Visualizing Multidimensional and Temporal Data at the Good Scale by Designing and Refining

Charles Perin

Visualizing Multidimensional and Temporal Data at the Good Scale by Designing and Refining

Charles Perin
INRIA, Univ. Paris-Sud, CNRS-LIMSI
Bat. 660, Univ. Paris-Sud
91405, Orsay, France
charles.perin@inria.fr

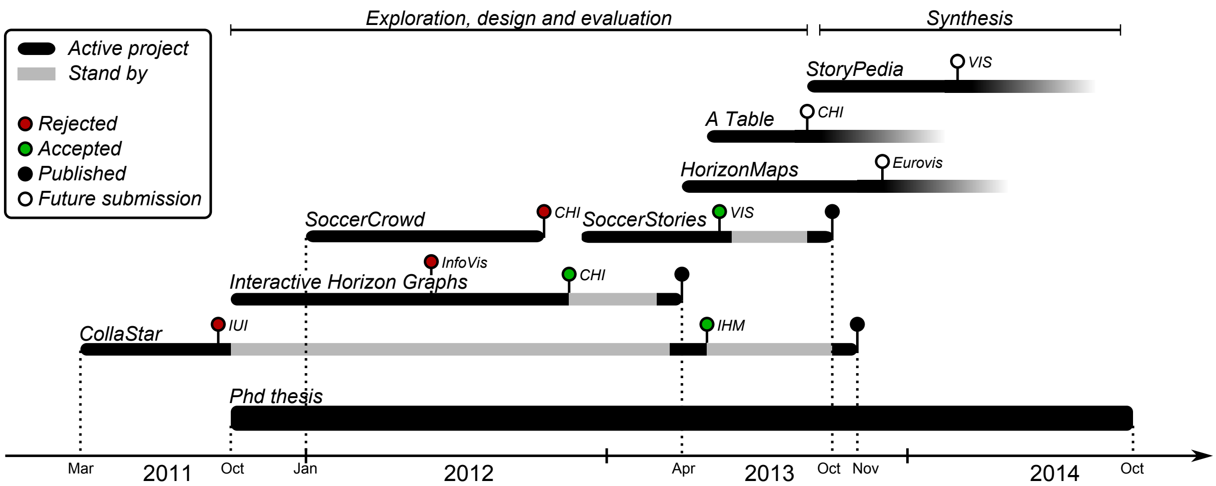

Figure 1. Timeline of the thesis.

RÉSUMÉ

This doctoral research is focused on the visualization and interaction with multidimensional and temporal data. We tackle the challenges of: 1) designing visualizations for different users' expertise and needs; and 2) refining existing visualizations to enhance their usability and genericness or make them best-suited to a specific domain, need, and user. In this research overview we state the research problem and our plan of research. We then describe the achieved projects, each introduced with its closest related works. Finally, we present our future work perspectives.

INTRODUCTION

Sets of quantitative values changing over time (*i. e.* time series) are predominant in a wide range of domains such as finance (*e. g.,* stock prices), sciences (*e. g.,* climate measurements, network logs), and popular media (*e. g.,* election polls, live sport events). With the variety of domains comes the variety of audiences, and of user's expertises. Because time series are almost always multidimensional, there is a need of going beyond the simplest ways of representing time series (*e. g.,* line charts) as well as finding the optimal scale at which the data should be visualized.

In this thesis, we tackle the challenge of proposing interactive visualization techniques for different domains and audiences where multidimensional and temporal data are prominent. Applying an existing visualization technique to a particular domain is everything but trivial; in particular, the scale at which the data are visualized is highly dependant of the task and the data. Although many visualization techniques exist and are designed every year, a few are really used. Our approach is then to identify pertinent visualization techniques for a particular domain or task and design a tool which uses one or several of these identified visualization. Instead of inventing more and more new visualization techniques, we first try to improve, refine and adapt the existing ones to show their usefulness in real-life applications. The choice of a visualization technique is for example based on the existing data for the domain, the user's needs, and the user's expertise.

