INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

Integrating Image and Spatial Data for Biodiversity Information Management

Ricardo da S. Torres *Claudia B. Medeiros*
Eric M. Hallerman *Marcos André Gonçalves*
Edward A. Fox

Technical Report - IC-04-012 - Relatório Técnico

October - 2004 - Outubro

Integrating Image and Spatial Data for Biodiversity Information Management

Ricardo da S. Torres* Claudia B. Medeiros* Eric M. Hallerman†

Marcos André Gonçalves ‡ Edward A. Fox‡

Abstract

Biologists gather many kinds of data for biodiversity studies; these data are managed by distinct types of information systems. GIS-based biodiversity systems support sophisticated spatial correlations on living beings and their habitats, and spatio-temporal ecosystem modeling. Image information systems allow content-based image retrieval, to help species identification based on similarity (e.g., shape and color characteristics). Different kinds of rule-based systems support species characterization. Unfortunately, these systems (and the underlying data) are independent of each other. This paper presents a solution that seamlessly combines these functionalities, supporting queries that merge textual descriptions, spatial correlations and content-based predicates. The solution is being implemented at Virginia Tech, for identification and data retrieval, supporting management of fish species. It takes advantage of innovations in Digital Library technology to combine networked collections of heterogeneous data under integrated management.

1 Introduction

Biodiversity Information Systems – BIS (e.g., [1, 3, 4]) – involve huge sets of geographic data as well as large databases concerning species' descriptions (e.g., taxonomic classifications). Most biodiversity systems are concerned with determining spatial distribution of one or more living species, and the spatio-temporal correlations and trends of these distributions. This requires combining data on species (when and where they are observed, by whom and how) with geographic data (characterizing the ecosystems where the species are observed). Data integration usually is based on spatial properties, and thus Geographic Information Systems (GIS) and geographic databases are essential to develop this kind of system.

An example of a standard spatial query in a biodiversity system is "Show the areas where the fish species *Percina rex* has been observed". A typical spatial correlation query

*Institute of Computing, University of Campinas, Av. Albert Einstein, 1251, CEP 13084-851, Campinas, SP, Brasil

†Department of Fisheries and Wildlife Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, 24061-0321, USA

‡Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA, 24061-0106, USA

requires combining information on species location and climatic conditions. Drawings and photos of species also may be used in this context. They are stored apart in data files, and treated as auxiliary documentation, usually being retrieved by species' name. One example of a query involving image files would be "Show all photos of fish species *Percina rex*". Images are accessed only via textual queries, ignoring content-based image retrieval. In these systems, scientists must always search for specific species by name.

Content-based image management, on the other hand, allows scientists to identify species using a given image (e.g., a photo) and search in a database for the "most similar" images. Geographic distribution, in this case, is stored as textual metadata (e.g., names of regions), and spatial correlations are infeasible.

The goal of this research is to provide biodiversity researchers with a BIS that combines these types of searching characteristics for exploratory querying. This BIS will help scientists to enhance or complete their knowledge and understanding about species and their habitats by combining textual, image content-based and geographical queries. An example of such a query might start by providing an image as input (e.g., a photo of a fish) and then asking the system to "Retrieve all database images containing fish whose fins are shaped like those of the fish in this photo". A combination of this query with textual and spatial predicates would consist of "Show the drainages where the fish species with "large eyes" coexists with fish whose fins are shaped like those of the fish in the photo".

Challenges involve work on two fronts: image processing and databases. Available systems do not attack these questions simultaneously – they either concentrate on image data or on spatial data. Indeed, GIS that support image queries are concerned with spatial correlations and not with image features (such as color or texture features). Our work, instead, combines these sources of evidence taking advantage of digital library facilities, which offer an organized infrastructure to integrate networked collections of heterogeneous data. These data consist of images of the living beings and geographic distribution, as well as maps and geographic, ecological, and image metadata. Our solution is being instantiated in a BIS for fish species in a real application. The goal is to help students, researchers, managers, and members of the general public to identify fish specimen by using retrieval techniques.

This text is organized as follows. Section 2 outlines the architecture of the BIS. Section 3 discusses the application scenario that instantiates the architecture for an ichthyology biodiversity system. Section 4 gives a brief introduction to related research. Finally, Section 5 presents conclusions and ongoing work.

