Mobile Analysis of Large Temporal Datasets for Exploration and Discovery

Andrew Huynh
Computer Science and Engineering
University of California, San Diego
a5huynh@ucsd.edu

Albert Yu-Min Lin
Qualcomm Institute
University of California, San Diego
a5lin@ucsd.edu

Abstract*—The increasing power and decreasing size of mobile devices and tablets provide a compelling new platform for mobile field geographic information system (GIS) operations with geospatial datasets. As these data sets become increasingly larger and more dynamic, it introduces the need for real time analysis and data collection in the field. Successful field research in the digital age requires computing power, functionality, and mobility. Location aware mobile computing devices enable new methods of knowledge synthesis by merging physical and virtual data layers. The geospatial dataset in this study is both large in scale and quickly transforms on a daily basis, requiring innovative strategies for effective application. This paper presents the novel combination of mobile tablets, GIS, and geospatial data to direct ground exploration and discovery of cultural heritage sites in Mongolia.** ***Keywords—Geographic Information Systems, mobile computing, geovisualization, spatio-temporal dataGeographic Information Systems, mobile computing, geovisualization, spatio-temporal data,

I. Introduction

Mobile technology continues to become increasingly powerful and small, while additionally featuring expanded functionality such as Global Positioning System (GPS) receivers, wireless internet, and touch capabilities all at reasonable cost. These advancements and the oncoming ubiquity of mobile computing devices introduce a powerful new platform for mobile GIS operations [1] that has the potential to replace conventional field methods with more efficient and capable alternatives.

Furthermore, high resolution satellite imagery have become increasingly available [2], [3]; improving GIS [4], [5] and enabling researchers to view and analyze geospatial data [6], [7] with ease. Conventionally, visualization and analysis of geospatial data is accomplished using normal computing devices or display systems specifically designed with this visualization and analysis in mind [8]; condensing this information for mobile computing device for field research is a difficult task. Constraints appear due to the small form factor and limited computation power, making it difficult to visualize and understand thousands, possibly millions, of geospatial data points in real-time. As such, real-time analysis and efficient interaction with geospatial data must be important considerations of any proposed technique.

The onset of increased data collection capabilities, such as high resolution satellite imagery, has led to crowdsourcing has become increasingly attractive as a source for scalable analytics. These methods have been applied on a large scale towards analyzing massive datasets in such instances as galaxy classification [9], image annotation [10], music annotation [11], protein folding [12], and transcription of ancient books [13]. Innovative methods have been designed to motivate large scale human participation through gameplay dynamics or monetary incentivisation. Taking this data into the field grants field researchers the ability to interact with a large, temporally changing dataset where they need it. This introduces interesting constraints on how the data must be updated, visualized and analyzed in the field and again must be important consideration of any proposed technique.

Fig. 1: Our system in use in the field as shown in a National Geographic documentary

In this paper we describe the novel use of mobile computing, GIS, and crowdsourced geospatial datasets as an aid for discovery and exploration (Figure 1). The geospatial data points are human annotations collected through the Valley of the Khans (VOTK) project, which asked participants to survey IKONOS and GEOEYE-1 satellite image data provided by the Geoeye Foundation for anomalies that may indicate archeological sites. The annotations create a dynamic geospatial dataset of over 2 million data points. Due to the enormous and dynamic dataset that was used in the field, our approach considered scalability and efficient analysis as key elements required for a successful system. Additionally, we report on the use of our system in the field and the positive results of the field expedition with the system.

II. Related Work

There have been a several attempts [14], [15], [16], [17] to introduce mobile computing as an aid for field research and data collection. Project Battuta [14], [16] approached the problem through a mobile field computing environment using GPS and GIS data to accurately display geospatial data to field researchers. Levy and Smith [15] present a digital field

978-1-4799-3170-5/13/$31.00 ©2013 IEEE

Authorized licensed use limited to: TU Wien Bibliothek. Downloaded on October 26,2024 at 11:11:32 UTC from IEEE Xplore. Restrictions apply.

research methodology encompassing the use the GPS and GIS data along with digital photography and topology. Tripcevich and Wernke [17] present real-time collection of excavation data using mobile GIS. Vanoni et al [18] present a mobile augmented reality interface, allowing users to interact with cultural heritage in novel, more meaningful ways. Thus far, these approaches have relied on custom built technologies, full computing systems, or are geared towards the collection of data in the field.

