PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON IMAGING SCIENCE, SYSTEMS, AND TECHNOLOGY

CISST'2000

Volume II

Editor:
Hamid R. Arabnia

Associate Editors:
F-X. Coudoux, Y. Mun, G. J. Power, M. Sarfraz, Q. Zhu

Las Vegas, Nevada, USA
June 26 - 29, 2000

Client-Server Based Real-Time Integration of Remotely Sensed and Digital Data*

Debra Lee Davis-Chu, Nagarajan Prabakar and Naphtali Rishe
High Performance Database Research Center (HPDRC)
School of Computer Science, Florida International University
University Park, Miami, FL 33199

Abstract

Spatial data is an important resource that has numerous applications. It can be used for environmental, commercial and educational purposes. It can also be complex and difficult to deal with, particularly when a user needs to integrate two different types of spatial data. This paper will describe a Web-based tool with which remotely-sensed data, such as aerial photography and satellite imagery, can be easily integrated with digital data, such as GNIS and US Census Tiger/Line data, while a user is seamlessly "flying" over the data. This real-time integration is performed automatically, on-the-fly at the user's request and is accessed via a standard Web browser. The main issues and the combination of solutions involved in its implementation and optimization are discussed.

Keywords: spatial data, semantic object-oriented database, geographical information, real-time integration

1. Introduction

The availability and use of remotely-sensed data has increased dramatically in the past several years. The varied types of information that can be extracted from remotely-sensed data is vast and has numerous applications. An inherent problem with this, however, is the complexity often involved in dealing with these types of spatial data. Spatial data sets are inherently very large and come in many different formats, making storage and retrieval of spatial data often cumbersome at best. This is particularly true when the average user needs to access and combine different types of spatial data, such as integrating remotely-sensed data (e.g., aerial photography, satellite imagery, etc.) with digital data sets (e.g. point, line, polygon and textual data). The time, expense and difficulty frequently involved in such manipulation are often beyond the ability or resources of the average user. Yet the information that can be gleaned from such integration is of such great potential for numerous areas of interest that a simple and easily accessible method for obtaining this information is needed.

To provide for this need, we have been developing a tool with which the average user can combine digital data, such as GNIS (Geographic Names Information System) and US Census Tiger/Line data, with remotely-sensed data, such as aerial photography and satellite imagery. This tool has been developed as an extension to our Web-based TerraFly system, a prototype interactive vehicle for flying over and manipulating remotely-sensed data. More specifically, we have been developing a tool with which users can seamlessly fly over remotely-sensed data while, at the same time, this data is automatically integrated with digital data on-the-fly. That is, as users are flying over remotely sensed data, associated digital data, such as the nearest place, place type and/or populated-area/county-subdivision, is continuously displayed. This capability can be further expanded by the creation of on-the-fly graphical overlays and distance calculations. Naturally, as users proceed over the earth, this information is constantly changing. The mechanism behind this automated, real-time

* This research was supported in part by NASA (under grants NAGW-4080, NAG5-5095, NAS5-97222, and NAG5-6830), NSF (CDA-9711582, IRI-9409661, HRD-9707076, and ANI-9876409), ARO (DAAH04-96-1-0049 and DAAH04-96-1-0278), AFRL (F30602-98-C-0037), BMDO (F49620-98-1-0130 and DAAH04-0024) DoI (CA-5280-4-9044), and State of Florida.

integration is transparent to the user. Further, the user does not need to install any special programs. The system can be run in any standard Web browser.

2. Background

The development of this project has tapped current and emerging technology. It extends HPDRC's Web-based TerraFly technology [1,2] and employs the Semantic Object-Oriented Database System (Sem-ODB) [4,6] to cache and interface data. This allows users to access data via the WWW browsers. It does not require the installation of additional software on the user's computer. We believe that this simplification greatly enhances the usability of our system and accessibility of spatial data since users already have some familiarity with browser technology. Some details of our Sem-ODB and Web TerraFly research projects are presented below.