2 Architecture

Figure 1 shows the basic architecture proposed for biodiversity information management. This architecture is composed by three main layers: data collections (see Section 2.1), search services (Section 2.2), and BIS Manager (Section 2.3).

Collections are organized in a digital library comprised of a set of search services. The BIS combines textual queries with image processing algorithms to extract image descriptors, and spatial data management in geographic databases based on location and on ecological features.

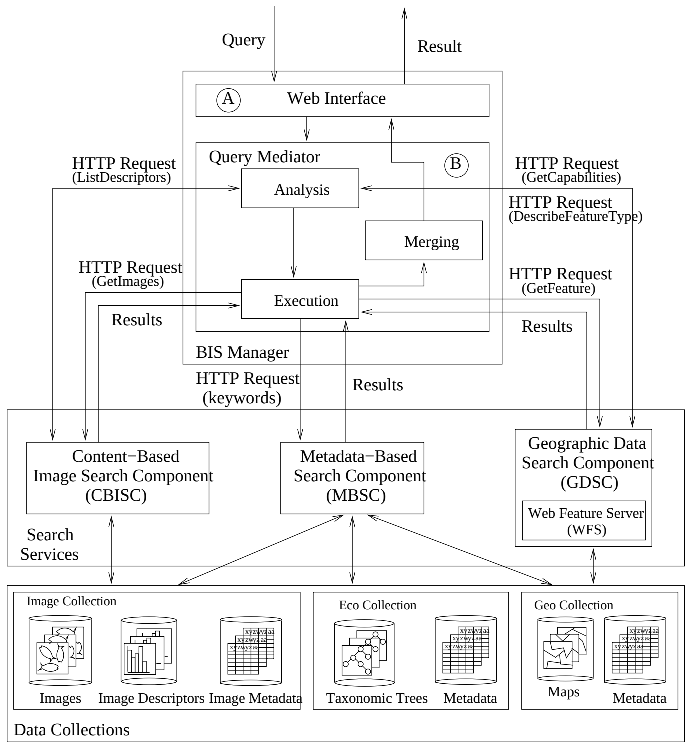

Figure 1: BIS System Architecture.

Although this architecture has been specified in a generic way, its implementation is being carried out for particular fish species. Thus, image data consists of fish photos, geographic data concerning areas where these fishes are likely to be found, and biodiversity metadata on fish and their ecosystems. As will be seen in Section 3, a considerable part of this architecture has already been implemented and tested. We are now working on the final integration with the geographic data search component.

2.1 Data Collections

This layer is responsible for database storage and low-level data management – image, geographic, and metadata databases. Sections 3.2 and 3.3 comment on the data collections used in the present implementation.

2.2 Search Services

Three search components are provided: a geographic data search component (Section 2.2.1), a content-based image search component called *CBISC* (Section 2.2.2), and a metadata-based search component called *ESSEX* (Section 2.2.3).

2.2.1 Geographical Data Search Component

The *Geographical Data Search Component (GDSC)* encapsulates a Web Feature Server (WFS) [45], an OpenGIS consortium [33] recommendation for fostering interoperability. It defines an interface allowing requests for geographical features across the Web, and uses the XML-based Geography Markup Language (GML) [21] for data exchange. GML utilizes XML to express geographical features. It can serve as a modeling language for geographic systems as well as an open interchange format for geographic data. We are using the GeoServer [20] free implementation of OpenGIS Consortium's WFS implementation specification.

The WFS submodule of the *GDSC* is responsible for performing queries on the data sources: it receives HTTP requests from the client (the *Execution* submodule in Figure 1) and returns results as a GML or XML document, depending on the request. A WFS request consists of a description of a query or a data transformation operation, applied to one or more features. Available operations include:

GetCapabilities: A WFS must be able to describe its capabilities. Specifically, it must indicate which feature types and what operations are supported on each feature type. For example, it could define that a *feature type* named *ekey:fishspecies* encoding occurrences of fish species within a specific region is available. It also could indicate supported operations on this feature type (such as operations based on spatial predicates – e.g., *Intersect*, *Within*, etc.).