Our approach focuses on the discovery and exploration of archeological sites. Data is collected and streamed in real-time to mobile computing devices where field researchers can make actionable decisions towards further analysis. User interaction especially plays an important role in analyzing annotations spread over a large geospatial domain. Many information visualization approaches [19], [20], [21] suggest the use of multiple linked views to further enable exploration and discovery of data. Fine-grained data discovery, however, was not used in our system as the focus is quick interaction such that discovery and exploration of large areas of interest could be done efficiently in the field.

III. Technical Approach

Our system employs a tablet computer with touch-based interaction as the basic platform of operation. Functionality provided with the mobile device includes a 9.5 x 7.31 inch touch screen, GPS, cellular internet access, and a compass. The screen was overlaid with georeferenced satellite images with geospatial data points overlaid and displayed to the field researcher. Using the GPS, cellular internet access, and the compass a field researcher has the opportunity to make actionable decisions or initiate georeferenced data collection based on the presented data.

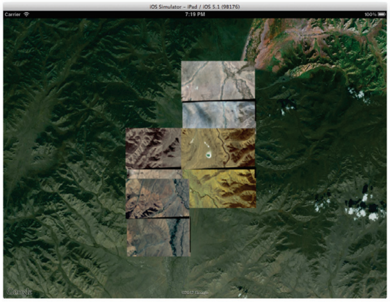
A. Geospatial Overlays

In addition to the publicly available satellite imagery, high resolution satellite imagery can be displayed as overlays as seen in Figure 2. The imagery has a resolution of 0.5 meters/pixel and is specially transformed using a stacked multi-resolution tiling scheme [22] which enables the device to increase the resolution displayed as we zoom into the image (Figure 2b). This tiling scheme enables data management strategies such that a minimal amount of computation power and network bandwidth is required to display the images. Upon interaction, a specific viewport of image tiles is requested by the device and rendered on screen. The tiling scheme performs well on mobile devices where limitations on memory and computation power may prevent the entirety of high resolution satellite imagery from being displayed.

All in all, approximately 50 Gigabytes of satellite image data is available to the device. Through smart bandwidth usage and localized caching, this large amount of data can be efficiently and quickly displayed to a field researcher.

B. Real-Time Data Acquisition

While it is useful to view static information, datasets can change dramatically over time. The ability to stream the latest data for real-time visualization and interaction is an important aspect of our system. A similar data tiling scheme to that



(a) Overview of interest region as seen on tablet device

(b) Zoomed in overview of interest region as seen on tablet device

Fig. 2: Georeferenced satellite imagery allow field researchers to pinpoint their exact location while also giving context to any geospatial data that is displayed. The basic overview is provided by publicly available satellite imagery. The georeferenced satellite imagery is provided by the Geoeye Foundation with a resolution of up to 0.5 meters/pixel. Additionally, using the built-in compass and GPS, a researcher is able to orient themselves on the display.

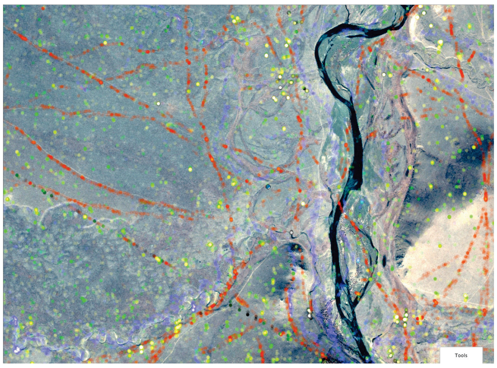


Fig. 3: This figure demonstrates how geospatial data overlays may appear to a field researcher. The data is streamed over the network if available or using a localized cache. Each color represents a different annotation received. Red indicates roads, blue indicates rivers and yellow indicates possible archaeological sites.

