Spatiotemporal Anomaly Visualization for Large-scale Dynamic Networks

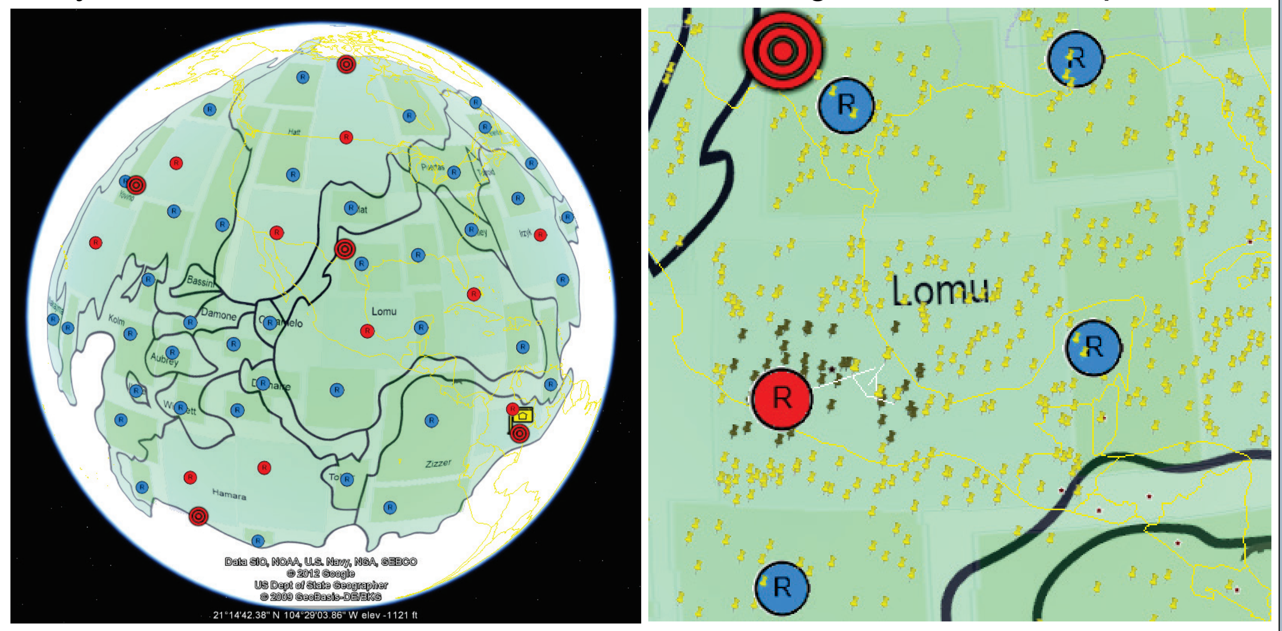
The 20th Annual Student Research & Creative Endeavors Exhibition (SRCEE), Central Michigan University, April 17, 2013

Tao Zhang, Dr. Qi Liao, Department of Computer Science, Central Michigan University

Problem

As we move to the internet of things (IoT), many devices (smart phones and sensors) may report data back with the most recent value at specific location and time. Even stationary spots like physical computer servers and data centers may log their performance data (e.g., number of traffic flow, system load, etc.) with time/location information.