METHODOLOGY AND PLAN OF RESEARCH

Our goal is to refine existing visualization techniques instead of inventing a new visualization each time a new problem is challenged. The timeline for each of our projects is similar: we first analyze the workflow of specific users dealing with multidimensional and temporal data and collect their needs. Then, we gather existing visualization techniques for such data and try to solve an identified problem by improving a technique (*e. g.,* by

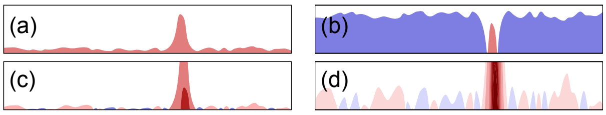

Figure 2. Four views of a time series illustrating the importance of the interactive settings of the baseline value and the zoom factor.

adding interactions). If no visualization technique is well-suited enough we design a new tool dedicated to a specific domain and specific tasks. This process involves collecting data, designing visualizations and interactions and evaluating the new tool. Depending on the domain, a massive part of the time might be allocated to discussions with users in order to identify their needs and the type of tool they may or may not use. For each project, the scale at which visualizing the data may or at which performing the task efficiently differs, and this must be considered in the process of designing and/or improving the visualization.

DESCRIPTION OF THE PROGRESS TO DATE

We describe the achieved projects of this thesis following our methodology: we first make a review of existing visualization techniques for specific data, domain and users; then we select one or several candidate(s) to improve; and finally refine the selected candidates and evaluate them.

Interactive Horizon Graphs

The first project of this thesis is Interactive horizon Graphs [15] (*IHG*), published at CHI'13.

Finance Data Users. The daily task of financial analysts or traders consists of observing, analyzing and making decisions on large numbers of time series.

Visualization Techniques. Since line charts have become widespread [19], many design exist for displaying data in the form of charts (*e. g.,* [1, 2, 23]) and for the comparison of graphical visualization techniques (*e. g.,* [17, 22]). However, using line charts to visualize multiple time series can be difficult because the limited vertical screen resolution can result in high visual clutter. Javed et al. classified visualization techniques for multiple time series into two categories [6]. In *shared-space* techniques, time series are overlaid in the same space and such techniques can support only a limited number of time series. In *split-space* techniques, the space is divided by the number of time series and each one occupies its own reduced space (*e. g., Reduced Line Charts* (*RLC*) [23], *Horizon Graphs* (*HG*) [5, 20]). Such techniques can support large numbers of time series, but need more screen real estate.

Upgrading Horizon Graphs. IHG are inspired by pan and zoom techniques and unify RLC and HG, two of the most effective techniques for visualizing multiple time series. We designed IHG to increase the number of time series one can monitor and explore efficiently by focusing on evaluating the benefits of adding interaction to an existing technique. Figure 2 illustrates how interactively setting the parameters impacts the efficiency of the visualization.

Results. We evaluated the benefits of IHG for standard tasks on time series visualizations in a controlled experiment and found that IHG outperform both RLC and HG for large numbers of time series and difficult tasks. This work showed that simple interactions can unify two visualization techniques and substantially improve their efficiency, illustrating that upgrading visualization techniques is certainly as much important as inventing new ones.

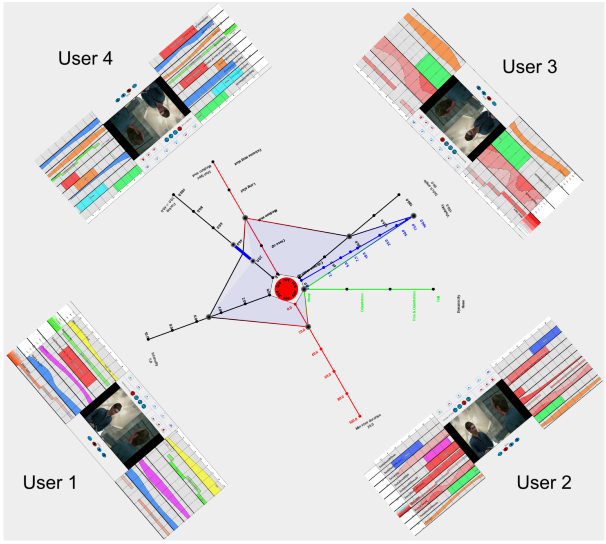

Figure 3. *CollaStar*: the collaborative star glyph in its center and several personalized Linear Walls all around.

CollaStar

The second project of this thesis is *CollaStar* [14] (Figure 3), to appear in IHM'13. It consists of designing an interface to control the CineSys software [9] whose role is to render a 2D movie from a 3D scene using a set of cinematography rules. A 2D image of a movie is a multidimensional data whose dimensions are for example the camera position, the lighting settings, relationships between actors and artistic rules.

Users. The users are movie creators. By movie we mean any 2D movie created from a 3D scene (*e. g.,* machinima, animation). We identified that these users 1) have a clear need of simpler tools than the existing ones (*e. g.,* 3DS Max, Autodesk Maya); and 2) have various expertises and need to work in collaboration.