Authorized licensed use limited to: TU Wien Bibliothek. Downloaded on October 26,2024 at 11:11:32 UTC from IEEE Xplore. Restrictions apply.

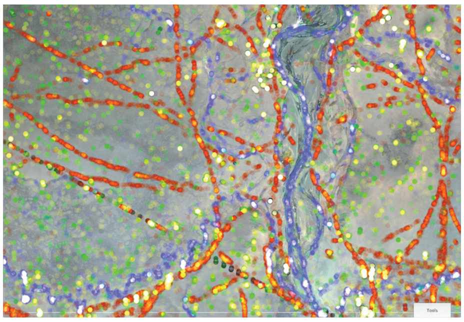
discussed in Section III-A can be used to selectively display geospatial data to field researchers(Figure 3). When a viewport of data is requested, geospatial queries are constructed to select only the data points within a region. This selective querying allows network bandwidth in the field to be minimized; a useful aspect in regions with limited or no connectivity.

Additionally, by using data tiling, data can also be stored in manageable segments. This allows sections to be reused, meaning that for small changes, such as translating the viewport, minimal data fetching is required. In turn, the data only needs to be fully reconstructed when zoom levels are significantly altered or when the viewport is translated substantially. Otherwise, the system is generally able to predicatively gather the data before it is needed on screen allowing for seamless panning of a geospatial dataset.

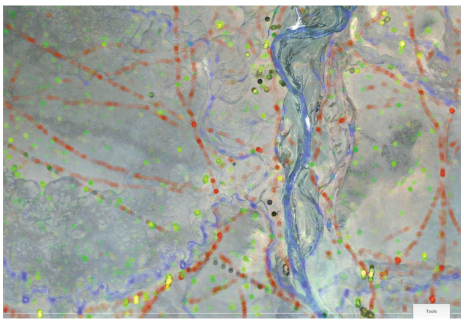
IV. Interaction and Analysis

Traditionally, researchers or explorers searching for interesting locations may spend countless man-hours looking through large geospatial areas in high resolution satellite images. In the field, this data may lack contextual clues or may be difficult to access. A mobile device featuring touch-based interaction presents the ideal environment for researchers to efficiently visualize and understand geospatial datasets in the field.

A. Filtering Data



(a) An example display without normalization filter



(b) An example display with normalization filter.

Fig. 4: The normalization filter allows areas of high human annotation activity to be clearly demarcated, revealing "hotspots" in the data that may be of importance to the field researcher. Areas with low annotation activity are rendered with a lower opacity and thus disappear from the display.

A researcher using our system often has a specific task in mind and the geospatial annotations may become cumbersome to analyze. As such, it is important to provide filtering tools to quickly modify how the data is visualized, allowing the field researcher to obtain the optimal amount of information for analysis. Opacity and size of the data points can be changed to quickly clarify certain locations during analysis. Further normalization based on the number of annotations in a given area is also provided to clearly demarcate areas of high human annotation activity, thus revealing "hotspots" to field researchers (Figure 4). This is apparent when comparing Figure 4b and Figure 4a. Figure 4a is incredibly complex and difficult to parse at a glance, a limitation that may become troublesome in the field. Normalization lowers the opacity of areas with low annotation activity reducing the amount of visual clutter displayed to a field research as seen in Figure 4b.

B. Data Collection

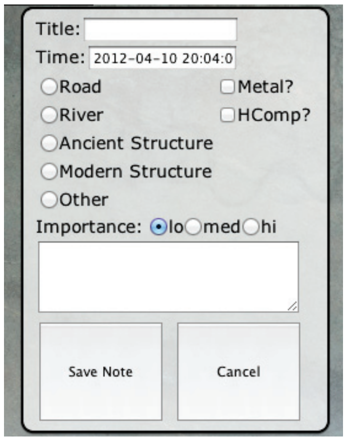


Fig. 5: Georeferenced data collection can occur by tapping on a location and entering the information into the box provided. While in our case it was used mainly to keep track of interesting locations or notes, it could be modified to accept any type of georeferenced data.