The Semantic Object-Oriented Database System (Sem-ODB)

Due to the inherently large size of spatial data, the storage and retrieval method used in this project is central to its success. The Sem-ODB technology under development at HPDRC has been designed to be efficient at dealing with spatial data and related products [4,6]. Sem-ODB is a general-purpose database management system (DBMS) designed to store varied types of data in an efficient and logical manner, and it easily deals with non-conventional data such as spatial data, as well as with different types of data in the same database, which many DBMS's can't do [5, 10]. It has some key advantages over current database technology. A few of Sem-ODB's other advantages are:

- Sem-ODB gives control to the user via an intuitive structure of information.
- The end-user is empowered to pose complex ad hoc queries.
- A conceptual data model of the enterprise is directly supported.
- Queries can be up to ten times shorter (and so easier to pose) than in relational databases.
- User programs for a semantic view are substantially shorter than for a relational view, achieving major improvements in the application software development cycle, maintenance and reliability.
- Data types are unlimited, strings can be of any length and numbers can be of unlimited length and precision.
- We have developed algorithms to provide very efficient full indexing, allowing fast access to every single fact.
- There is no need for NULL attributes. Sparse tables in relational databases waste space and processing time.
- There is no need for tables and indices, reducing the space allocation required.
- No keys are needed. Referential integrity constraints are supported automatically by the semantic database.

Sem-ODB can easily handle Terabytes of data. To further improve performance and flexibility, Sem-ODB can be used as a distributed database, as it is in this project. In this way, data could be stored on multiple servers and in multiple locations, and retrieved simultaneously from the various locations [7].

TerraFly

The current project is an extension of our TerraFly technology, a prototype interactive vehicle for flying over remotely-sensed data via the Internet. The TerraFly system applies Sem-ODB technology for storage and retrieval of in-house data used by the system and allows its users to fly over and manipulate the retrieved data. The multimedia database used by TerraFly currently contains text, remotely-sensed raster data [3] and graphical data (graphical maps). A friendly graphical user interface is provided for ease of use. A screen shot depicting TerraFly's user interface can be seen in Figure 1.

Some of the features currently available in TerraFly are:

- Internet capable through the use of any browser
- Multiple Data Types & Multiple Flight Windows
- Fine Flight Direction Control
- Varied Flight Speed and Refresh Rate
- Informational Textboxes
- Go-To Coordinate, Place and Address Functions

- Sensor Band Controls for multi-spectral data types:
  - Pre-defined Three-Band Combinations
  - Advanced Custom Three-Band Combinations
- RGB Intensity Control
- Data Dispensing Capability