Visualization techniques. After a thorough review of visualization techniques for multidimensional and temporal data, we chose 1) to handle the multidimensional data in real time using a star glyph [4, 7], and 2) tu use what we call a *Linear Wall* for real time monitoring of the temporal dimensions, inspired from DIVA [10] and the perspective wall [11]. Indeed, two different scales are needed to perform the identified task: a scale at which a specific time only is visualized (the current time) using the star glyph; and a range-scale of a short but zoomable interval of time to monitor the creative process using the Linear Wall.

Upgrading the star glyph. We chose the star glyph as the collaborative interface because it is well-suited to the type of data to manipulate. Contrary to parallel coordinates, it allows users to be all around the same multi-touch table and manipulate different branches of the same star glyph, while observing what their co-workers do. The star Glyph is made interactive by providing interactive star branches. Each branch is a slider with one or several thumbs (dis-

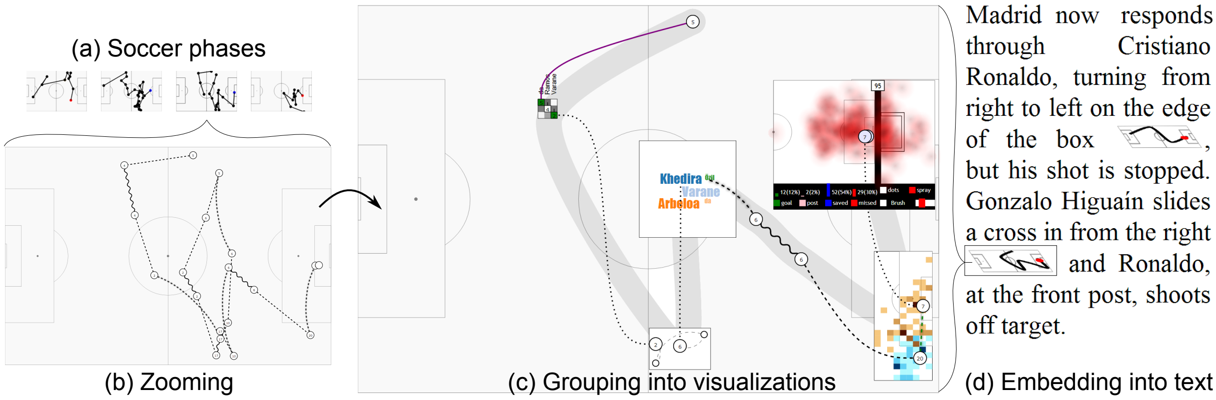

Figure 4. Using SoccerStories: (a) navigating among soccer *phases* of a game; (b) mapping a phase on a focus soccer field; (c) exploring the phase by grouping actions into tailored visualizations; and (d) communicating using *SportLines* embed into text.

crete or continuous). Branches can be added, removed, or moved around the center according to the users' needs.

Challenge. The main challenge was to design a tool to communicate with a specific system: CineSys. We had to translate the cinematography rules into high level controllers (star branches) and handle the collaborative interaction, involving considering dependencies between controllers. The technical part was also challenging. Beyond developing a multi-user interface on a multi-touch table, we had to make a bidirectional bridge for CollaStar (Java) and CineSys (C++) to communicate: 1) CollaStar sends the high level parameters values in real time to CineSys; 2) CineSys computes the associated style and applies it to the 3D scene; 3) CineSys sends the current frame to every linear Wall of the interface as the focus view.

Results. The main result from the discussions we had with cinematography experts is that CollaStar+CineSys allows a quick prototyping of 2D movies from 3D scenes. Indeed, it is far less precise than the existing professional tools, which means far less tedious to learn. They saw in the system an alternative to *story boards*, which are used for quickly prototyping movies. The strength of our tool is that it allows: prototyping in real time (WISIWYG); users with various expertise working in collaboration; and post-editing if needed, using the Linear Wall edition mode.

SoccerStories

The third published project of this thesis is *SoccerStories* [16] (Figure 4), a visualization interface to support analysts in exploring soccer data and communicating insights. The core idea consists of extracting *phases* of interest from soccer games to analyze and communicate. Phases are sequences of events leading to an event of interest, and we discovered it is an appropriate scale to visualize the data which has not been explored yet.