DescribeFeatureType: A WFS must be able, upon request, to describe the structure of any feature it can service.

GetFeature: A WFS must be able to answer a request, and retrieve feature instances.

2.2.2 Content-Based Image Search Component

One of the most common approach to retrieval image is based on the so-called *Content-Based Image Retrieval (CBIR)* systems. Basically, these systems try to retrieve images similar to a user-defined pattern (e.g., image example). Their goal is to support image retrieval based on *content* properties (e.g., shape, color, or texture), which are often encoded in terms of *image descriptors*.

The *Content-Based Image Search Component (CBISC)* is a new search engine recently developed to support content-based queries on image collections [12]. It supports retrieval using color, shape, and texture descriptors, with 1D or 2D feature vectors. *CBISC* encapsulates multidimensional index structures [6] to speed up the search process.

CBISC is based on the Open Archives Initiative (OAI) [31, 41] principles. The OAI develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content.

Our *CBISC component* is an OAI-like search component that supports queries on image content. As in the OAI protocol [28, 31], queries are given by way of HTTP requests. However, we generalize to have an extended OAI (XOAI) protocol for image search that fits into the Open Digital Library (ODL) framework [41, 42]. As is typical with XOAI protocols, each request specifies the Internet host of the HTTP server and gives a list of key-value pairs. Two different requests ("verbs") are supported by this image search component:

ListDescriptors: Used to retrieve the list of image descriptors supported. No arguments are required for this verb.

GetImages: Used to retrieve a set of images by taking into account their contents. Required arguments specify the query image, the descriptor to be used, and the kind of query. The present version of *CBISC* supports two kinds of queries: *K-nearest neighbor query (KNNQ)* and *range query (RQ)* [6].

2.2.3 Metadata-Based Search Component (MBSC)

The *ESSEX* search engine [14] is being used as our metadata-based search component. *ESSEX* is a componentized vector-space search engine optimized for digital libraries. *ESSEX* acts as the core portion of an Open Digital Library (ODL [41]) search component, answering requests transmitted through an extended OAI (XOAI) protocol. *ESSEX*, available as open source software, was primarily developed for the CITIDEL (Computing and Information Technology Interactive Digital Educational Library) project [7], and also is being used in the *PlanetMath* project [34]. In *ESSEX*, all information is indexed in "chunks" associated with field names, where chunks may correspond to XML elements in a metadata record. Its high speed is the result of both keeping index structures in memory and using a background daemon model based on socket communication with the DL application.

2.3 BIS Manager

This module comprises a Web interface and a query mediator.

A) Web Interface:

This interface supplies query specification and visualization of results. The user will be able to formulate textual queries, interactive queries on maps, queries for image content, or a combination of these.

B) Query Mediator:

The search services are supported by a *Query Mediator* implemented as a server, which combines query mechanisms in metadata, image, and geographic data collections.

Its *Analysis* submodule receives as input a specification in terms of a query image, query terms, and/or rectangle coordinates in a map – and parses it. The parsing process takes advantage of previous knowledge of the *GDSC* and *CBISC*. In the former case, this information is obtained in the form of XML and XML schema documents, obtained each time it performs *GetCapabilities* and *DescribeFeatureType* requests on the *GDSC*. In the latter case, the *Analysis* submodule performs a *ListDescriptor* request on *CBISC* to obtain the enumeration of the image descriptors supported. Note that this information also can be used to guide the query optimization process.

The *Execution* submodule is in charge of forwarding the sub-queries to the appropriate search component (*CBISC*, *MBSC*, or *GDSC*). Finally, the *Merging* submodule combines the obtained results by using an appropriate combination scheme, and returns a ranked list containing the "most" similar objects matching the original specification.

3 Application Scenario and Implementation

The application scenario concerns the instantiation of the proposed architecture to support the creation of a BIS for fish species in a real application. The goal is to help fisheries students, researchers, managers, and the general public to identify fish specimens by using search retrieval techniques. This system will be used by students in ichthyology courses of the Department of Fisheries and Sciences at Virginia Tech.

3.1 Problem

Given a mixed collection of specimens from a river, ichthyologists face the problem of identifying which fish species are present in that collection. Their aim is to determine the taxonomic classification (e.g., family, genus, species) of each given specimen. The traditional approach is based on the use of dichotomous *keys* – basically, rules defining a decision tree that is traversed until one reaches an identification (e.g., [27]).

