MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11 - 13, 2002, Nara - ken New Public Hall, Nara, Japan

3—19

An XML General Framework for Image Annotation, Storing and Retrieval

Maria Grazia Albanesi, Emanuele Alberto Benzi, Marco Ferretti
mariagrazia.albanesi@unipv.it, benzi@unipv.it, marco.ferretti@unipv.it
Dipartimento di Informatica e Sistemistica
University of Pavia
Via Ferrata 1, 27100 Pavia, Italy

Abstract

In this paper we describe a new framework based on XML for image characterization, description and storing for fast content-based retrieval. The novelty is mainly in the construction of a unified method for describing images and the related information extracted by means of a generic image processing algorithm. The XML description can be associated to well known image standards or to other customized application-dependent formats. On this level of image abstraction and description, a method for dynamic querying has been developed, which allows the semantic annotation of the image for a fast retrieval in the database. A prototype of the entire system for image description, storing and retrieval has been developed for GeoTIFF and DIG-35 image formats.

1 Introduction

Image search engines are nowadays really common over the Internet [1,6]. Almost every general-purpose search engine has its own section specialized on photo searches. Images are user-submitted or looked up on the Internet using web crawlers. Then images are usually indexed using keywords such as title, authors and so on. These approaches are probably the most effective solution for common users but are inadequate to be used in scientific contexts.

Dealing with scientific images (e.g. astronomical photos coming from observatories and satellites) is never a simple task because you never know before-hand which search criteria will be useful in the future so you never know which information you have to create indexes on. It's even harder creating a common front end for every need.

Continuous studies on database management and computer vision have given quite good results producing a number of standard image file formats, a huge set of data structures (to store images or their related information), and even ad hoc databases to create query by image content [2].

The biggest drawback of these methods (which are commonly used more or less obliviously in many research laboratories) is that they all require static formats, static data structures, custom software and custom databases.

2 A Flexible Annotation Architecture

We are proposing a completely new method based on a standard, general purpose and already widely used technology, that is XML (eXtensible Mark-up Language) [10]. This technology was designed to describe data focusing on what data are, not on how data look or on how to deal with them. XML is a mark-up language, so it puts no constraints on the meaning of the data it represents. By exploiting XML technology to create and run queries, we can create semantic queries; that is, we can focus on the meaning of the data we are querying and not on its representation.

The basic idea is to provide a framework to describe the elements in an images database. If it is possible to create metadata describing the images, then this framework should make it possible to use these data to query the database, and browse from an image to the other.

Structuring metadata in XML documents and storing these documents along with their schemas makes it possible to create interactive queries.

If someone or something (e.g. a pattern recognition algorithm) generates information about an image, then it can organize this information into an XML document (many software tools already provide XML as exporting format), creating an XML schema (to tell the framework how to read data) and generating an XML document for each image.

When users submit queries, they are prompted to choose query criteria based on the schemas contained in the database. Schemas will be read by a software that can transform them into HTML pages with combo boxes, radio buttons or check boxes so user friendly that the user could be oblivious to the underlying XML schema. This way we can also add new information and even new schemas at run time, without having to re-design any software or query because we are only adding new XML documents.

The general framework we have designed is described in Figure 1. It is based on an annotated image database and on an XML based infrastructure that maps various forms of user interaction onto the stored dataset; the technology items involved are the XML schemas, that store the semantic criteria for the annotation, image converters that map existing formats into the schema annotated ones, image processing functions that produce metadata for schema instantiations, and schema for end user querying.

3 Images and Metadata

On a lower level of abstraction, our annotation architecture is based on the extended use of metadata.

Metadata are usually defined as data about data, that is data used to locate and manage other data. A typical example of this is found in the libraries: we usually search a book by Title, Author, Keywords. That is we search the main information (the book) using metadata (Title etc.).

The same kind of metadata are used to index and retrieve images in general purpose databases. For example Internet search engines store images along with their original file name, their caption[1] and the context (that is the text around the image). Dealing with scientific images is somehow more complex because many types of metadata can be found. We classified scientific metadata as follows:

Context based: Information that can only be stored during acquisition (e.g. georeference [3]). It can refer to the image or to its subject.

Classifying: Produced manually by an operator, they are used to classify the subject of the image and usually are based on a standard dictionary (e.g. biological classification).

Computed: Produced by image processing algorithm (e.g. color distribution).

Metadata are not a novelty in digital imaging [6]: almost every file format provides ad hoc space to store additional information about the image. The standard approach is to define a set of possible information fields (often referred to as TAGs) to be stored in the file; these tags are then assigned a standard label or a fixed offset in the file structure.

The obvious drawback of this simple approach is the scarce flexibility: the meaning of the information fields we want to use has to be known by the application we use to read the file. Even in those file formats where custom tags can be defined[2], their meaning and their structure cannot be stored along with the tag value and must be known beforehand by the application. Moreover if two users produce the same custom tag, they cause a conflict.