Data collection in the field is often troublesome and much of the previous work in digital field research [14], [15], [16] have focused on improving the strategies used in this process. We sought to streamline the collection of data by ensuring every piece of information is georeferenced and timestamped using the GPS provided by the mobile device. A timeline of data collection and exploration can now easily be constructed allowing field researchers to have a precise log of their activities.

The system begins data collection when a researcher taps on a location within the display. This point is converted from screen coordinates into a geocoordinate based on the researcher's current detected location. Afterwards, an information box prompts the researcher for data as demonstrated in Figure 5 with the current time and location stored. In our proposed system, the data collection is mainly used to keep track of interesting locations or notes but could easily be extended to handle any georeferenced data collection.

Authorized licensed use limited to: TU Wien Bibliothek. Downloaded on October 26,2024 at 11:11:32 UTC from IEEE Xplore. Restrictions apply.

V. Evaluation and Field Results

The goals of the presented approach were to: (1) employ mobile GIS techniques to extract information from a large, unfiltered crowdsourced dataset, and (2) facilitate the exploration and discovery of archaeological sites through these analytics.

We deployed our mobile GIS system over a course of an expedition to aid field researchers in the discovery and exploration of cultural heritage sites in Northern Mongolia. Due to the mobile form factor of the devices used, researchers could take their data on foot without any reliance on a central source of information in the field. This ability for field researchers to orient themselves on a map, retrieve real-time information, and do basic analyses for potential cultural heritage sites proved to be effective in quickly discovering hotspots in which the field researchers could then explore. The analysis enabled by the mobile GIS system helped field researches discover over 50 archaeological sites.

VI. Conclusion and Future Work

In this paper we propose and demonstrate an approach to efficiently visualizing GIS and geospatial data on mobile computing devices to aid in the discovery and exploration of archaeological sites. The mobile devices enabled researchers to quickly visualize, filter, and analyze the crowdsourced geospatial data, which was updated in real-time through cellular internet access. Furthermore, researchers were able to make actionable decisions and record data in the field making field annotation of important site information easy and automatically associated through geo-referencing with the site of interest.

Our proposed system and strategies are not only limited to human annotation datasets; it can be applied to any case in which massive datasets of geospatial context are generated and visualized. As mobile computing devices become more powerful and include greater functionality the focus turns to a search for proper methodologies and strategies to improve the efficiency and power of mobile GIS.

VII. Acknowledgments

We would like to thank Prof. S. Bira and Prof. T. Ishdorj of the International Association for Mongol Studies, Dr. F. Hiebert of the National Geographic Society, Prof. Gert Lanckreit and Prof. Serge Belongie and Luke Barrington of UC San Diego and the entire Valley of the Khans Project team for their collaboration. We would also like to thank the Geoeye Foundation, Digitaria, and the National Geographic Digital Media team for their collaborative support in the crowdsourcing effort that generated the data used in this study. This effort was made possible by support from the National Science Foundation through the IGERT award number DGE-0966375 and EAGER award number ISS-1145291 grants.

References

[1] A. J. Wagtendonk and R. A. M. D. Jeu, “Sensible Field Computing : Evaluating the Use of Mobile GIS Methods in Scientific Fieldwork,” *Photogrammetric Engineering & Remote Sensing*, vol. 73, no. 6, pp. 651–662, 2007.

[2] D. Bell, F. Kuehnel, and C. Maxwell, “NASA World Wind: Opensource GIS for mission operations,” *Aerospace*, 2007.

[3] R. Stockli, E. Vermote, N. Saleous, R. Simmon, D. Herring, and R. Stöckli, “The Blue Marble Next Generation-A true color earth dataset including seasonal dynamics from MODIS,” *Earth*, 2005.

[4] D. Butler, “The web-wide world,” *Earth*, vol. 439, no. February, 2006.