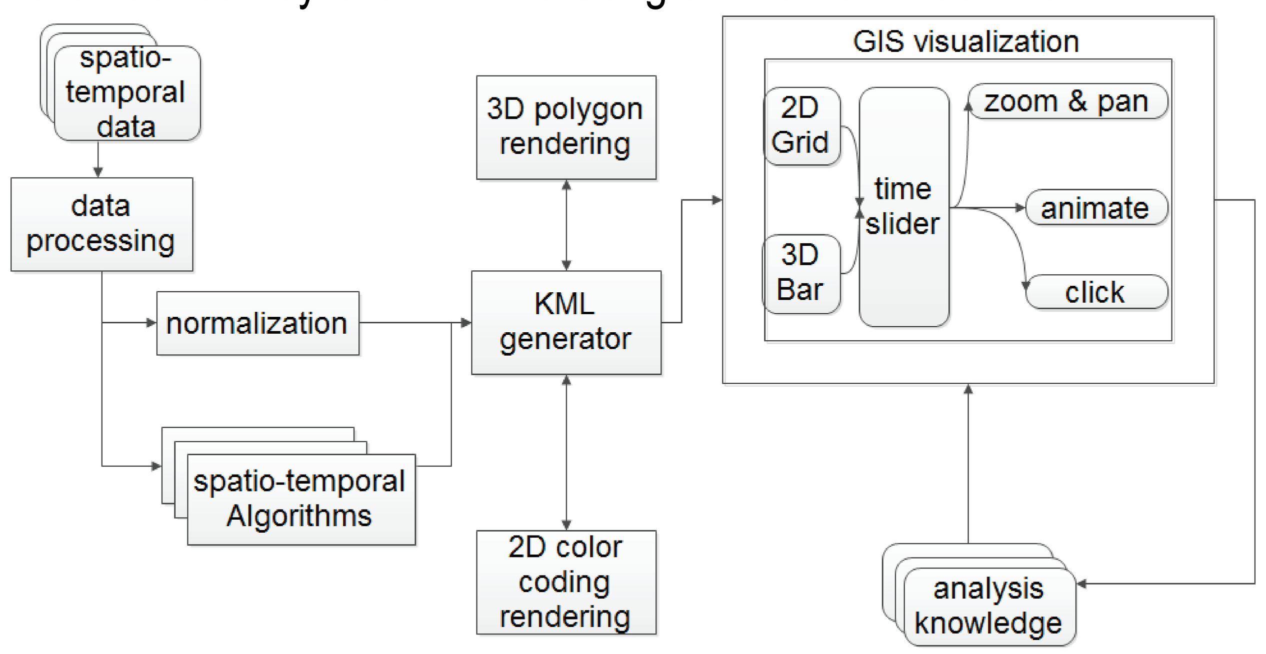
Monitoring and understanding these spatiotemporal data[1] is challenging because not only the data grows much larger in size but also more complex in nature. This is further complicated by the fact that the data values are usually very dynamic and they change not only across different areas but over time as well. It is difficult for humans to understand the dynamics and correlation of events among the time and space.



VC12 MC1 Data Example by Google Earth

To that end, we developed a visual analytic system that allows users or domain experts to interactively explore spatiotemporal datasets and their anomalous