A solution to this problem is the use of XML to store metadata. This way the value of the tag is stored along with its structure definition in the form of a DTD or a Schema. This structure definition uses a standard syntax to describe custom data structures, allowing every program, application or search engine to read and understand even not standard information.

This innovative approach is also adopted by international organizations such as the Digital Imaging Group that proposed XML as a standard format to store metadata inside image files [4]. This proposal applies both to the new coming JPEG 2000 and to the standard JPEG for which a new marker (APP4 application container, XML data container) has been defined.

The standardization effort of the Digital Imaging Group is a good example of how a standard can be flexible using XML technologies: every time a new feature was added, a new version of the standard was required, but earlier version schemas are still available on line to help everyone and everything to read the existent files.

Generalizing this example, we can understand that creating a new set of metadata becomes an easy job using XML: we only have to publish on the Internet the schema of our metadata and store the URL[3] of the schema in every image that contains our set of metadata. It doesn't matter if the metadata are context based, classifying or computed. They are just metadata.

1 Retrieved from the HTML tag <IMG ALT="caption">

2 For example JPEG markers and TIFF tags

3 In XML documents we usually refer to URIs [9] instead of URLs meaning a generalization of the Uniform Resource Location. For example a URI of a file could be its ID inside a database, not just it's Internet address.

4 XML and Databases

The use of XML as a transport and storing medium for metadata raises the issue of indexing and data retrieval. Actually the XML document we use to store metadata could be viewed as a data itself. So the question is: if we store the document in a database, how do we manage its content?

The information we can retrieve from an XML document is not only the content of the various elements, but also the structure of the document itself, the schemas it is linked to and the order of the elements[4].

These complex data are therefore stored and retrieved using a new database engines based on a semi-structured data model. This model manages the self-describing data and let the user access data ad every level of the structure (document, elements, attributes of the elements). The databases designed on this model are often referred to as Native XML Databases (NXD) [7].

The NXDs are really important nowadays because they are replacing traditional databases in many common tasks. Their efficiency and speed is therefore always increasing and new indexing algorithms are developed continuously.

5 Current Framework

Within this project we have succeeded in developing the query engine that creates the user interface starting from the XML schema. The documents are stored in a XML Native Database that creates indexes based on XML elements. The software developed is a collection of dynamic HTML. Only a small part of these pages depends on the database we chose, while the most part is platform independent. At the client side we need nothing but a XML enabled browser. This gives the maximum flexibility to the user. Using this query engine, we can let the user search for images even if they are annotated by not yet existing metadata collections. Figures 2 and 3 show the interface of the environment for image querying and retrieval. To test the environment, a small image database of GeoTIFF annotated images has been mapped into the new structure; also, a few samples of images stored according to the DIG-35 specifications have been imported.

6 Future Work

A common platform is currently under development to let algorithms compute metadata from the images. When an algorithm is stored, it should read every image in the database, one by one, and compute related metadata. The algorithm should be run in the background and offline. New technologies such as DOM, SAX and all those connected to semantic web and XML should be evaluated in order realize the best platform [8].

The project aims at building a set of tools for the semantic annotations of images and an easy environment for semantic querying. In the latter field, we are currently investigating solutions that allow to express queries against a variety of different annotations classes: it will be the task of the system to map the query onto the existing semantically annotated XML repositories, with the proper

4 It is useful to remember that the fields and records are not ordered in the traditional relational databases.

transformations in order to match the diverse, yet corresponding annotation classes.

References

[1] Image Retrieval: Current Techniques, Promising Directions and Open Issues - Y. Rui, T.S. Huang, S.G. Chang - in Journal of Visual Communication and Image Representation, 10 - 1999

[2] Multimedia Database Management, Requirements and Issues - D.A. Adjeroh, K.C. Nwosu - IEEE Multimedia – July September 1997

[3] GeoTIFF Format Specification Revision 1.5 (DRAFT) - Niles Ritter, Mike Ruth - geoTIFF.org - 14 June 2001

[4] DIG 35 Specification (Metadata for Digital Images) Version 1.1 - Digital Imaging Group, Inc. - www.i3a.org - 18 June 2001

[5] Using Metadata to Simplify Digital Photography - James Milch, Ken Parulski - Eastman Kodak Company - 2000

[6] Semantic Models for Multimedia Database Searching and Browsing - Shu-Ching Chen, R.L. Kashyap, Arif Ghafoor - Kluwer Academic Publishers - June 2000

[7] Mapping DTDs to Databases - Ronald Bourret - http://www.xml.com - 9 May 2001

[8] The Semantic Web - Tim Berners-Lee, James Hendler, Ora Lassila - Scientific American - May 2001

