

Grand Challenge Award 2008: Support for Diverse Analytic Techniques - nSpace2 and GeoTime Visual Analytics

Lynn Chien, Annie Tat, Pascale Proulx, Adeel Khamisa, William Wright*

Oculus Info Inc.

ABSTRACT

GeoTime and nSpace2 are interactive visual analytics tools that were used to examine and interpret all four of the 2008 VAST Challenge datasets. GeoTime excels in visualizing event patterns in time and space, or in time and any abstract landscape, while nSpace2 is a web-based analytical tool designed to support every step of the analytical process. nSpace2 is an integrating analytic environment. This paper highlights the VAST analytical experience with these tools that contributed to the success of these tools and this team for the third consecutive year.

CR Categories: H.5.2 [Information Interfaces & Presentations]: User Interfaces – Graphical User Interfaces (GUI); I.3.6 [Methodology and Techniques]: Interaction Techniques.

Additional Keywords: visual analytics, human information interaction, geo-spatial information systems, temporal analysis.

1 OVERVIEW

GeoTime and nSpace2 are analytical visualization applications that facilitate massive data analysis. Both applications were developed in collaboration with analysts and are being used and evaluated on a day-to-day basis. The tools were used for first analyzing all four Mini Challenges and then the Grand Challenge.

1.1 nSpace2

nSpace2 is the web version of nSpace, an environment supporting the whole analytical workflow. It has two main components: TRIST, focused on information triage, and the Sandbox [3], for evidence marshaling and analytical sense-making. nSpace2 (still in beta) is revolutionary in terms of what it can do on the web; however, it currently has only a subset of the capabilities that its parent nSpace has. Nevertheless, its core initial capabilities and in particular its strength in supporting multiple analysts on related projects were definitively key to this team's analytical process.

1.2 GeoTime

GeoTime supports the visualization and analysis of entities and events over time and geography within a fused 3D space. GeoTime supports the extension Configurable Spaces [2] that extends GeoTime's X, Y, T coordinate space for temporal analysis to any arbitrary diagrammatic workspace by replacing a geographic map with a diagram.

2 ANALYTICAL PROCESS

2.1 Sharing Insights from Mini Challenges

Two junior and two experienced analysts worked on the VAST 2008 Challenge. Analysts working on the Mini Challenges summarized and shared all of their material using nSpace2, making it available to the analyst working on the Grand Challenge.

* email: {lynn.chien, annie.tat, pascale.proulx, adeel.khamisa, bill.wright}@oculusinfo.com

nSpace2 allows analysts to share TRIST and Sandbox files in a web environment. Multiple Sandboxes can be made and opened in different browser windows so the analyst can interact with different facets of information at the same time, as shown in Figure 2. The Sandbox is an integrating analytic tool. Results from each of the quite different mini-challenges were combined in the Sandbox environment.

The Sandbox is a flexible and expressive thinking environment where ideas are unrestricted and thoughts can flow freely and be recorded by pointing and typing anywhere [3]. The Pasteboard sits at the bottom panel of the browser and relevant information such as entities, evidence and an analyst's hypotheses can be copied to it and then transferred to other Sandboxes to be assembled according to a different perspective for example.



Figure 1. Using two different browser windows to analyse different aspects of the analysis. Results from different analytic tasks can be combined in one or more Sandboxes.

2.2 Assembling and Making Sense of Gathered Data

The Sandbox supports put-this-there cognition, allowing the analyst to re-organize and rearrange gathered data in a way that makes sense to her. A minimum of buttons and no dialog boxes or forms are used to create, edit, place, arrange, re-arrange, group, emphasize, highlight and compare [3], which makes it easier to stay focused on the analysis.

Once the Grand Challenge analyst finished looking through the analyses for the Mini Challenges, she created new Sandboxes corresponding to how she wanted to combine the evidence and conclusions from the various sources. She ended up with the following Sandboxes: Key Players and their Relationships, Key Players' Activities and Timeline of Key Events. Using the nSpace web pasteboard, she copied and transferred all relevant data into her new Sandboxes. The analyst re-organized and rearranged this saved data using various methods provided by the Sandbox [3].

2.3 Marshaling Evidence and Hypothesis Generation

The Sandbox supports analysts in the development and assessment of meaningful hypotheses, which are captured as

assertions. Evidence gathered from other Sandboxes, including observations, notes, events, ideas, and concepts, can be dragged into assertions from the right as supporting or from the left as refuting evidence for these hypotheses. The weight of the evidence can be changed or adjusted when needed.

The analyst hypothesized many possibilities of how the challenges may link together and placed them as assertions in yet another Sandbox in a new browser window. Using the evidence and notes she gathered from the mini-challenges, she evaluated her hypotheses and then brainstormed ideas for how to prove them.

2.4 Cross-Referencing by Layering the Datasets

GeoTime helps the analyst see patterns of events, relationships, and interactions over time within a geospatial (or any conceptual or diagrammatic) context [1] and was used to analyze all of the mini challenges. Users can import multiple datasets in different layers so information can be combined or filtered depending on the needs of the analysis so commonalities across datasets can be detected easily.

