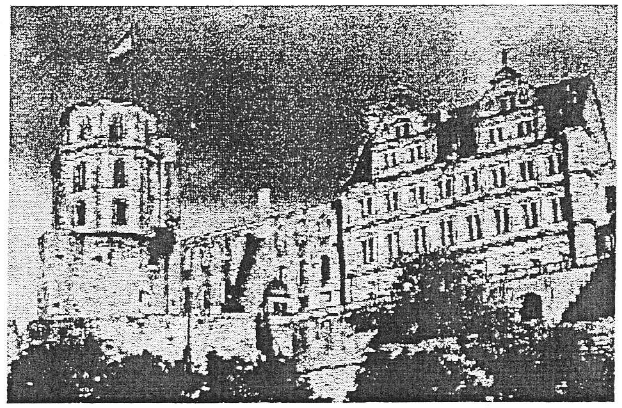
[01-TH]
f=papers

01



ICDE 2001

The 17th International Conference on
Data Engineering
April 2-6, 2001
Heidelberg, Germany

Demo Sessions

Abstracts

TerraFly: A High-Performance Web-based Digital Library System for Spatial Data Access*

Naphtali Rishe, Shu-Ching Chen, Nagarajan Prabakar, Mark Allen Weiss[1], Wei Sun

Andriy Selivonenko, Debra Lee Davis-Chu

High-Performance Database Research Center
School of Computer Science
Florida International University
Miami, FL 33199
{rishen, chens, prabu, weiss, weisun, selivona, davisd}@cs.fiu.edu

Abstract

This paper describes a high performance web-based digital library system, called TerraFly, for spatial data access. This system utilizes a database engine, web technology, dynamic mosaicking of digital images, and a multithread server design developed by our group at the NASA Regional Application Center (NASA RAC) in the High Performance Database Research Center (HPDRC) at Florida International University (FIU). We will demonstrate that this TerraFly system can manage a huge amount of spatial data and provide advanced functionalities.

1. Introduction

With the exponential growth of the World Wide Web (WWW), there are more domains open to GIS applications. The Internet can provide information to a multitude of users, making GIS available to a wider range of public users than ever before. This change of domain users from GIS experts to the general public requires the re-evaluation of design issues and creation of new features for GIS. For instance, system architecture needs to be refined for Internet usage, taking into consideration different platforms across the Internet, network speed, and so on. A major objective is to make a GIS system accessible to the general public, who has little knowledge of spatial data, and allow them to interact with the system to manipulate and retrieve the information they need. In order to address these issues, we, at the High-Performance Database Research Center (HPDRC) [5], have adopted systematic methodologies to solve several serious distributed digital library system problems. This research is based upon TerraFly, an HPDRC web-based spatial data access system that handles different remote sensed data sets. One ultimate goal of our research is to achieve a high-performance distributed digital library system which provides spatial data access.

2. System Architecture

The Web-TerraFly system is based on a client/server model, with a data-less thin client, one CGI proxy server, one multithreaded information server, and a spatial database system. TerraFly's system architecture is shown in Figure 1.

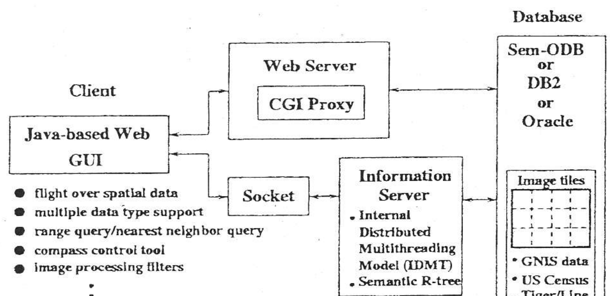


Figure 1: TerraFly system architecture.

Client Side User Interface: The TerraFly client is a Java-based web application, and a data-less graphical user interface (GUI). The client uses two different communication channels to communicate with the data server and information server. The client sends requests to the data server to retrieve images via a CGI proxy server using the HTTP protocol. The client issues a query (either a range query or a nearest neighbor query) through a TCP/IP socket. A communication module is developed for the client to synchronize data transfer from the data server and information

*This research was supported in part by NASA (under grants NAG5-9478, NAGW-4080, NAG5-5095, NAS5-97222, and NAG5-6830) and NSF (CDA-9711582, IRI-9409661, HRD-9707076, and ANI-9876409).

