flexible customization and combination of tools.

In this special issue, two of the four research papers deal with time-dependent phenomena and, consequently, time-referenced data (Maciejewski et al., Schumann & Tominski). The paper by Schumann & Tominski deals with complex data structures including spatial, temporal, and categorical hierarchies, which have a relation to the issues of spatial and temporal scales. Regrettfully, the user-related challenges were not addressed in any of the submitted papers, perhaps, due to the focus of the call and the journal outlet. This does not mean that there is no research in geovisual analytics related to the user issues; thus, VAST 2010 had multiple papers with a focus on users, and some of them dealt with users analyzing geographical information. Still, it is important to attract the attention of a larger number of geovisual researchers to these issues.

Research papers of this issue
The four papers published in this special issue have been selected from seven submissions. The papers exemplify different research problems dealt with in the area of geovisual analytics and different approaches to solving them.

One of the problems is analysis of multivariate geographically referenced data. To enable the understanding of relationships between geographic space and multi-dimensional attribute space, Skupin and Esperbé have developed an approach based on the use of the self-organizing map (SOM), a computational technique that combines clustering with dimensionality reduction. The authors describe a methodology for obtaining high-resolution SOMs and suggest innovative techniques for linking the representations of the geographic and attribute spaces. The work of the approach is demonstrated by analyzing a dataset in which more than 200,000 U.S. census block groups are characterized in terms of 69 attributes. The authors have been able to analyze the attributes much more comprehensively than it was possible with the previous approaches and to arrive at a holistic regionalization in the geographic space, i.e., division into regions that are internally homogeneous in terms of the multiple attributes.

Schumann and Tominski consider complex data structures involving temporal, spatial, and categorical hierarchies. As a basis for visualizing such data, the authors extend the established concept of focus + context by distinguishing between map focus and context, on the one side, and data focus and context, on the other side. They suggest a strategy for building data representations depending on the current map and data foci and contexts and show how this strategy applies to representing multi-hierarchical data on maps. The strategy implies building data abstractions, which can be achieved by means of computational techniques. The paper discusses one of the possible data abstractions that can be obtained using the method of association rules from data mining and the ways of visualizing its results on maps.

The remaining two papers are about visual analytics support to decision making. Guo and Jin deal with the problem of optimal division of a territory into geographically contiguous districts with
certain desired properties such as equal population and compact shapes. Human judgment is needed to balance multiple conflicting criteria and to account for vaguely defined or tacit criteria, which cannot be used in an automatic optimization algorithm. The authors present an approach in which computational optimization is controlled by the user in an interactive and iterative process. As a result, not only user’s preferences can be properly taken into account but also the quality of the outcomes is improved.

Maciejewski et al. consider a decision problem that is not only spatial but also temporal. Preparedness of public health organizations to a potential pandemic requires them to understand how the disease may spread in space over time and how this process may be affected by possible countermeasures. Simulation models are employed to obtain plausible scenarios of the pandemic spread under different conditions. The authors have developed a visual analytics toolkit that allows public health officials not only to observe the evolution of the pandemic and its effects on the population but also to investigate the likely impacts of possible countermeasures (such as social distancing or vaccination) taken at different time moments. The tools can be used for developing plans for pandemic response as well as for education purposes.

This sample of papers indicates the breadth of the scope of the geovisual analytics research. Quite logically, all researchers deal with data having a spatial (geographic) component, but there is a great variety of types and structures of such data. All visual analytics approaches, including those presented in this issue, combine interactive visualization with computational data processing. On the visualization side, the use of cartographic maps is indispensable for enabling human analysts to deal with spatial information (Andrienko et al. 2007, 2008). However, the power of maps is still limited, especially in representing multi-attribute and dynamic data. To compensate for the limitations, researchers combine maps with other visual representations, as can be seen in the papers of this issue. On the computational side, a large array of techniques is used including clustering, projection, aggregation, pattern extraction, optimization, and simulation. The analytical processes geovisual analytics strives to support range from exploration of data and general comprehension of the underlying phenomena to deep penetration into specific aspects and features and to deriving plans for practical actions. The internal diversity of the geovisual analytics research field is the consequence of the variety and complexity of the space-related phenomena and problems.

Acknowledgment
We are grateful to Heidrun Schumann, Christian Tominski, and André Skupin for providing valuable comments and suggestions for improvement. We thank the authors of all submitted papers and reviewers for their work on insuring the scientific quality of the special issue.

References


Keim, D., and Kriegel, H.-P. (1996): Visualization techniques for mining large databases: a comparison. IEEE Transactions on Knowledge and Data Engineering (Special Issue on Data Mining) 8, 923-938.


Workshop “Visualization, Analytics & Spatial Decision Support” at the GIScience 2006 conference, September 20-23, 2006, Münster, Germany


Workshop “GeoVA(t) - Geospatial Visual Analytics: Focus on Time” at the AGILE 2010 conference, May 10-11, 2010, Guimarães, Portugal