Soccer analysts. We worked with a data expert (data provider, from Opta [12]) to get an access to the company's data and identify the needs of their customers. We worked with two journalists and a sports trainer to understand their workflow and needs. Currently, most analyses on soccer data relate to statistics on individual players or teams. Analysts we collaborated with consider that quantitative analysis alone does not convey the right picture of the game, as context, player positions and phases of player actions are the most relevant aspects.

Visualization techniques. We designed SoccerStories to support the current practice of soccer analysts and enrich it, both in the analysis and communication stages. Our system provides an overview+detail interface of game phases, and their aggregation into a series of connected visualizations. Each visualization, called a *faceted view*, is tailored for actions such as a series of passes or a goal attempt. Figure 5 illustrate the set of faceted views we extracted from the literature and implemented. The main difficulty is due to the lack of academic research for visualizing sport events (examples are [3, 8, 13, 18, 21]); thus we had to consider web-based visualizations to extract best practices for sport analysis and communication.

Results. We evaluate our tool with two qualitative user studies, with data from one of the world's leading live sports data providers. The first study resulted in a series of four articles on soccer tactics, by a soccer analyst, who said he would not have been able to write these otherwise. The second study consisted in an exploratory follow-up to investigate design alternatives for embedding soccer phases into word-sized graphics. For both experiments, we received a very enthusiastic feedback and participants consider further use of SoccerStories to enhance their current workflow. We conclude the following. 1) Phases are adequate for soccer exploration, well understood and easy to extract. 2) The soccer field is the workspace for understanding the phases and for communication. 3) Using advanced visualizations in a domain such as soccer requires particular care although their standardization is mandatory to increase the visualization literacy of both sport analysts and their audience. 4) Spatio-temporal thumbnails or fingerprints are a promising alternative for timelines, but their design still needs further investigation.

We are particularly proud of this project because it is certainly our biggest achievement. Indeed, although it was accepted for its first submission, this article is the outcome of a two years project. It is mature because we tried many directions, encountered many obstacles, and finally converge to SoccerStories.

FUTURE WORK

In present and future work, we plan to keep the same methodology. We already started some projects, others are

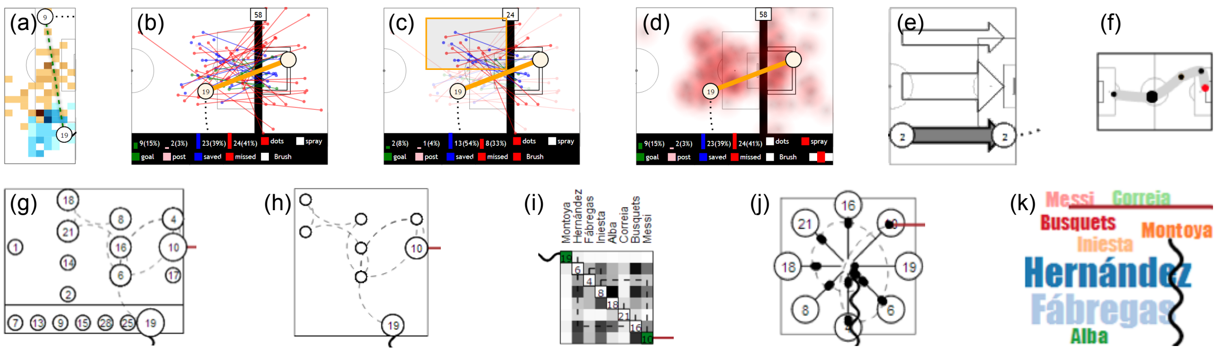

Figure 5. Faceted views for soccer: (a) corner kicks and crosses, (b-d) shots, (e) long runs, (f) global flow, (g-k) different facets for pass clusters.

planned, and some will remain side projects (Figure 1). We consider exploring ways of visualizing *very large* time series. By very large, we mean time series that can't be stored in memory or visualized on a screen and need for example a database access, aggregation strategies, and multi-resolution visualization techniques. We also want to do the future work we planned in the previous projects. We will keep working on sport data and investigate the case of fingerprints for small spatio-temporal time series.

One of the main challenge is to determine for each domain, user, and data, the good scale to visualize and explore the data. For example, a project on soccer may have various scales: single game, season, history of a championship; and can be about teams, players or competitions. The data to consider and the associated visualizations are highly dependent on this scale, which has to be determined with users. The last six to eight months will be devoted to the synthesis and the writing of the thesis, in parallel with publishing the last projects, attending to conferences and looking for a post-doctoral position.