Operationally, this approach suffers from several problems. First, while an experienced scientist knows how to answer technical questions on subtle features of fish anatomy in order to use a dichotomous key, a student or non-scientist will find it difficult or impossible to correctly answer those questions. Second, dichotomous keys often lack images to support their use by non-experts. Third, dichotomous keys invariably lack reference to geographic distributions of fishes, although geographic data can prove highly useful for fish identification. For example, knowing only where they collected a specimen, novices often have difficulty making species identifications. Access to geographically explicit information on fishes occurring in a watershed – especially if related to information on shapes and other appearance-related characters of the respective species – can aid in fish identification, reliably to genus and often to species. Often, a worker has a preliminary idea of the genus and species of a specimen. Knowing where the fish was collected, identification of the specimen is facilitated by access to information on the particular species occurring in that watershed. Certain families of freshwater fishes, for example, the sculpins (Family *Cottidae*), contain a number of cryptic species that are difficult to differentiate. Species identification is supported by drawing spatial correlations among fish observations.

Our BIS tries to solve these problems, starting from the key-based approach, by creating a fish identification system that instead of being merely based on textual definitions, improves the fish identification process by allowing users to perform successive queries based on fish shape information, textual descriptions, and geographical data.

3.2 Present Implementation Stage

The *MBSC* and *CBISC* architecture components are fully operational for the goal application, being already tested though a Web-based application interface [11]. In particular, we have tested different kinds of queries on a collection of 11000 fish images [9]. The present version is to be used for fish identification in the Commonwealth of Virginia, and thus is restricted to fishes found in this area.

Our data collections comprise an image database (with fish photos); a geographic database (containing spatial data characterizing the regions in which the fish have been observed); and a database with metadata on the fish species and on the geographic data, dichotomous keys for identifying fishes and fish taxonomic trees. Metadata help query processing and are stored in a PostgreSQL database, while the current *CBISC* version manages image content description as XML documents.

We are now working on both implementing the BIS Manager modules with respect to the geographic data handling and organizing the Geo collection database. The latter uses PostgreSQL [36] database system and PostGIS [35]. PostGIS can be seen as an OpenGIS-conformant extension to the PostgreSQL, which allows geographic information systems objects to be stored into the database.

3.3 Data Sources

The fish-related data were obtained from [27] and from a site recently created to help students in the fish identification process [23]. Fish keys and metadata include data about over 200 species found in the Commonwealth of Virginia, USA. A subset of these data, covering 183 species and 187 images, is being used in this work.

Biodiversity Metadata: The biodiversity-related metadata include data about fish taxonomic classification (species, genus, family), common names, reproductive and food habits, metabolism, habitat description, information about similar species, and morphological descriptions.

Image Description: Current experiments configured *CBISC* to use several shape descriptors [2, 10]. We will further extend it to support queries on color information [43].

Geographic Data: The geographic data include maps (encoded in the ShapeFile format), spatial, and conventional data characterizing the regions in which the fish have been observed. Coordinates referring to the locations of occurrence of fish species also are stored. Data are being obtained from the Conservation Management Institute (CMI) at Virginia Tech. The CMI's Fish and Wildlife Information Exchange (FWIE) Division works as a technical assistance center, data analysis center, and information clearinghouse for fish, wildlife, and land management agencies and organizations.

3.4 Identifying a Specimen

An example of a query including textual, geographic data and image descriptor information is: *"retrieve fish descriptions of all fish whose shape is similar to that shown in Figure 2, which belong to genus 'Notropis', which have 'large eyes' and 'dorsal stripe', and have been observed within the catchments of the 'Tennessee' river"*. Notice that the first part of this query (shape similarity) is typical of image information systems; genus and physical characteristics are extracted from metadata-based systems; the last part is typical of GIS-based biodiversity systems (using the "within" spatial operation). The geographic component of the query is typically processed using a buffer operator or a user-specified rectangle encompassing part of the Tennessee drainage.