changes. Our system is built on top of one popular geographic information system (GIS), i.e., Google Earth (GE) and utilize a generic data format, i.e., Keyhole Markup Language (KML).

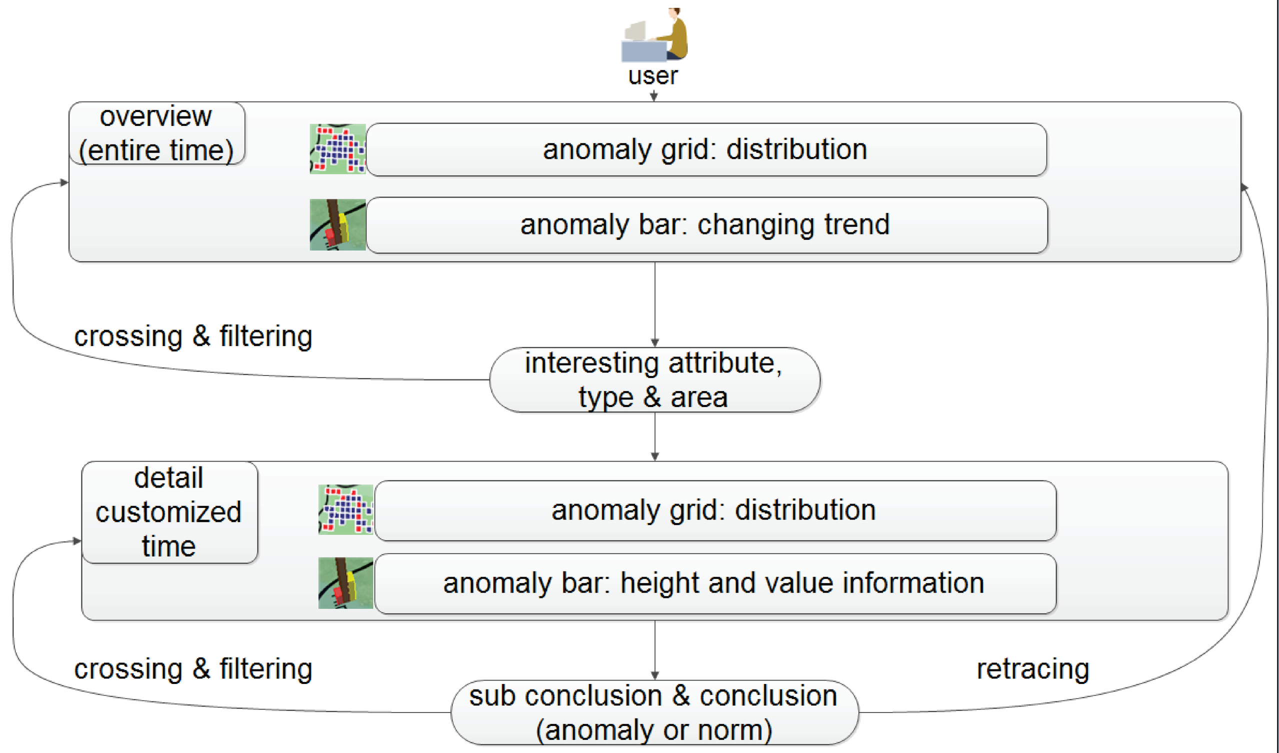
In our system, data is handled with various processing methods such as scanning, aggregation, and normalization. The result will be parsed by KML generator, which writes all parsed data into files with KML format. 2D and 3D rendering solution will be used separately to deal with grids and bars information result together with attached timestamp. The processed KML files will then be read by GIS such as Google Earth for visualization.