[9] Uniform Resource Identifiers (URI) - Tim Berners-Lee (World Wide Web Consortium) - http://www.ietf.org/rfc/rfc2396.txt - August 1998

[10] Extensible Markup Language (XML)1.0 Recommendation -World Wide Web Consortium (W3C) - http://www.w3.org/TR/1998/REC-xml-19980210 - 10 February 1998

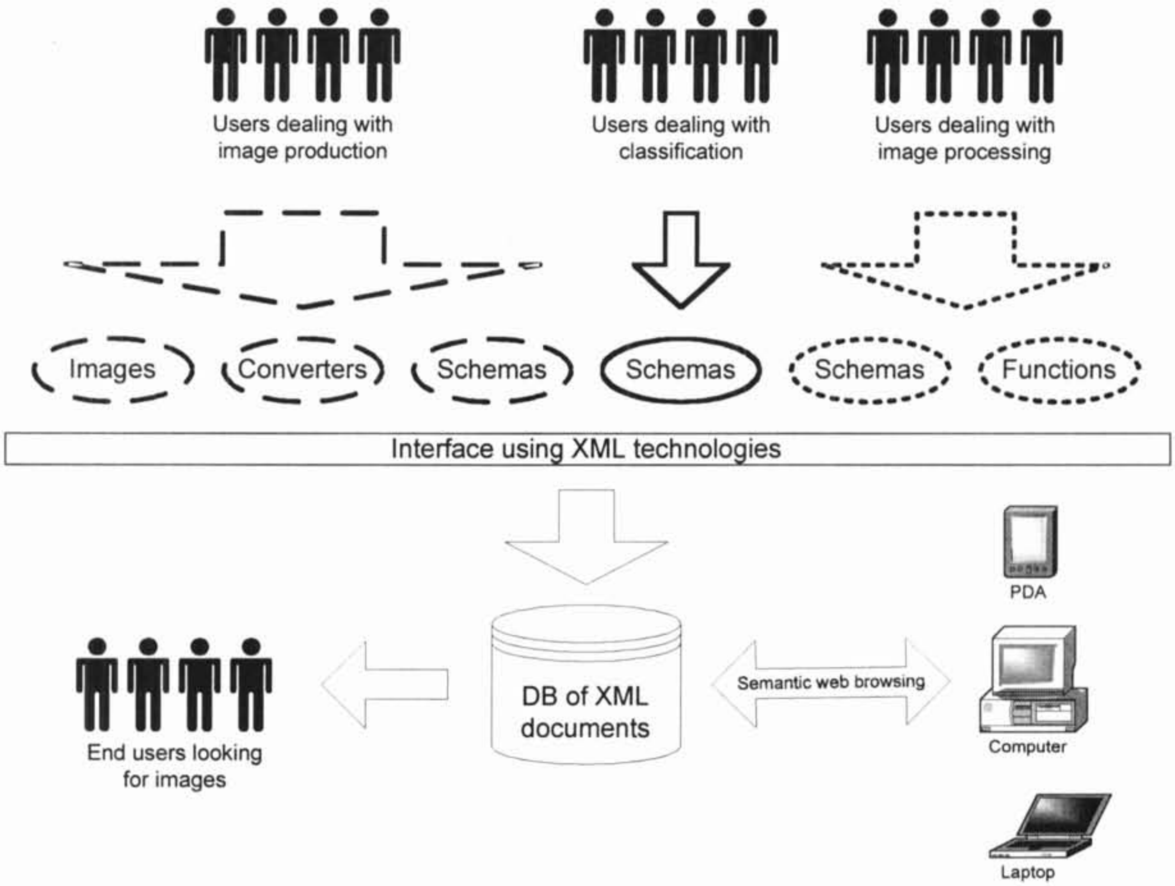