The analyst began to connect the Mini Challenges by combining the migrant data in Challenge 2 and the cell phone data in Challenge 3. One of the patterns discovered was that the calls by the entities of interest are concentrated in the network towers closest to the three major points from which the migrations were being launched. That observation was captured in a GeoTime report that was imported into the Sandbox to be used as further evidence in the various hypotheses.

2.5 Leveraging the Synergy between GeoTime and nSpace2

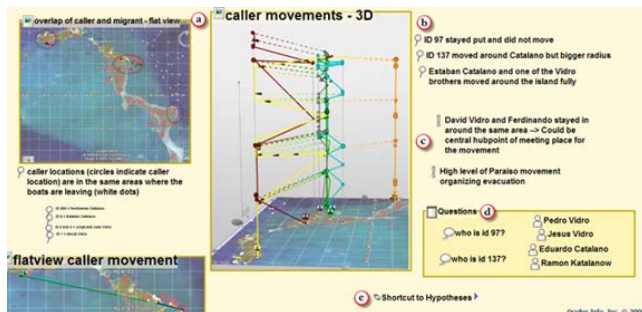


Figure 2. Portion of a Sandbox used to tie two different observations from two different Challenges to find new observations.

GeoTime and nSpace2 provide complementary support for analysis. nSpace2 supports uploading images and files directly from the desktop and important documents and findings made from any application can be easily uploaded and used to support current investigations and for further analysis. Work can be shared by multiple analysts in real-time so new observations can be timely passed to another analyst for verification or support.

Observations and hypotheses generated from analysis conducted in GeoTime for the Mini Challenges were captured in GeoTime reports, which are 2D annotated snapshots that were augmented and combined with additional evidence and placed into assertions in Sandboxes. The analyst working on the Grand Challenge used these observations as evidence to help develop the initial hypotheses, and with her additional finding from layering the datasets, was able to prove that the migration and cell data are

indeed linked. That led her to form the new hypothesis that the high-level Paraiso members are organizing the illegal migrations out of the island, as seen in Figure 2.

2.6 Creating Social Networks and Structuring Information Hierarchically

Information in the Sandbox is represented by icons and text descriptions. Links can be drawn and labelled between objects to represent their relationships so that a social network can be easily formed. Information can also be organized in a hierarchical manner. By using Shortcuts, multiple Sandboxes can be hyperlinked so that information is displayed in consumable segments. When analysts need to drill down on a point, they can use a Shortcut to open a Sandbox that contains the details.

When the analyst began analyzing the main entities of the challenges and their relationships, a social network diagram emerged showing strong connections between the Vidro and Catalano family, and fewer connections with the Tafoya family, allowing an assertion to be made that there are three major families in the movement. The analyst did not provide a detailed profile for each character but made shortcuts to a Sandbox that contains their profiles in case users wanted to know more about the individuals.

3 LESSONS LEARNED

Analysts were able to use both nSpace2 and GeoTime hand-in-hand to perform the analysis for the VAST challenges. GeoTime was more effective for the data distributed by the VAST contest committee this year. The tool easily compiled and visualized the migrant data in Challenge # 2 and the RFID data of Challenge #4. The GeoTime extension, Configurable Spaces, was effectively utilized for detecting patterns and factions in Challenge #1 and #3. Capturing the results of the analysis is not as strong in GeoTime as in nSpace, however, but that part of the workflow was easily pursued in nSpace2. Currently the reporting aspects are being enhanced for GeoTime, such as implementing the ability to construct narratives and stories [4]. The browser-based nSpace2 proved to be useful in a team-oriented analytical environment as analysts were able to share and edit the analyses as they worked. The data was not as a good fit for the type of information triage nSpace is good at and many aspects of the analysis require manual layout, which has advantages, but could be made more efficient. Automatic ways to extract social networks and work with geographical data are currently under development for nSpace2.

4 CONCLUSION

nSpace2 and GeoTime proved to be powerful systems that enabled collaboration of multiple analysts to discover more insights in a variety of analytical problems. The 2008 VAST Challenge helped us test the new nSpace2 technology against real-world data and allowed us to find ways to better enhance collaboration across multiple analysts and applications.

REFERENCES

- [1] Kapler, Thomas, and William Wright, GeoTime Information Visualization, IEEE Symposium on Information Visualization, 2004.
- [2] Kapler, T., R. Eccles, R. Harper, W. Wright, Configurable Spaces: Temporal Analysis in Diagrammatic Contexts, paper accepted for IEEE VAST 2008.
- [3] Wright, William, D. Schroh, P. Proulx, A. Skaburskis and B. Cort, The Sandbox for Analysis - Concepts and Methods, paper accepted for ACM CHI 2006.
- [4] Eccles, R., Kapler, T., Harper, R., Wright, W., Stories in GeoTime. Information Visualization 7, 3-17, 2007.