System architecture on spatiotemporal data process, analysis and visualization.

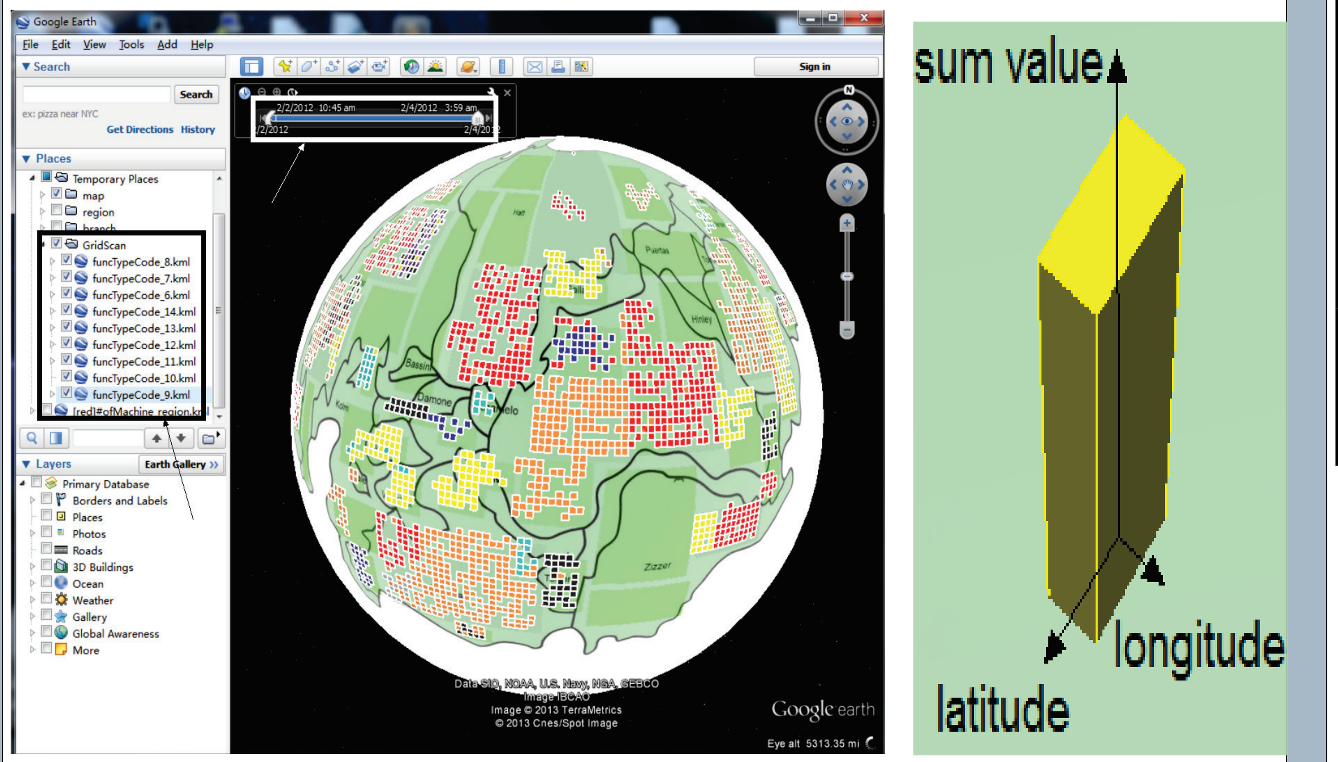
Data Analysis and Visualization

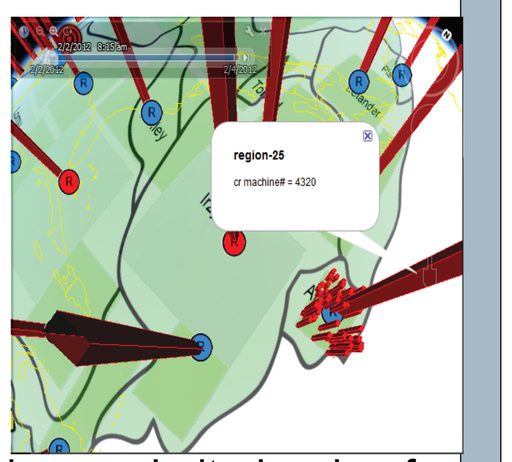
In the visualization part, that repetition of analyzing between two main levels is available. Analyst could find and study all interested areas by switching the visualization scale. Investigators could make conclusions by clearly observing from both static big pictures and dynamic moving trends. The amount of details are dynamically adjusted based on the zoom levels and time window selection.



The general procedure and flow chart of detecting and analyzing the spatiotemporal anomalies using the visualization interface.

We use a few simple yet effective techniques, i.e., 3D bars for representing the value dynamics. By combining data mining algorithms such as GridScan[2] as an overlay 2D grids on top of map, it is possible that the tool may guide users effectively to find potential anomalies.