This query requires processing data from a variety of heterogeneous sources, stored in different formats. These sources are composed of images, image metadata and content descriptors, ecology-related data (species description and taxonomic trees), and geographical

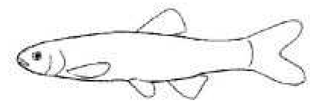

Figure 2: Example of shape outline used to define a query.

information (spatial data and metadata).

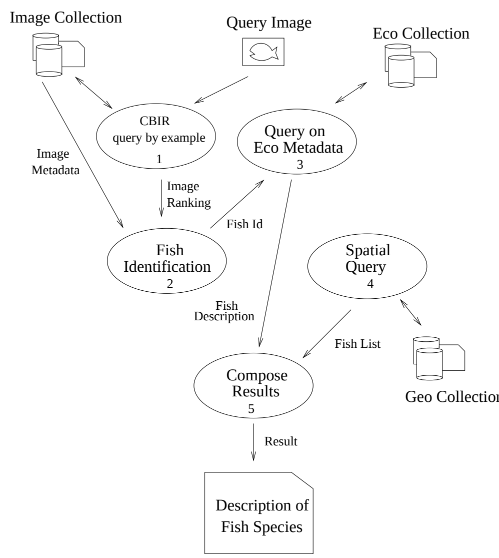

Figure 3: Execution plan for identifying fish species.

Figure 3 illustrates a possible execution plan for the proposed query within our BIS. It is composed of several steps, represented by ellipses. First, a content-based image retrieval process is executed in an image collection. Here, the *Execution* submodule of the *Query Mediator* (see Figure 1) performs a *GetImage* HTTP request on *CBISC*, using the image showed in Figure 2 as input and a pre-defined shape descriptor (1). CBISC will return a list of images, ranked by similarity to the input image.

The list of similar images is next used to retrieve fish identification parameters (2) for each image. Next, a query is performed on the *MBSC* to return fish species that belongs to genus “Notropis”, whose morphological description include terms like “large eyes” and “dorsal stripe”, and whose identification parameters match those returned by *CBISC* (3).

In the following, a spatial query is executed in order to identify which fishes have been observed within the catchments of the “Tennessee river” (4). This query is performed using a *GetFeature* HTTP request on the *GDSC*. By considering the rectangle-based query, this HTTP request might be encoded in XML as shown in Figure 4, where the query Filter is a box. The result of the query is a list of fish species’ names and scientific names (parameters Property Name).

Finally, the results of (3) and (4) are combined by the *Merging submodule* and the descriptions of the most relevant fish species are returned (5).

```
<wfs:GetFeature service="WFS" version="1.0.0"
  outputFormat="GML2"
  xmlns:topp="http://www.openplans.org/topp"
  xmlns:wfs="http://www.opengis.net/wfs"
  xmlns:ogc="http://www.opengis.net/ogc"
  xmlns:gml="http://www.opengis.net/gml"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.opengis.net/wfs
                  http://schemas.opengis.net/wfs/1.0.0/WFS-basic.xsd">
  <wfs:Query typeName="ekey:fishspecies">
    <wfs:PropertyName>ekey:scientific_name</wfs:PropertyName>
    <wfs:PropertyName>ekey:common_name</wfs:PropertyName>
    <ogc:Filter>
      <ogc:BBOX>
        <ogc:PropertyName>the_geom</ogc:PropertyName>
        <gml:Box srsName="http://www.opengis.net/gml/srs/epsg.xml#27345">
          <gml:coordinates>489154,5433017 505234,5448023</gml:coordinates>
        </gml:Box>
      </ogc:BBOX>
    </ogc:Filter>
  </wfs:Query>
</wfs:GetFeature>
```


Figure 4: Example of WFS XML request which fetches fish species (feature) with a bounding box filter.

This is, of course, one possible processing strategy. Another alternative would be to start with a spatial query that would limit the set of fish species to those observed within a certain range of the catchments. Next, content-based retrieval would be applied only to those species.

The existence of alternatives to query processing concern another issue, that of optimization. Our work is not yet concerned with performance aspects, and so assumes a predefined query processing strategy.

4 Related Work

The work proposed here involves combining research on image databases, geographic databases, and digital libraries for biodiversity information management. The following subsections outline related work in these areas.