[5] S. R. J. Sheppard and P. Cizek, “The ethics of Google Earth: crossing thresholds from spatial data to landscape visualisation.” *Journal of environmental management*, vol. 90, no. 6, pp. 2102–17, May 2009.

[6] M. N. K. Boulos, “Web GIS in practice III: creating a simple interactive map of England’s Strategic Health Authorities using Google Maps API, Google Earth KML, and MSN Virtual Earth Map Control.” *International journal of health geographics*, vol. 4, p. 22, Sep. 2005.

[7] M. Gibin, A. Singleton, R. Milton, P. Mateos, and P. Longley, “An Exploratory Cartographic Visualisation of London through the Google Maps API,” *Applied Spatial Analysis and Policy*, vol. 1, no. 2, pp. 85–97, Apr. 2008.

[8] A. Huynh, K. Ponto, and A. Y.-m. Lin, “Visual Analytics of Inherently Noisy Crowdsourced Data on Ultra High Resolution Displays,” in *2013 Aerospace Conference Proceedings*, 2013.

[9] C. J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. J. Raddick, R. C. Nichol, A. Szalay, D. Andreescu, P. Murray, and J. Vandenberg, “Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey ,” *Monthly Notices of the Royal Astronomical Society*, vol. 389, no. 3, pp. 1179–1189, Sep. 2008.

[10] L. von Ahn and L. Dabbish, “Labeling images with a computer game,” *Proceedings of the 2004 conference on Human factors in computing systems - CHI ’04*, vol. 6, no. 1, pp. 319–326, 2004.

[11] D. Turnbull, R. Liu, and L. Barrington, “A game-based approach for collecting semantic annotations of music,” *Conference on Music*, pp. 2–5, 2007.

[12] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popović, and F. Players, “Predicting protein structures with a multiplayer online game.” *Nature*, vol. 466, no. 7307, pp. 756–760, 2010.

[13] L. von Ahn, “Designing games with a purpose,” *Communications of the ACM*, vol. 51, no. 8, p. 57, Aug. 2008.

[14] K. K. C. Andrea, A. Nuernberger, T. Pingel, and D. Qingyun, “User Interface Design for a Wearable Field Computer,” *Computer. National*, no. 1997, pp. 1–5, 2002.

[15] T. E. Levy and N. G. Smith, “On-Site GIS Digital Archaeology,” *American Contributions to the Archaeology of*, 2007.

[16] S. Nusser, L. Miller, K. Clarke, and M. Goodchild, “Geospatial IT for mobile field data collection,” *Communications of the ACM*, vol. 46, no. 1, pp. 1–2, 2003.

[17] N. Tripcevich and S. a. Wernke, “On-Site Recording of Excavation Data Using Mobile GIS,” *Journal of Field Archaeology*, vol. 35, no. 4, pp. 380–397, Dec. 2010.

[18] D. Vanoni, M. Seracini, and F. Kuester, “ARtifact: Tablet-Based Augmented Reality for Interactive Analysis of Cultural Artifacts,” *2012 IEEE International Symposium on Multimedia*, pp. 44–49, Dec. 2012.

[19] D. Guo, M. Gahegan, and A. MacEachren, “Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach,” *Cartography and Geographic Information Science*, vol. 32, no. 2, pp. 113–132, 2005.

[20] M. Jern, “”GeoAnalytics” - Exploring spatio-temporal and multivariate data,” *Information Visualization, 2006. IV 2006. Tenth International Conference on,*, pp. 25–31, 2006.

[21] J. Roberts, “Exploratory visualization with multiple linked views,” *Exploring Geovisualization*, no. November 2004, pp. 0–8, 2004.

[22] J. Dollner, K. Baumann, and K. Hinrichs, “Texturing techniques for terrain visualization,” *Visualization 2000. Proceedings*, pp. 227–234,, 2000.

Authorized licensed use limited to: TU Wien Bibliothek. Downloaded on October 26,2024 at 11:11:32 UTC from IEEE Xplore. Restrictions apply.