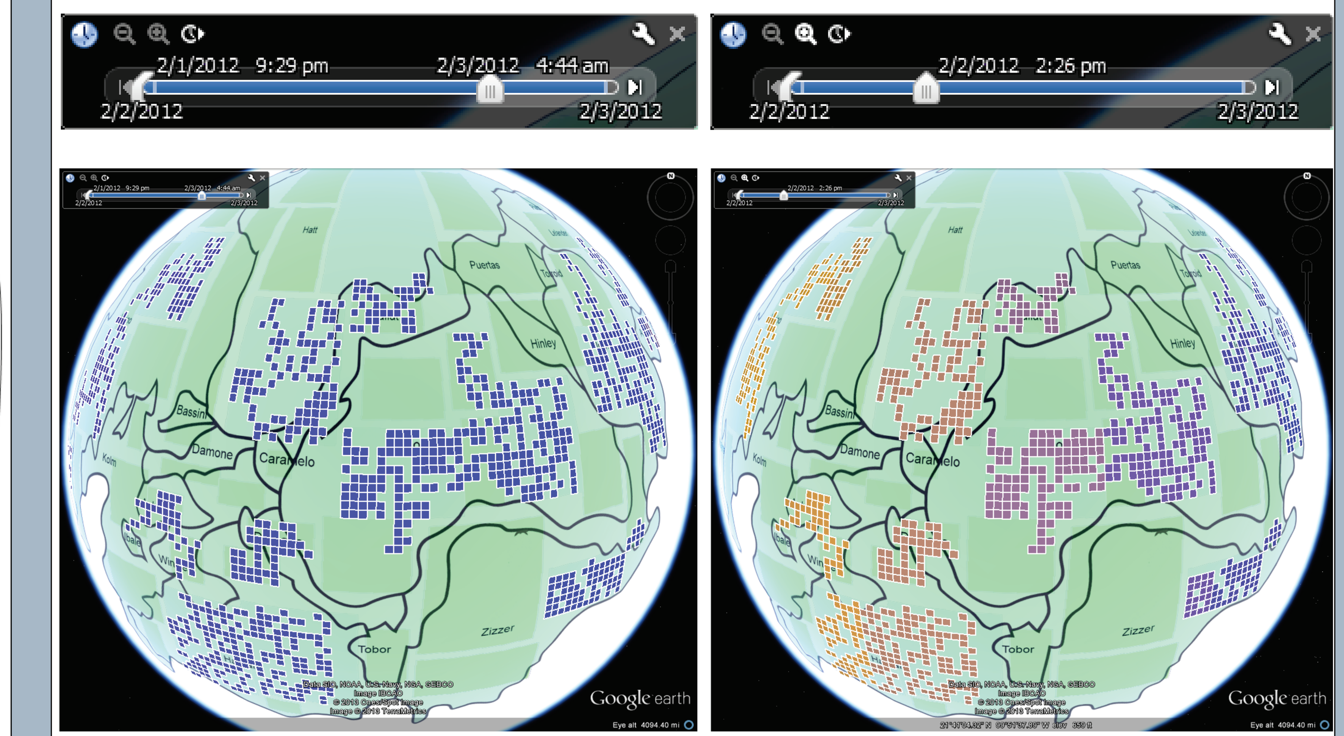


The UI of Google Earth with visualization: the left panel includes various filters for the all input files based on different solutions. Grids could be seen in main window; the right windows are 3D bars visualization with interactions such as bubble with detail information by clicking.

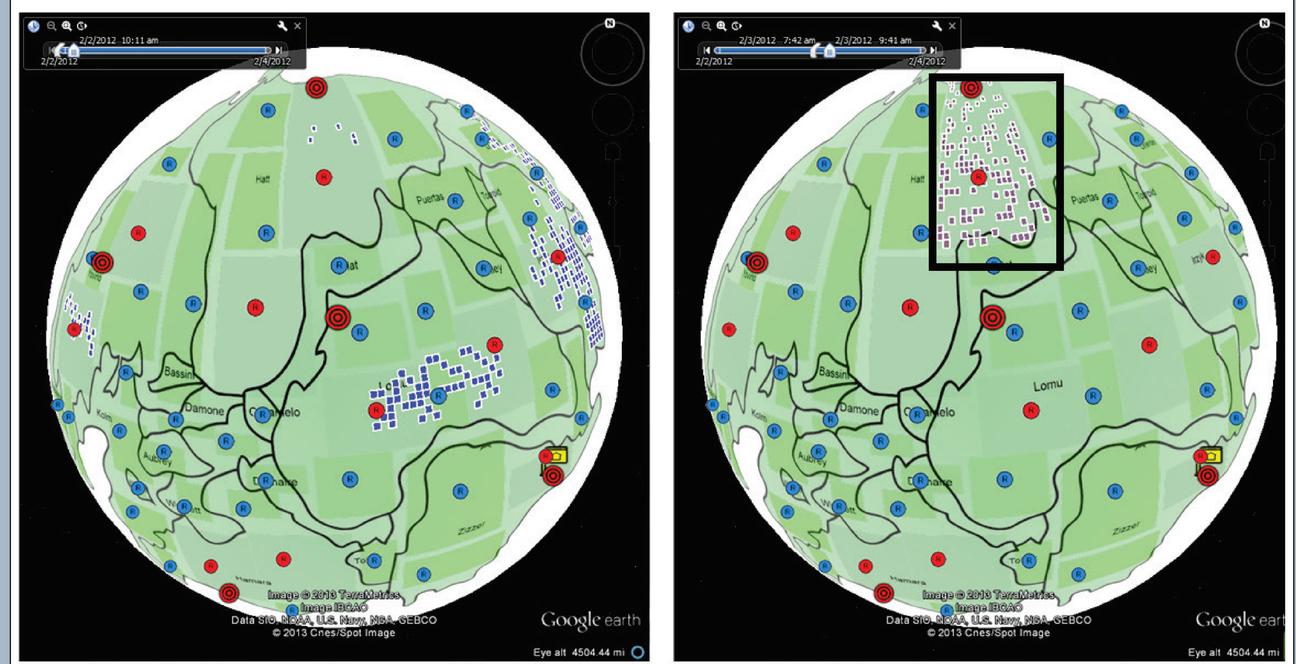
The design of user interaction is two fold, granularity levels of data values based on zoom-in/out level, and granularity levels of colors based on start-end time slider selection by users. For trend presentation, we could animate the moving trend and view the temporal variation. a color encoding method is another solution besides animation to detect moving trend.