Biodiversity Information Systems and GIS: There are several initiatives for the development of biodiversity information systems. Many of these initiatives are being linked to a worldwide project called GBIF [19] – Global Biodiversity Information Facility. GBIF intends to set up an interoperable network of biodiversity databases and database management tools that will allow Web users to navigate and query across these databases. Other initiatives are being conducted at smaller scales. Most of these systems are very new, and still under construction. Considerable effort is being applied to creating databases for species' taxonomic descriptions (e.g., [8, 15, 26]), and software on these databases, but still with little help from GIS (e.g., [16]).

Another trend is to process species' spatial distributions using GIS, for a more reduced set of species (e.g., [1, 3]). Efficient spatial data management and retrieval, query processing, interface design, and interoperability are among the many problems faced in the design and development of such systems. Spatial databases pose several research and implementation challenges (e.g., [22]). Some of these challenges are motivated by the intrinsic nature of the geographic data – they are location-sensitive and vary in time.

Another difficulty is that the spatial dimension introduces questions related to spatial integrity constraints and spatial query processing [37], involving topological, metric, and directional queries [22]. Our work is not concerned with solving specific problems within the geographic database realm. Rather, we have taken advantage of existing solutions in spatial query processing and combine them with our image processing mechanisms.

A particular issue faced by our approach is that of interoperability. Interoperability problems occur in the GIS context (e.g., [5, 17, 25, 29, 30, 33]). In fact, new geographic applications appear every day, and cover several space-time scales and distinct kinds of objects and phenomena. Moreover, the data are gathered in massive volumes, and proceed from different sources with distinct levels of generalization and incompatible scales.

Several approaches have been discussed to provide geographic systems interoperability and data integration/conversion. Our problem, however, is that of promoting interoperability across systems of different natures – i.e., textual descriptions, image content, and spatial data management. As far as we know, ours is the first proposal that promotes this kind of interoperability.

Image Databases: Image databases (e.g., [18, 32]) combine research on databases and image processing, involving problems that vary from storage issues to friendly interfaces [40]. Images are particularly complex to manage – besides the volume they occupy, retrieval is application- and context-dependent [38]. Even though many other content-based retrieval systems exist [18, 32, 44], they do not take advantage of the component philosophy. Thus, they are not easily amenable to reuse in distinct situations. Our proposal has the advantage of encapsulating CBIR functionality into a DL component, thereby ensuring its reusability and coupling to other DL-based systems.

Digital Libraries There are several DL initiatives that cover topics related to our research. One example is the digital museum of butterflies [24], which aims at building a digital collection of Taiwanese butterflies. This digital library contains 6 modules: XML-based information organization of digitized butterfly collections, content-based image retrieval of butterflies, synchronized multimedia exhibition, compositional FAQ, interactive games regarding butterfly ecosystems, and on-line courseware on butterflies. Queries based on butterfly spatial location are not supported. Another example is floristic digital libraries (FDL) [39]. These are distributed virtual spaces comprising botanical data repositories and a variety of services offered to library patrons to facilitate the use and extension of existing knowledge about plants. FDL uses an agent-based infrastructure to manage information about taxonomic keys, distribution maps, illustrations, and treatments (morphological descriptions). Content-based retrieval is not supported.

5 Conclusions

This paper presented results of an ongoing project for biodiversity information management, that combines work in image databases with that of geographic distribution of species and their ecosystems. Its originality lies not so much in solving issues in geographic or in image systems, but in providing a solution that combines features from both systems. It relies on a system architecture which extends spatial query processing with retrieval based on image content and textual descriptions, thereby proposing a new class of georeferenced queries. In this context, the main challenges to be considered are: the necessity of interaction mechanisms to allow users to easily formulate queries; the difficulty of combining mechanisms of content-based retrieval in image databases and queries of geographical databases; and the complexity of the management of such heterogeneous data. Images, metadata, and maps are stored in databases and are to be retrieved according to a set of predicates based on combining textual and visual descriptors of image content, spatial operators and metadata. A key issue in this architecture is that several query-processing techniques must be investigated, according to user profiles and to the way images and spatial data are preprocessed before being stored.