Figure 1: Project overview with every entity dealing with images and documents

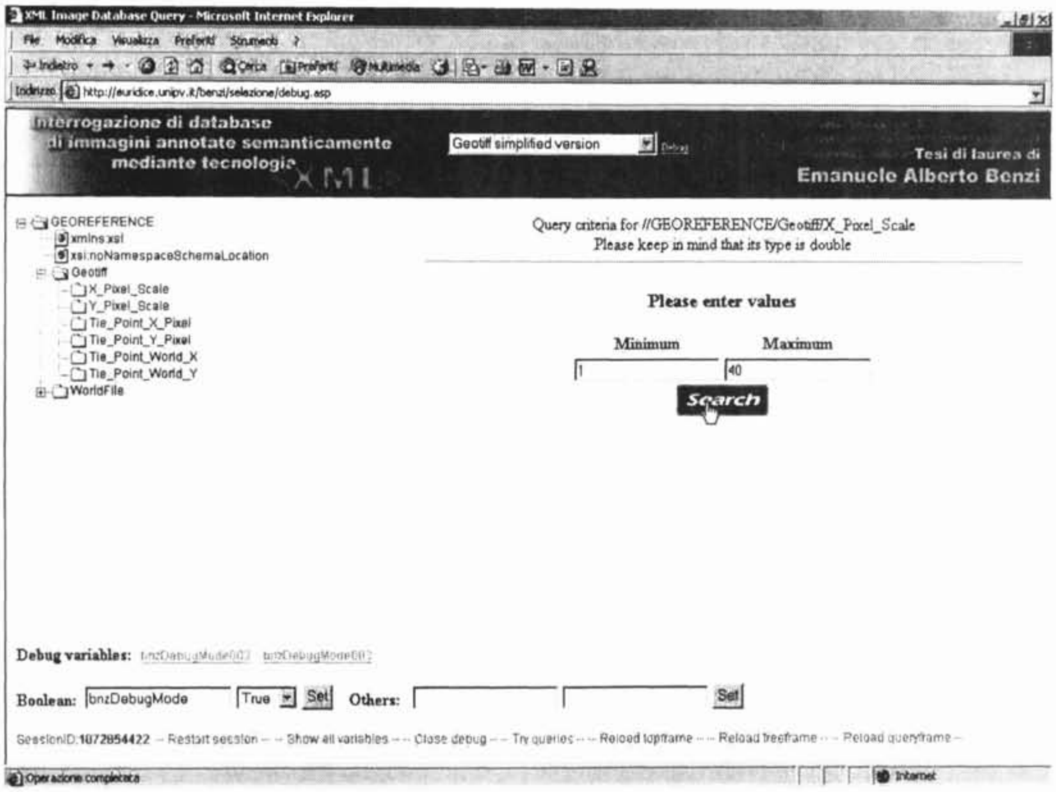

Figure 2: A snapshot of the interface; the schema on the left and the dynamic form on the right

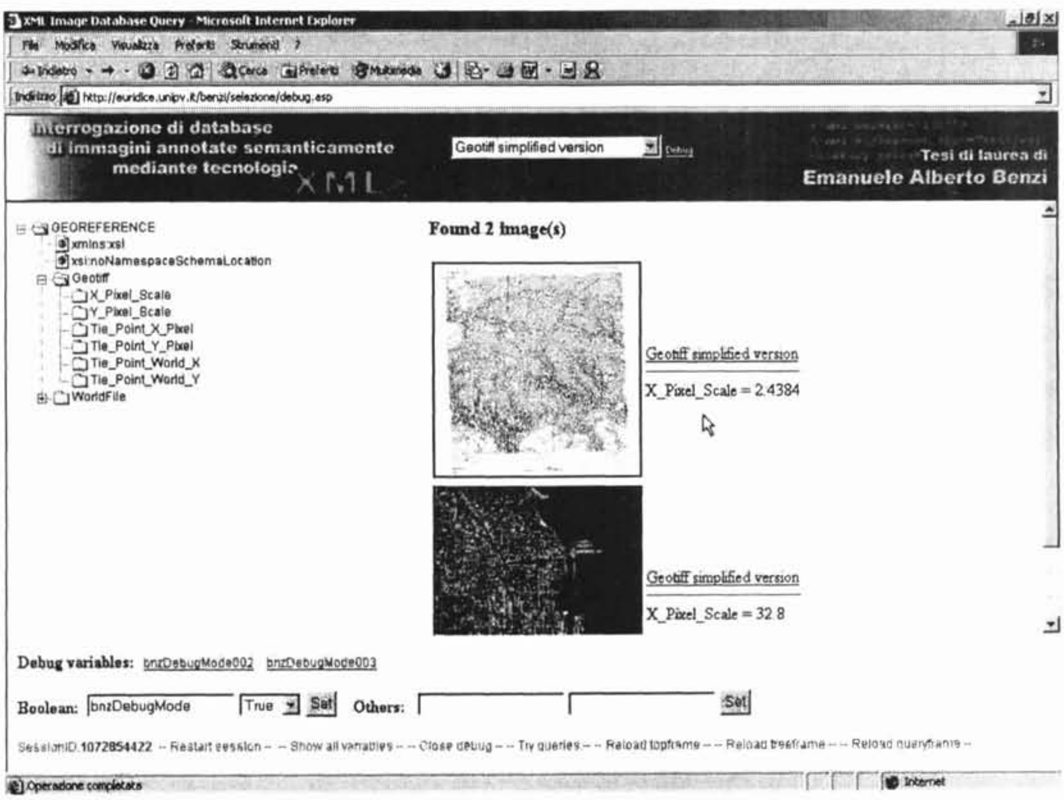

Figure 3: A snapshot of the interface with the result of a dynamic query on a GeoTIFF image database