INTRODUCTION

Spatio-temporal data play an important role in science, both as observed natural phenomena like temperature curves or satellite imagery and as artificially generated data such as simulation results or statistical data derived from spatio-temporal phenomena. As a coarse but common classification, spatio-temporal data can be grouped into discretized and conceptually continuous data (Figure 1). The first category allows points, lines, areas, and bodies to have any coordinate value, while the latter category has all data values sitting at the crosspoints of equidistant grids. When dealing with maps, these categories are called vector and raster data, respectively, while in Computational Fluid Dynamics (CFD), for example, the terms general mesh and regular mesh are in use.

We will use the term raster/sampled/discretized data or Multidimensional Discrete Data (MDD) interchangeably with the definition that such an object consists of a set of point/value pairs where the points fill an axis-parallel rectangular area in the Euclidean space $\mathbb{Z}^d$ for some dimension $d \geq 1$. Obviously, this structure is equivalent to an array in the programming language sense.

MDD appear in a large variety of applications. Examples for multidimensional raster data are 1-D scalar measurements like temperature and radioactivity, 2-D satellite imagery spanning large seamless maps of the Earth’s surface, 3-D image time series ($x/y/t$) and geophysical data ($x/y/z$), and 4-D climate models ($x/y/z/t$). Figure 2 shows some of today’s most important areas; each field in turn usually has many diverse subfields, as the example of mapping/cartography shows.

The kind of services required on MDD can be summarized as fast, flexible selection on huge raster data assets.

For example, in Figure 3, the aerial image of Bavaria is depicted, consisting of 950,000 x 1,000,000 RGB (red/green/blue) pixels; this represents a raw data volume of 2.85 Terabyte (TB). A common task on such geo-imagery is to interactively select and display the overall image or a small cutout scaled to the client’s window size (see Figure 3), something which occupies, say, 30 kB as a JPEG image. Further operations involve analysis like deriving the vegetation index from multi-band satellite images.

Recent hardware development allows holding of such large objects available for online access, and there is an increasing demand for Web services including MDD support, such as aerial/satellite image archives (see www.ceos.org, for example). Consequently, database systems (DBMSs) in future will have to provide storage, query, and optimization support to accommodate MDDs side by side with the traditional data types — in database speak, MDDs must become first-class citizens in the database. In the sequel, we discuss how this can be accomplished and what issues are still waiting to be solved.

Figure 2. Raster application fields

Figure 3. Aerial image of Bavaria: downscaled (thumbnail) overall view (left) and zoomed cutout (right)
BACKGROUND

Differentiation to Other Fields

A related but different field where databases also handle images is Multimedia Databases. Multimedia database systems rely on content recognition techniques to extract semantic knowledge from the image beforehand and henceforth perform all querying on this semantic net leaving the imagery untouched except for displaying them unchanged. Raster databases, conversely, work on the pixel level and do not attempt to understand the contents; rather, they allow quick navigation on and selection from very large data objects.

Another related field is image processing. While the set of operations known there exceeds raster database functionality by far, data sets in imaging systems traditionally have to fit into main memory; raster databases, on the other hand, focus on data selection on objects which may well exceed main memory capacity by a factor of a thousand or more.

History and State of the Art

Work on raster databases—whether industrial or scientific—falls into two basic categories: statistics and sensor/image databases. The field of statistical databases has received outstanding thrust through data warehousing and online analytical processing (OLAP) which use model business data as cell values (“facts”) allocated in multidimensional spaces (“data cubes”) described by abstract dimensions (“features”).

Completely separate from this, sensor and image data have been investigated. Traditionally, images have been stored in BLOBs (binary large objects), that is, byte strings without any further semantics, introduced as “long fields” by Lorie (1982). First approaches to add more semantics include Tamura (1980) where a set of imaging functions was added to the programming interface, but not to the query language; there was no conceptual background justifying the operations chosen. A first image query language was proposed with PICDMS (Chock, Cardenas & Klinger, 1984; Joseph & Cardenas, 1988). However, many queries were dependent on the operation sequence, and no architectural support for large objects was indicated. In Vandenberg and DeWitt (1991), a general-purpose conceptual database model was extended with simple array capabilities. The quest for support of non-trivial array operations in a query language was first phrased in Buneman (1993). A conceptual raster model with a declarative, optimizable query language based on an algebraic framework was presented in Baumann (1994, 1999); this approach has been implemented in the rasdaman system (www.rasdaman.com) which is in worldwide commercial use. Other algebras which have been implemented to a lesser extent are Libkin, Machlin, and Wong (1996) and Marathe and Salem (1997, 1999).

An example for domain-specific DBMS extensions to accommodate, in this case, 3D medical imagery is described in Arya et al. (1994). Requirements for supercomputing data management have been stated in Kleese (2000) from an application point of view.

With their version 10g, Oracle (www.oracle.com) has released raster support for large 2-D geographic imagery.

The main difference between statistical and sensor/image databases does not lie in the data structure (both deal with multidimensional grids), nor are operations substantially different (an OLAP roll-up from days to weeks is mathematically close to scaling an image by a factor of 7). The essential difference lies in the sparsity, that is, the percentage of cells in the data space considered which actually carry a value. Statistical databases are sparsely populated (on the average about 2%, maximum 5%), while image data usually are densely populated (usually between 60% and 100%). Technology to handle such data is completely different for both cases. However, given the far-going similarity between both, it seems promising to research ways for an integrated approach.

Relevant Bodies

- SQL/MM defines database handling of imagery in the context of the SQL standard.
- The OpenGIS Consortium (OGC, www.opengis.org) standardizes interfaces for Web-based services on geographic information, among them, multidimensional raster (“coverage”) data in the Web Coverage Service (WCS) standard.
- CODATA (Committee on Data for Science and Technology, www.codata.org) is a user-driven international organization whose goal is to enhance accessibility of scientific data.
- ERCOFTAC (European Research Community On Flow, Turbulence And Combustion, www.ercoftac.org) coordinates data management research and provides sample data sets in the field of Computational Fluid Dynamics (CFD).

A SAMPLE RASTER SERVER

As an example for multidimensional raster database support, we sketch the rasdaman system which has been implemented in the course of several European-funded research projects and has been commercialized mainly in the field of geographic image map services. The underly-
Related Content

Evolutionary Database: State of the Art and Issues
[www.igi-global.com/chapter/evolutionary-database-state-art-issues/20693?camid=4v1a](www.igi-global.com/chapter/evolutionary-database-state-art-issues/20693?camid=4v1a)

Semantic Enrichment of Geographical Databases
[www.igi-global.com/chapter/semantic-enrichment-geographical-databases/11209?camid=4v1a](www.igi-global.com/chapter/semantic-enrichment-geographical-databases/11209?camid=4v1a)

RORIB: An Economic and Efficient Solution for Real-Time Online Remote Information Backup
Scott J. Lloyd, Joan Peckham, Qing Yang and Jian Li (2003). *Journal of Database Management* (pp. 56-73).
[www.igi-global.com/article/rorib-economic-efficient-solution-real/3299?camid=4v1a](www.igi-global.com/article/rorib-economic-efficient-solution-real/3299?camid=4v1a)

Towards Code Reuse and Refactoring as a Practice within Extreme Programming
[www.igi-global.com/chapter/towards-code-reuse-refactoring-practice/4292?camid=4v1a](www.igi-global.com/chapter/towards-code-reuse-refactoring-practice/4292?camid=4v1a)