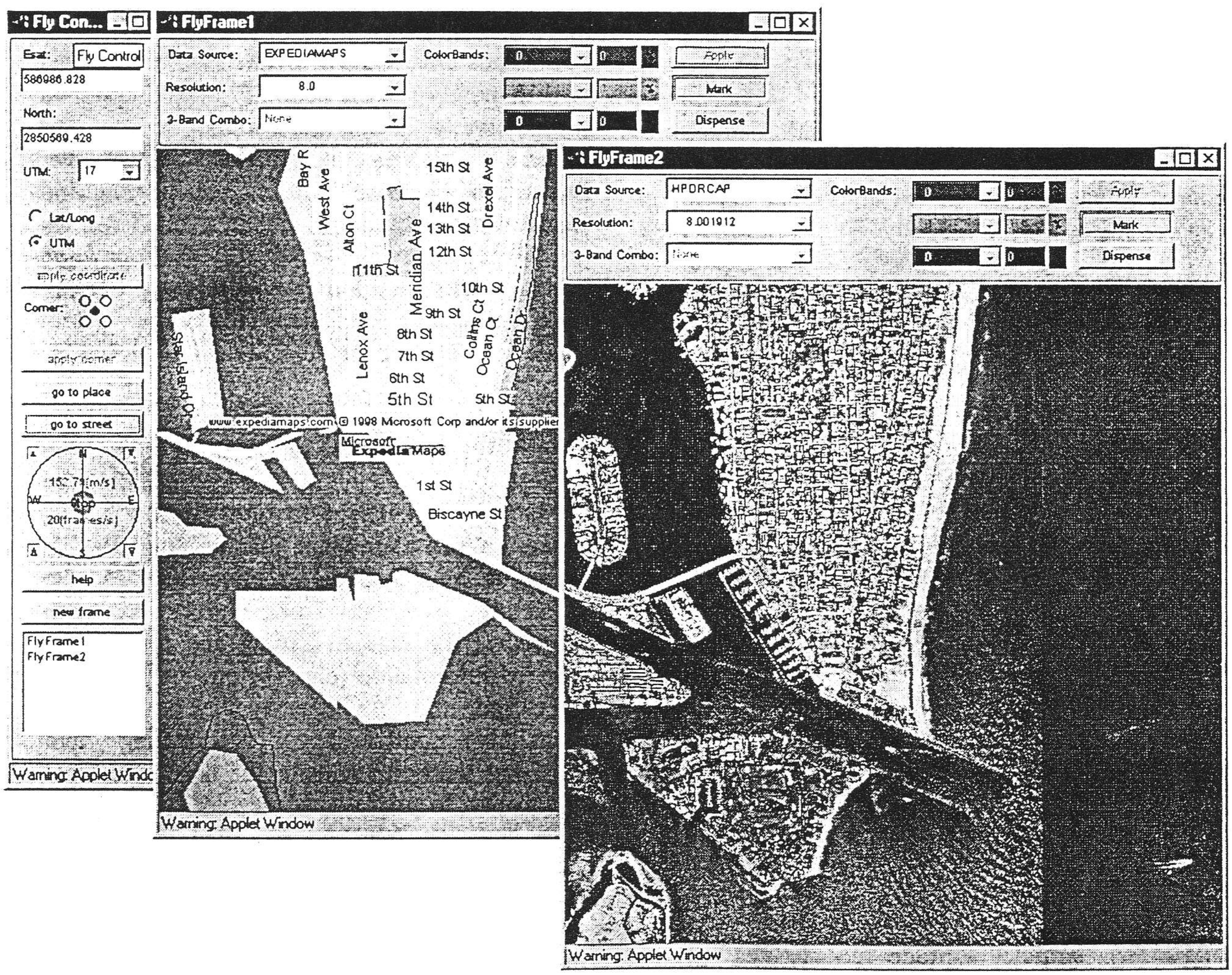

Figure 1: Web TerraFly Interface

3. Digital Data Integration

This project is designed as an extension of the Web TerraFly system to better allow users to combine and understand data of interest over the Internet. More specifically, using any standard browser, a user can 'fly' over remotely-sensed data such as Aerial Photography or Landsat imagery while associated digital information is continuously retrieved and displayed in real time. The information provided includes the name of the closest place, place type and populated-area/county-subdivision, distance from the center point of the image to the closest place and area, as well as on-the-fly graphical overlay of street, place and/or area information. This project uses a semantic database schema, data structure and associated algorithms which can be used to efficiently store, retrieve, manipulate and combine georeferenced digital data with remotely-sensed data. Further, this approach can be used in conjunction with remotely-sensed data at varying resolutions.

Digital Data

There are many types of georeferenced digital data (e.g. point, line, polygon and textual data) available on the market. Two of the most popular are GNIS (Geographic Names Information System) [9] and US Census Tiger/Line files [8]. GNIS data primarily consists of names and types of places along with associated coordinate point information. The GNIS data currently used in this project includes data for the entire US and its territories.

US Census Tiger/Line data consists of point, line and polygon data which provides information such as feature types, address ranges and ZIP Codes, codes for legal and statistical entities, landmark point features, area landmarks, latitude/longitude coordinates of linear and point features, key geographic features and area boundaries. The Tiger/Line data currently used in this project includes information regarding US highways, major roads, streets/addresses and populated-area/county-subdivision shape coordinate points data. The US highways, major roads and street/address data is line data, with each line segment consisting of a series of ordered points with separate beginning and ending points. The populated-area/county-subdivision data is polygon data consisting of a series of ordered points that form a polygon shape.

Project Design

As an extension of the Web TerraFly system, this project is designed to fit seamlessly into the system while maintaining a modular design, which can easily be plugged in and out of the main system. From the user's perspective, only a few simple GUI-based commands are required to integrate digital data with remotely-sensed data. From the system's perspective, it is not this simple. In sum, the process involves continuous client/server interaction, database retrieval, data sorting and search procedures, and calculation functions.

When users choose to integrate the remotely-sensed data they are currently viewing with digital data, the client side of the system sends a request for the desired data to the server. This request includes information such as the type of data needed, resolution and geographical coordinates. When the server side receives the request, it retrieves the needed information from the database, packages its and sends it to the client. When the client receives the data, it places it into a data structure that can be efficiently searched and proceeds with processing the data to extract and display the required information (e.g., the closest place and place type, etc.). As users fly over the data, the client is continuously recalculating and searching through the data for the appropriate information to display. More specifically, each place and area has an associated "boundary" which, in conjunction with its coordinates, is used to determine its distance from the center point of the image. As new tiles of remotely-sensed data are requested from the server, new packages of associated digital data are also requested. In this way, only the data that is actually needed is transferred to the client.