Case 1: Corporate Computer Networks

The VAST 2012 mini-challenge 1 data embodies a large-scale enterprise network with network traffic log and geographic information. The log data come from locations all over a fictitious Bank of Money (BoM) facilities that contains close one million IP addresses. The main goal for our visualization solution is to create large-scale cyber situation awareness.

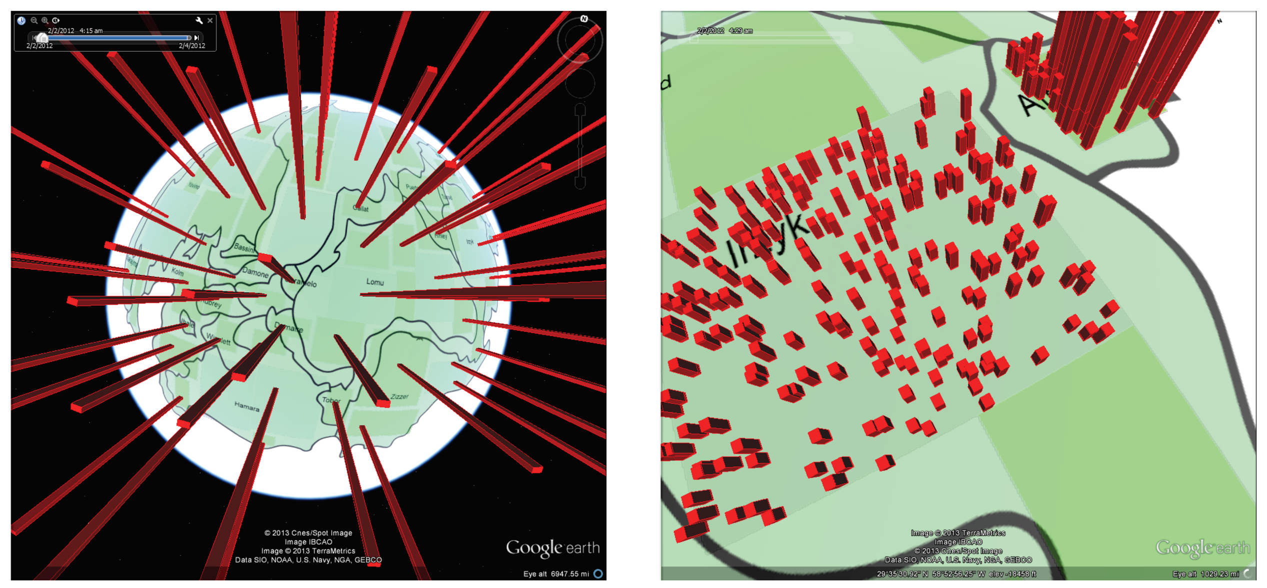


Dynamic color spectrum on grids and bars through interaction. Colors are dynamically defined based on user's selection of start and end time for better and clearer visualization.



