A High-Performance Database System for Managing Large Multi-resolution Medical Images *

Tahsin Kurc, Ph.D.†, Michael Beynon†, Chialin Chang,†, Renato Ferreira†, Ben Bederson, Ph.D.†, Joel Saltz, M.D., Ph.D.†,‡, Alan Sussman, Ph.D.†,‡†

†Dept. of Computer Science
University of Maryland
College Park, MD 20742
{kurc,beynon,chialin,renato,bederson,saltz,als}@cs.umd.edu

‡Dept. of Pathology
Johns Hopkins Medical Institutions
Baltimore, MD 21287

In this work we address the design of a database system to explore, process, and visualize very large (multi-terabyte) multi-resolution image datasets, obtained from MRI, CT and ultrasound, and digitized microscopy images. The basic requirements for such a database management system include (1) support for adding and managing user-defined processing functions, (2) managing datasets stored in disk farms and in tertiary storage, (3) mechanisms to register multi-resolution images into the database, and (4) support for optimized processing on a parallel machine.

We have developed an infrastructure, called the Active Data Repository (ADR) [1], for building parallel database systems that enables integration of storage, retrieval and processing of multi-dimensional datasets on a parallel machine. ADR allows for custom processing of these datasets by applications, while providing support for index generation, data retrieval, memory management, scheduling of processing across a parallel machine and user interaction. The system architecture of ADR provides much of the desired functionality for managing large image datasets. The ADR infrastructure is designed to make it possible to carry out data aggregation operations efficiently in parallel. ADR achieves its primary advantage from the ability to integrate data retrieval and processing for a wide variety of applications and from the ability to maintain and jointly process multiple datasets.

Using ADR as a base, we are building a new database management system, called OR-ELSE DBMS (Object Relational Extemely Large ScalE Database System), to handle very large multi-dimensional and multi-resolution image datasets. OR-ELSE DBMS targets data intensive applications from several domains including Earth Science and Petroleum Engineering, as well as Radiology and Digital Microscopy. The ADR infrastructure is being extended to allow subset operations on very large datasets, some of which may reside on tertiary storage devices, as well as to operate in interactive disk cache mode. The OR-ELSE DBMS will provide support for multi-scale, multi-resolution datasets composed of multiple structured and unstructured grid patches. We are coupling the OR-ELSE DBMS to a relational database, which is used to store metadata describing attributes of OR-ELSE datasets, and to index those attributes. OR-ELSE DBMS also addresses registration of multi-resolution images.

The core visualization component of our database system is based on a zooming engine, called JPad, a new implementation of a zoomable user interface (ZUI) based on Pad++ [2]. ZUIs connect information graphically, and present details without losing context because the context is always a quick zoom away. ZUIs use screen real estate effectively, and have great potential even on small screens. We are adapting the JPad ZUI paradigm to give users the ability to interactively explore large multi-scale, multi-resolution datasets. ZUIs will be used to interactively generate queries into OR-ELSE DBMS and visualize data products derived from several OR-ELSE datasets. The generalized JPad ZUI will be the front-end of a complete application with a customized OR-ELSE DBMS instance as the back-end.

References