[1] Mark Allen Weiss was supported in part by NSF (EIA-9906600)

server. One benefit of setting up an independent communication module is to isolate communication functionality from the main program and to make the connection as transparent to the main program as possible. This transparency makes future research work on communication easier. The source code for the TerraFly GUI is about 22,000 lines of Java code.

Database Management System (Database Engine): The server connects to a spatial database system built by HPDRC using the Semantic Object-Oriented Database Management System (Sem-ODB) [4], which is based on the Semantic Binary Object-Oriented Model (Sem-ODM) [6, 7]. Sem-ODM combines the advantages of relational and object-oriented data models. It not only has the ability to handle alphanumeric data, but also can manage data that covers multi-dimensional spaces such as image data (maps). Currently, the database contains semantic/textual, spatial/remote sensed and digital data including Digital Orthophoto Quad (DOQ) (Arial photograph) and Landsat data. The source code for the Sem-ODB database engine itself is about 75,500 lines using C++. The TerraFly system can also use any major commercial database system such as DB2, Oracle, Microsoft SQL Server to store spatial data. Corresponding APIs have been developed to support DB2 and Oracle database systems.

Proxy Server: A proxy server is a bridge used to relay data requested by the clients to the database server, and to transmit the data retrieved by Sem-ODB back to the clients. This proxy uses two different protocols, one to interface with the clients, and the other one to interface with the Sem-ODB server. After the database connection is established, all queries can be performed using APIs provided by Sem-ODB.

Information Server: For a distributed GIS system, the objective is to maximize the utilization of resources and to improve performance. To achieve this goal, we designed a model called the "Internally Distributed Multithreading Model (IDMT) [2]." In this model, componentization and distribution of threads are based on the different functionality each thread may have, enabling better usage of the server's CPU and other resources. For example, unlike a traditional GIS system, TerraFly can answer a specific query such as "Find the nearest airport." These kinds of information are very important to users, but they are not used at all in constructing a spatial data structure such as an R-tree. As shown in [2], this information is built into the proposed Semantic R-trees as a partial optimization when it is created for use with GNIS (Geographic Names Information System) [3] and US census Tiger/Line data (street and area information). This data structure provides significant savings in response time. Our experimental results showed that the Semantic R-tree outperforms the well-known R-trees in providing specific information that users request [2].

Dynamic Mosaicking of Digital Images: A dynamic mosaicking scheme is used to remove overlaps and segment images into tiles for effective accessibility [1]. These tiles are stored into our Sem-ODB as georeferenced objects that allow continuous searches, retrievals and display. This design helps to reduce the tremendous network traffic, huge disk space, and memory requirements on the client side.

3. Demonstration

We will demonstrate our web-based semantic database technology by emphasizing its advantages. We will focus on the Semantic Binary Object-Oriented Data Model, Semantic Binary Object-Oriented Database, dynamic mosaicking of heterogeneous digital images, and the IDMT model. Following are some highlights of our demonstration:

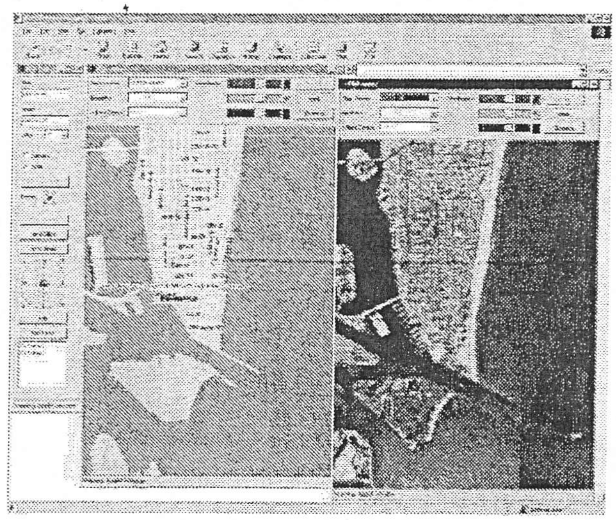


Figure 2: TerraFly User Interface with multiple windows.