The solution proposed is based on using new or recently developed DL components. This architecture is easily extensible, and provides users a considerable degree of flexibility in data management. Furthermore, the implementation we provide complies with both digital library (e.g., OAI) and OpenGIS standards (e.g., GML and WFS). Our solution solves many current problems in this kind of system, allowing handling of images, geographic data, and textual information in an integrated fashion. This architecture was conceived to be applied to several domains. In order to show its feasibility, this paper describes a specific implementation of the architecture to build a fish species biodiversity information system. In particular, this system will be used by students in ichthyology courses at the Department of Fisheries and Wildlife Sciences, Virginia Tech.

Ongoing work concerns the investigation of query optimization techniques to speed up query evaluation across the different sources of evidence. For this part of the work we will take advantage of previous research and development conducted at the University of Campinas in biodiversity query processing [13].

Acknowledgements

This work was supported by FAPESP (Proc. 01/02788-7), CAPES (Proc. BEX844/03-9), the Brazilian WebMaps, PRONEX/SAI and AgroFlow projects, CNPq, and NSF (DUE-0121679 and ITR-0325579). Marcos André Gonçalves is supported by a fellowship from AOL.

References

[1] S. Analyst, 2004. http://www.speciesanalyst.net (as of October 2004).

[2] N. Arica and F. T. Y. Vural. BAS: A Perceptual Shape Descriptor Based on the Beam Angle Statistics. *Pattern Recognition Letters*, 24(9-10):1627–1639, June 2003.

[3] BioGIS. BioGIS - Israel Biodiversity Information System. http://www.eti.uva.nl/ (as of October 2004).

[4] BIOTA/FAPESP. SinBiota (Sao Paulo State's Biodiversity Information System). http ://www.biota.org.br/sia (as of October 2004).

[5] G. Camara, R. Thome, U. Freitas, and A. V. Monteiro. Interoperability in Practice: Problems in Semantic Conversion from Current Technology to OpenGIS. www.dpi.inpe.br/gilberto/papers/interop99.pdf (July 2004).

[6] P. Ciaccia, M. Patella, and P. Zezula. M-tree: An Efficient Access Method for Similarity Search in Metric Spaces. In *Proceedings of 23rd International Conference on Very Large Data Bases*, pages 426–435, Athens, Greece, 1997.

[7] CITIDEL. Computing and Information Technology Interactive Digital Educational Library, 2003. http://www.citidel.org (as of October 2004).

[8] CMI. Conservation Management Institute, http://fwie.fw.vt.edu/indexnew.html (as of October 2004).

[9] R. da S. Torres, C. G. da Silva, C. B. Medeiros, and H. V. da Rocha. Visual Structures for Image Browsing. In *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, pages 167–174, New Orleans, Lousiana, USA, November 2003. ACM Press.

[10] R. da S. Torres, A. X. Falcão, and L. da F. Costa. A Graph-based Approach for Multiscale Shape Analysis. *Pattern Recognition*, 37(6):1163–1174, June 2004.

[11] R. da S. Torres, E. Halleman, R. E. Jenkins, and N. M. Burkhead. EKEY, http://virginia.cc.vt.edu:8086/ekey (as of October 2004).

[12] R. da S. Torres, C. B. Medeiros, M. A. Goncalves, and E. A. Fox. An OAI Compliant Content-Based Image Search Component. In *Proceedings of the 4th Joint ACM/IEEE Conference on Digital Libraries*, page 418, Tucson, AZ, June 2004.

[13] S. de P. Matias. Query Processing in BIOTA Biodiversity Database. Master's thesis, Institute of Computing, UNICAMP, Campinas, SP, Brazil, 2000.

[14] ESSEX, 2004. http://br.endernet.org/∼akrowne/elaine/essex/index.html (as of October 2004).

[15] ETI. World Biodiversity Database, 2002. http://www.eti.uva.nl/Database/WBD.html (as of October 2004).

[16] ETI. Expert Center for Taxonomic Identification GCMD - The Linnaeus System, 2004. http://www.biogis.huji.ac.il/ (as of October 2004).

[17] R. Fileto. Issues on Interoperability and Integration of Heterogeneous Geographical Data. In *Proceedings of the Third Brazilian Symposium on Geoinformatics*, pages 129–136, 2001.