CONCLUSION

With the variety of projects fitting into this research topic, I took the opportunity to investigate a broad range of domains, the diversity of the projects leading to difficulties. Indeed, I am not sure yet what the final message of my thesis is and I expect to clarify this during the doctoral meeting. If some Phd students have a straightforward thesis due to a very specific Phd topic, mine is broad and blurry and I sometimes feel like making scrapbooking.

On the other hand, each project having its own specificities, I learnt many different things, from system architecture and programming language to user's needs and domain-specific visualization literacy issues. By avoiding focusing on a specific topic, I have been able to go back and forth, switching from one project, to another. It resulted in finding relationships between projects, reuse knowledge across them, and have a broader overview of the domain. I intend to synthesize this knowledge in practical guidelines for the design of multidimensional and temporal visualization tools in a generic pipeline generalizing several heterogeneous successful projects.

BIBLIOGRAPHIE

1. J. Bertin. *Semiology of graphics*. University of Wisconsin Press, 1983.
2. W. S. Cleveland. *The elements of graphing data*. Wadsworth Publ. Co., Belmont, CA, USA, 1985.
3. A. Cox and J. Stasko. Sportvis: Discovering meaning in sports statistics through information visualization, 2006.
4. E. Fanea, S. Carpendale, and T. Isenberg. An interactive 3d integration of parallel coordinates and star glyphs. In *TVCG*, pages 20–, 2005.
5. S. Few. Time on the horizon. available online at `http://www.perceptualedge.com/articles/visual_business_intelligence/time_on_the_horizon.pdf`, 2008.
6. W. Javed, B. McDonnel, and N. Elmqvist. Graphical perception of multiple time series. *TVCG*, 16(6), 2010.
7. A. Klippel, F. Hardisty, R. Li, and C. Weaver. Colour-enhanced star plot glyphs: Can salient shape characteristics be overcome? In *Cartographica, volume 44*, pages 217–231, 2009.
8. P. Legg, D. Chung, M. Parry, M. Jones, R. Long, I. Griffiths, and M. Chen. Matchpad: Interactive glyph-based visualization for real-time sports performance analysis. In *Computer Graphics Forum*, volume 31, pages 1255–1264, 2012.
9. C. Lino, M. Christie, F. Lamarche, G. Schofield, and P. Olivier. A real-time cinematography system for interactive 3d environments. In *Proc. SCA'10*, pages 139–148, 2010.
10. W. Mackay and M. Beaudouin-Lafon. Diva: Exploratory data analysis with multimedia streams. In *Proc. CHI'98*, pages 416–423, 1998.
11. J. D. Mackinlay, G. G. Robertson, and S. K. Card. The perspective wall: detail and context smoothly integrated. In *Proc. CHI'91*, pages 173–176, 1991.
12. Opta. March 2013. `http://www.optasports.com/`.
13. M. Page and A. V. Moere. Towards classifying visualization in team sports. In *CGIV'06*, pages 24–29, 2006.
14. C. Perin, M. Christie, F. Vernier, and C. Lino. Collastar: Interaction collaborative avec des données multidimensionnelles et temporelles. In *TO APPEAR IN IHM'13*, 2013.
15. C. Perin, F. Vernier, and J.-D. Fekete. Interactive horizon graphs: improving the compact visualization of multiple time series. In *Proc. CHI '13*, pages 3217–3226, 2013.
16. C. Perin, R. Vuillemot, and J.-D. Fekete. SoccerStories: A Kick-off for Visual Soccer Analysis. *TVCG*, Oct. 2013.
17. L. Peterson and W. Schramm. How accurately are different kinds of graphs read? *Educational Technology Research and Development*, 2:178–189, 1954.
18. H. Pileggi, C. Stolper, J. Boyle, and J. Stasko. Snapshot: Visualization to propel ice hockey analytics. *TVCG*, 18(12):2819–2828, 2012.
19. W. Playfair. *The Commercial and Political Atlas*. London, 1786.
20. H. Reijner. The development of the horizon graph. available online at `http://www.stonesc.com/Vis08_Workshop/DVD/Reijner_submission.pdf`, 2008.
21. A. Rusu, D. Stoica, and E. Burns. Analyzing soccer goalkeeper performance using a metaphor-based visualization. In *IV'11*, pages 194–199, 2011.
22. D. Simkin and R. Hastie. An Information-Processing Analysis of Graph Perception. *Journal of the American Statistical Association*, 82(398), 1987.
23. E. R. Tufte. *The visual display of quantitative information*. Graphics Press, USA, 1986.