2/2/2012 8:12am – 10:11am   2/3/2012 7:42pm – 9:41pm

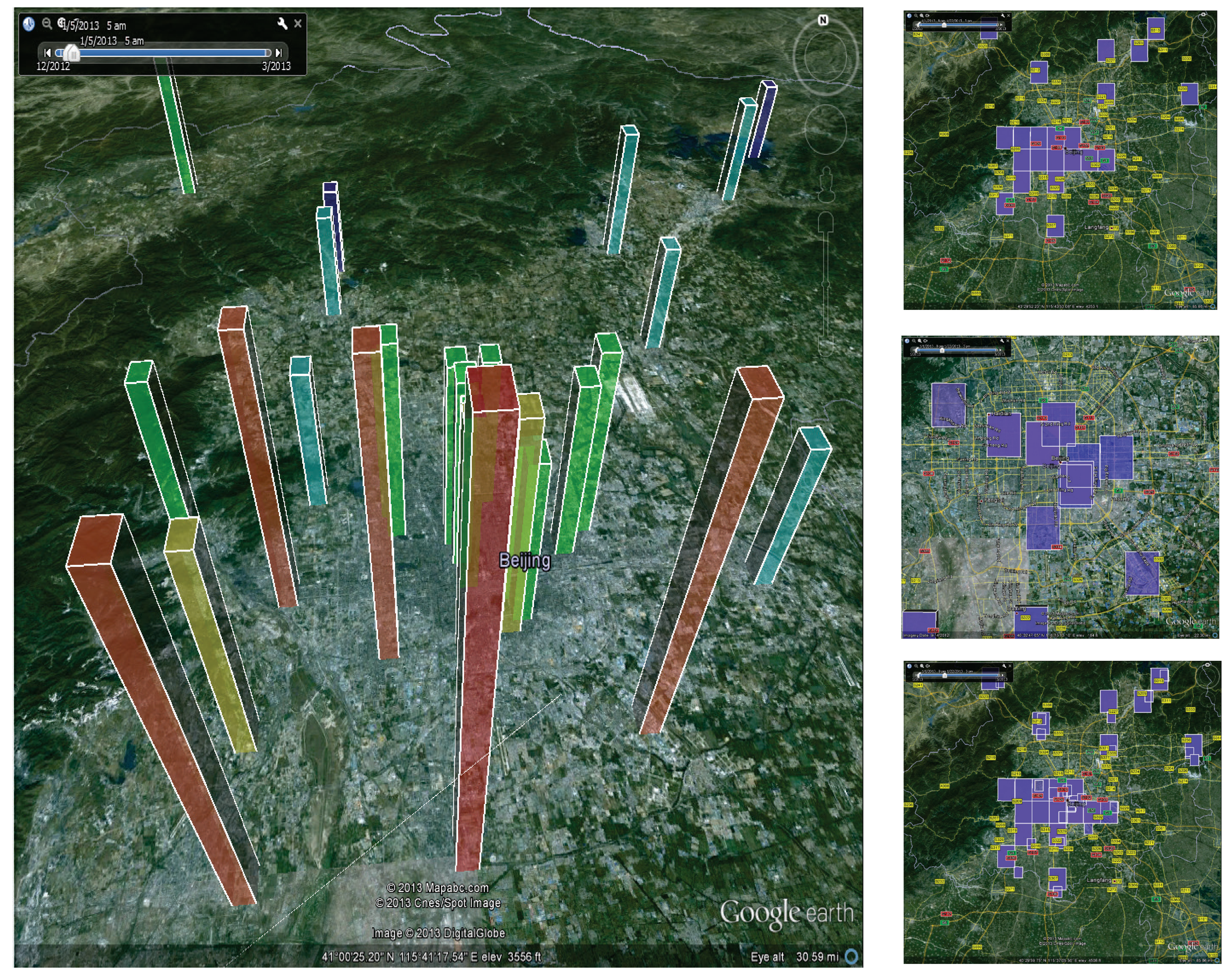
Left picture shows that clusters are in different regions from the first day 8:12am. to 10:11am. The box in the right picture shows that clusters are in one region from the second day 7:42am. To 9:41am.