TerraFly User Interface: A user-friendly client GUI has been developed to aid both experts and general public users alike to easily interact with the system and to explore all functionalities. A screen snapshot of TerraFly's user interface can be found in Figure 2. This user interface provides many features such as:

- Smooth Flight Over Spatial Data: Data is mosaicked during preprocessing and our proprietary algorithm allows for efficient data retrieval from any DBMS.
- Multiple Data Type Support: IKONOS, Landsat, IRS, Maps, GNIS, US Census, Aerial Photography, and

more.

- Range Query/Nearest Neighbor Query/Semantic Query: Semantic R-trees are used to answer these three types of queries.
- Multiple Windows: TerraFly allows multiple windows to be displayed within the web browser, as shown in Figure 2.
- Web-enabled: Standard Web Browsers can be used for access to the TerraFly system.
- Geolocation ID: The image's geographical coordinates are updated as users "fly" over the spatial data. A GO-TO Coordinate feature is provided in this system, and users can choose 5 point-of-interest locations for the displayed coordinates.
- Compass Control Tool: This feature lets the user to control flight speed, direction, and refresh rate with the click of the mouse.
- Spectral Band Control: Provides the user with the capability to create false-color images on-the-fly by combining spectral bands.
- Go-To Place: Allows users to Go-To a specific place of interest using its name.
- Place Identifier: Finds the exact coordinates of any point in the image, even during flight.
- Street Address Lookup: Finds the coordinates and allows users to Go-To a specific street address.
- Zoom In/Zoom Out: Users can view data at different resolutions using our Zoom feature.
- Data Delivery Capabilities: Users can select an area of interest with our easy to use GUI, and receive the data according to their choice of file format and media.
- Information Overlay: Features from the database can be highlighted and data can be overlayed with different data sources.
- Image Process Filters: Image filters are available to enhance image/features.
- Animated Tutorial: A tutorial is provided to help users learn to use the system.

Dynamic Mosaicking of Heterogeneous Digital Images: We will demonstrate that a source image can be reconstructed from mosaicked tiles on the client side. This dynamic approach reuses digital images upon demand and generates mosaicked tiles only for the required region according to user's requirements such as resolution, temporal range, target bands, etc.

Sem-ODB and Native APIs: We will demonstrate Sem-ODB including its expressive data model capabilities such as inheritance, relations (1:m relations, m:m relations, multi-valued attributes), and other features (including native APIs that provide basic database access).

6. References

[1] Andriy Selivonenko, Nagarajan Prabakar, Naphtali Rishe, and Debra Lee Davis-Chu, "Dynamic Mosaicking of Heterogeneous Digital Images," *ISCA 2nd International Conference On Information Reuse And Integration* (IRI-2000), pp. 86-90, November 1-3, 2000, Honolulu, Hawaii, USA.

[2] Shu-Ching Chen, Xinran Wang, Naphtali Rishe, and Mark Allen Weiss, "A high-Performance Web-Based System Design for Spatial Data Access," *Eighth Symposium of ACM GIS*, pp. 33-38, November 10-11, Washington D.C. USA.

[3] http://mapping.usgs.gov/www/gnis/gnisftp.html.

[4] Naphtali Rishe, *A Semantic Approach to Database Design: the Semantic Modeling Approach*, McGraw-Hill, 1992.

[5] N. Rishe, W. Sun, D. Barton, Y. Deng, C. Orji, M. Alexopoulos, L. Loureiro, C. Ordonez, M. Sanchez, A. Shaposhnikov, "Florida International University High Performance Database Research Center," In *SIGMOD Record*, 24(1995), 3, pp. 71-76.

[6] Naphtali Rishe, Jun Yuan, Rukshan Athauda, Shu-Ching Chen, et al., "SemanticAccess: Semantic Interface for Querying Databases," in *Proceeding of the VLDB conf.*, pp. 591-594, September 10-14, 2000, Cairo, Egypt.

[7] Naphtali Rishe, Alexander Vaschillo, Dmitry Vasilevsky, Artyom Shaposhnikov, Shu-Ching Chen, "A Benchmarking Technique for DBMS's with Advanced Data Models," *ACM ADBIS-DASFAA Symposium on Advances in Databases and Information Systems*, pp. 138-149, September 5-8, 2000, Prague, Czech Republic.