Relevant Implementation Issues

At first glance, it may seem that the integration of digital data and remotely-sensed data is straight forward and easy to accomplish. When one looks at the details involved in automated, real-time, Internet-capable integration as discussed above, however, one discovers that it involves a number of rather complex issues that must be addressed. In short, the main issues are:

- Constant searching involving a large amount of digital data – Because we are dealing with constant retrieval of a large amount of digital data in conjunction with associated remotely-sensed data, efficient storage and retrieval capabilities are extremely important if we want to maintain continual, smooth flight while constantly retrieving the required data.
- Internet bandwidth and browser limitations – Because remotely sense data is inherently large, and memory and CPU intensive, every new feature and each piece of additional data can potentially slow down the system dramatically. We must find a way of effectively dealing with sending large amounts of data over the limited bandwidth currently available over the Internet, as well as dealing with the memory limitations of current browser technology.
- CPU and memory intensive functions – Both imagery and digital data are constantly updated and information is constantly recalculated as users 'fly' over the data. This is highly CPU and memory intensive and must be dealt with appropriately to maintain stable yet accurate data flow.
- Various remotely-sensed data types and resolutions – This issue is particularly relevant to the integration of different types of spatial data. In TerraFly, users can fly over data from resolutions ranging from 1meter to 1 kilometer or more. At different resolutions, different features and feature types are more apparent and relevant to the data. We must have a way of effectively dealing with parsing out ineffective or irrelevant data, as the resolution involved becomes coarser. For example, if a user is flying over 4 meter resolution data, the number of schools that can be seen in one tile (256 x 256 pixels) of data will be relatively small, perhaps 2-3 at most. If, however, the user is flying over the same area at 1 kilometer resolution, it is possible that hundreds of schools can be within the viewable area, yet they are too small to be seen and would often seem to overlap. Situations such as this must be dealt with effectively to avoid unusable data as well as to control unnecessary, excessive data flow and computation.

Optimization

To deal effectively with the complex issues involved in the implementation of this project, we employed a number of solutions, which in combination can satisfactorily resolve these issues. Our solutions are as follows:

- Employ client/server technology where data processing is done on both the client and server sides, with the bulk of the work on the client side.
  We have found that by using efficient data retrieval techniques on the server side, we can reduce the response time of the system. However, the greatest lag time is associated with data transmission over the Internet. Thus, to improve performance, we must make our data processing and retrieval very efficient. To further improve data retrieval time, we are employing pre-fetching combined with a partial sort routine on the server side. This way, the data is already waiting for the next request from the client. Once the client receives the requested data, the client must then insert and sort the digital data into an efficient data structure such as a splay tree. The client is then responsible for all data processing and calculations.
- Incorporate the use of SemODB technology and design a logical and efficient semantic database schema for better retrieval.
  Our database schema has been designed to store remotely-sensed and digital data in the same database while allowing for fast and efficient retrieval of the data. The semantic subschema for the digital data in TerraFly's database is shown in Figure 2. For ease of understanding the association between the specific data we are using and the database design, the category names in the schema reflect the particular data we are addressing in this project. This schema, however, can

easily be modified to contain any type of point, line or polygon data by changing the names of some of the categories.

- Employ the use of a highly efficient data structure on the client side, which can quickly be updated on the fly.

  As users fly, the center geographical coordinate that they are flying over constantly changes. Because of this, the places and areas closest to that center coordinate will also be constantly changing. Thus, in order to provide accurate information, the client employs a highly efficient data structure which aids in the ability to quickly and efficiently search through the digital data and perform needed calculations.

- Preprocess some of the data to provide identification of object boundaries that can then be used for more efficient data retrieval and calculations.

  This preprocessing will aid in the efficiency of both the client and server sides. When a request for new data is sent to the server, one of the pieces of information included in that request is the desired resolution of the remotely-sensed data. This information can then be used in conjunction with object size and boundary information to limit the data retrieved. Thus, when flying over coarse resolution, more relevant and visible data is retrieved. On the client side, some of the preprocessed information is used to more efficiently organize the various place and area objects and provide additional information needed for the calculations used to determine information such as object distance from the center point of the image.