[18] M. Flickner, H. Sawhney, W. Niblack, Q. H. J. Ashley, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by Image and Video Content: the QBIC System. *IEEE Computer*, 28(9):23–32, Sep 1995.

[19] GBIF. Global Biodiversity Information Facility (GBIF). http://www.gbif.org (as of October 2004).

[20] GeoServer. http://geoserver.sourceforge.net (as of October 2004).

[21] GML. Geography Markup Language (GML). http://opengis.net/gml (as of October 2004).

[22] R. H. Guting. An Introduction to Spatial Database Systems. *The International Journal on Very Large Data Bases*, 4(3):357–399, October 1994.

[23] L. Helfrich, T. Newcomb, E. Halleman, and K. Stein. EFISH, http://www.cnr.edu/efish (as of October 2004).

[24] J. S. Hong, H. Chen, and J. Hsiang. A Digital Museum of Taiwanese Butterflies. In *Proceedings of the Fifth ACM Conference on Digital Libraries*, pages 260–261, San Antonio, Texas, United States, 2000. ACM Press.

[25] Interop. International Conference and Workshop on Interoperating Geographic Information Systems. Web site http://www.ncgia.ucsb.edu/conf/interop97, 12 1997. Site address valid as of 07/99.

[26] IUBS, CODATA, and IUMS. Species 2000. http://www.sp2000.org (as of October 2004).

[27] R. E. Jenkins and N. M. Burkhead. *Freshwater Fishes of Virginia*. American Fisheries Society, Bethesda, Maryland, 1993.

[28] C. Lagoze and H. V. de Sompel. The Open Archives Initiative: Building a Low-Barrier Interoperability Framework. In *Proceedings of the Joint Conference on Digital Libraries*, pages 54–62, Roanoke, VA, USA, June 2001.

[29] V. Mesev. Remote Sensing of Urban Systems: Hierarchical Integration with GIS. *Computers, Environment and Urban Systems*, 21(3/4):175–187, 1997.

[30] V. Noronha. Towards ITS Map Database Interoperability - Database Error and Rectification. *GeoInformatica*, 4(2):201–213, 2000.

[31] OAI. Open Archives Initiative, 2003. http://www. openarchives.org (as of October 2004).

[32] V. E. Ogle and M. Stonebraker. Chabot: Retrieval from Relational Database of Images. IEEE *Computer*, 28(9):40–48, Sep 1995.

[33] OpenGIS. Open GIS Consortium (OGC). http://opengis.org (as of October 2004).

[34] PlanetMath. http://planetmath.org/ (as of October 2004).

[35] PostGIS. http://postgis.refractions.net (as of October 2004).

[36] PostgreSQL. http://www.us.postgresql.org (as of October 2004).

[37] P. Rigaux, A. Voisard, and M. Scholl. *Spatial Databases. With Application to GIS*. Morgan Kaufmann, San Fracisco, CA, USA, 2001.

[38] Y. Rui, T. S. Huang, and S. F. Chang. Image Retrieval: Current Techniques, Promising Directions, and Open Issues. *Journal of Communications and Image Representation*, 10(1):39–62, March 1999.

[39] J. A. Sanchez, C. A. Flores, and J. L. Schnase. Mutant: Agents as Guides for Multiple Taxonomies in the Floristic Digital Library. In *Proceedings of the Fourth ACM Conference on Digital Libraries*, pages 244–245, Berkeley, California, USA, 1999. ACM Press.

[40] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.

[41] H. Suleman. *Open Digital Libraries*. PhD thesis, Computer Science Department, Virginia Tech, Blacksburg, VA, 2002. http:// scholar.lib.vt.edu/theses/available/etd-11222002-155624/unrestricted/odl.pdf.

[42] H. Suleman, E. A. Fox, R. Kelapure, A. Krowne, and M. Luo. Building Digital Libraries from Simple Building Blocks. *Online Information Review*, 27(5):301–310, 2003.

[43] M. Swain and D. Ballard. Color Indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[44] K. Vu, K. A. Hua, and W. Tavanapong. Image Retrieval Based on Regions of Interest. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):1045–1049, July/August 2003.

[45] WFS. Web Feature Server (WFS). http://www.opengis.org/specs (as of October 2004).