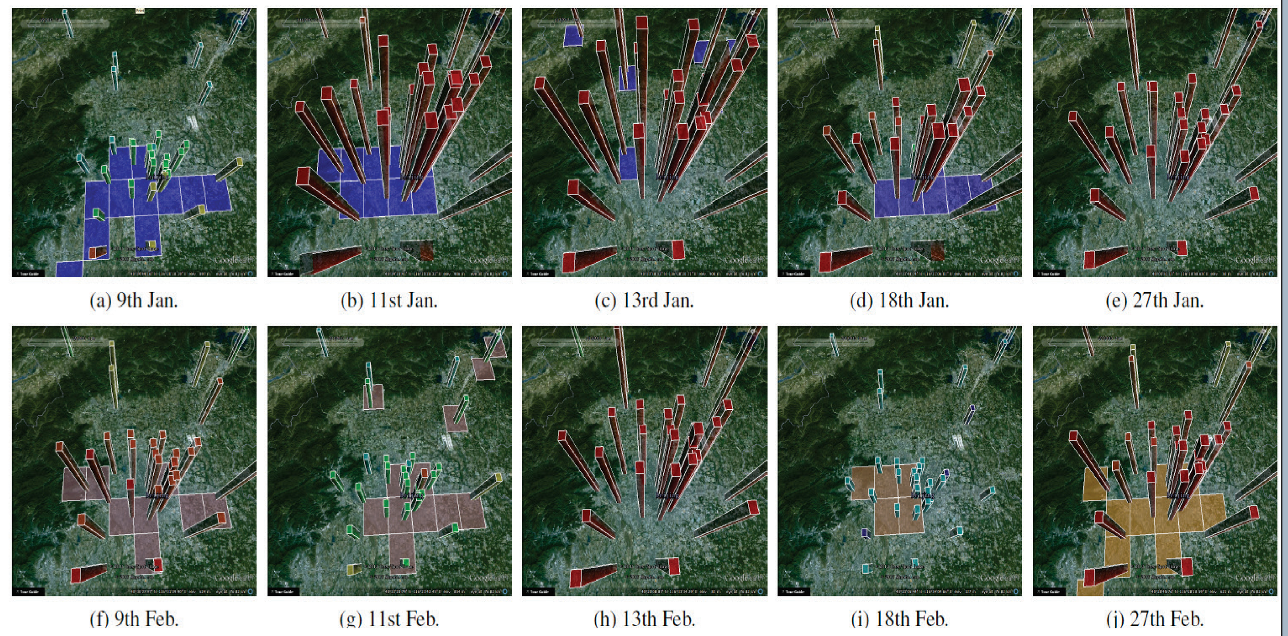
Multi-level granularity of 3D bars through interaction. (left: region level, right branch level)

Case 2: Environmental Quality

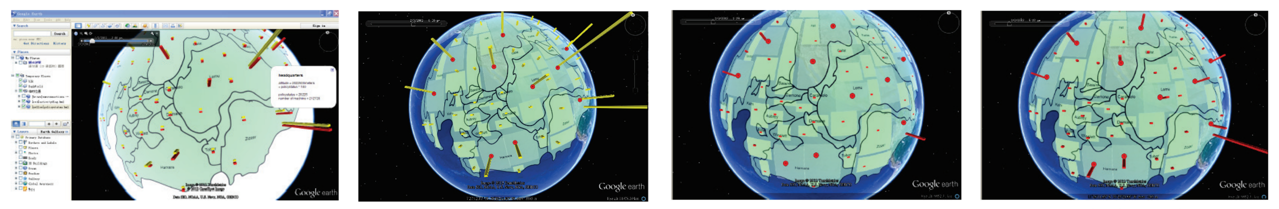
Air pollution problem in developing countries such as China becomes increasing challenging. Beijing has can provide AQI data. We collect the AQI data from more than 34 sensor stations in Beijing provided by Beijing Municipal Environmental Protection Bureau from January to March of 2013, and combine with relative sensor stations' location information.



Left: Bars representing data dimensions (colors imply AQI level) Right: Multi-level granularity of 2D grids through interaction.



Air Quality Index (AQI) in Beijing during January and February 2013



An interactive demonstration can be viewed in the scene of SRCEE

Related Paper

[1] N. Mamoulis, H. Cao, G. Kollios, M. Hadjieleftheriou, Y. Tao, and D. W. Cheung. Mining, indexing, and querying historical spatiotemporal data. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '04, pages 236–245, New York, NY, USA, 2004.

[2] W. Dong, X. Zhang, L. Li, C. Sun, L. Shi, and W. Sun. Detecting irregularly shaped significant spatial and spatio-temporal clusters. In SDM, pages 732–743. SIAM / Omnipress, 2012.

Acknowledgement

Special thanks to Dr. Lei Shi, Dr. Weishan Dong and Lucheng Lv