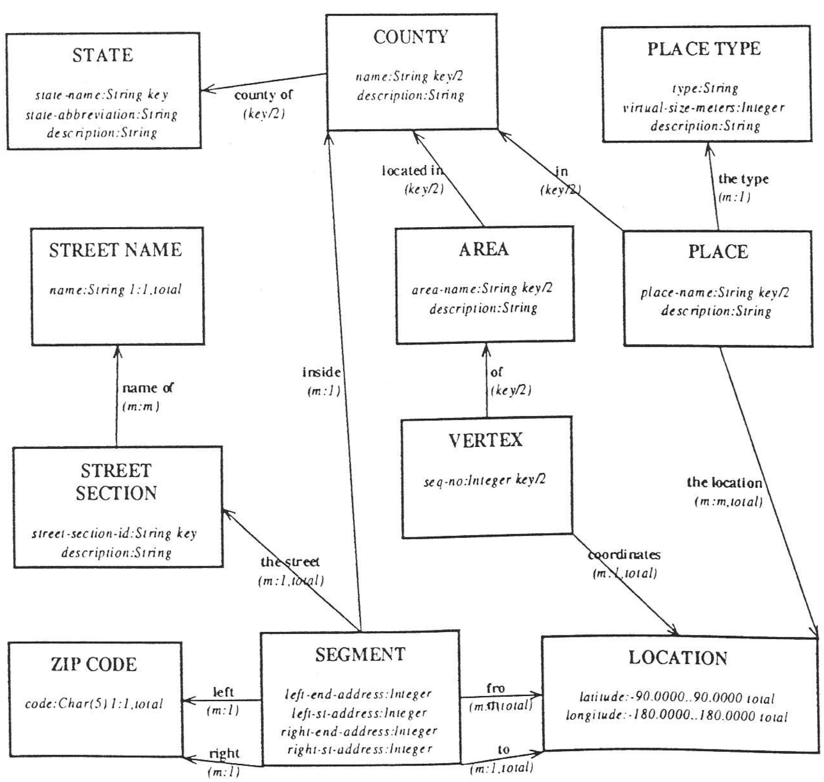

Figure 2: Digital Data Semantic Subschema of the TerraFly Database

4. Conclusion

We have presented a technique for automatic integration of remotely-sensed and digital data in real time to provide users with a more complete understanding of spatial data. This work is an extension of our Web-based TerraFly system and runs through any standard browser. It allows the average user to better use and integrate spatial data while still providing an efficient and user friendly interface. More specifically, our tool allows users to seamlessly fly over remotely-sensed data while this data is automatically integrated with digital data on-the-fly. This is accomplished through the use of a number of approaches including appropriately distributing the workload between the client and the server, providing an efficient database schema, providing an efficient data structure and preprocessing of the data. This is an important and valuable tool, which reduces the time, expense and difficulty frequently involved in dealing with spatial data.

References

1. Alvarez, E. and Rishe, N. "Multimedia Spatial Databases." *Proceedings of the First Workshop on Next Generation Database Design and Applications*, Miami, FL, April 30 - May 1, 1998.
2. Davis-Chu, D.L., Alvarez, E.L. and Rishe, N. "The Creation of a System for 3D Satellite and Terrain Imagery" In *Proceedings of the 13th International Conference in Applied Geologic Remote Sensing*, Vancouver, British Columbia, Canada, March 1-3, 1999, pp. II-329 - II-336.
3. Muffin, G. "Raster versus Vector Data Encoding and Handling: A Commentary", *Photogrammetric Engineering and Remote Sensing*, Vol. 53, No. 10, pp.1397-1398, 1987.
4. Rishe, N. "A Database Design: The Semantic Modeling Approach", McGraw-Hill, 1992.
5. Rishe, N. and Li, Q. "Storage of Spatial Data in Semantic Databases." In *Proceedings of the 1994 ASME International Computer in Engineering Conference*, Minneapolis, MN, pp. 793-800, Sept 11-14, 1994.
6. Rishe, N., Barton, D., Chekmasov, M., Madhyanapu, K., Graham, S. and Chekmasova, M. "Everglades Data Integration using a Semantic Database System." *International Conference Geospatial Information in Agriculture and Forestry*, Lake Buena Vista, FL, pp. I-567 - I-573, June 1-3, 1998a.
7. Rishe, N., Barton, D., Urban, F., Chekmasov, M., Martinez, M., Alvarez, E., Gutierrez, M. and Pardo, P. "High Performance Database Management for Earth Sciences." *NASA University Research Center Technical Conference*, Huntsville, AL., pp. 539-544, Feb 22-25, 1998b.
8. Tiger Overview, URL: http://www.census.gov/geo/www/tiger/overview.html
9. USGS mapping Information: Geographic Names Information System (GNIS), URL: http://mapping.usgs.gov/www/gnis/
10. Waugh, T.C. and Healey, R.G. "The GEOVIEW Design: A Relational Database Approach to Geographical Data Handling", International Journal of Geographical Informal Systems, Vol. 1, No. 2, pp. 101-118, 1987.

Volume I, ISBN: 1-892512-53-x
Volume II, ISBN: 1-892512-54-8
Set, ISBN: 1-892512-55-6