An OAI Compliant Content-Based Image Search Component

Ricardo da S. Torres and
Claudia Bauzer Medeiros
IC, University of Campinas, Campinas, SP, Brazil
{rtorres,cmbm}@ic.unicamp.br

Marcos André Gonçalves and
Edward A. Fox
Virginia Tech, Blacksburg, VA, USA
{mgoncalv,fox}@vt.edu

ABSTRACT

Advances in data storage and image acquisition technologies have enabled the creation of large image datasets. In order to deal with these data, appropriate information systems (e.g., image digital libraries) have been developed to efficiently manage such collections. One of the most common retrieval approaches is to employ 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*.

This demonstration presents a new CBIR system based on configurable components. The main novelty resides in its *Content-Based Image Search Component (CBISC)* that supports queries on image collections. *CBISC* is based on the OAI [3, 4] principles, and thus provides an easy-to-install search engine to support querying images by content. As with the OAI protocol, queries are posed via HTTP requests and the responses are encoded in XML.

Figure 1: *CBISC Configuration Tool* screen shot.

CBISC encapsulates multidimensional index structures [1] to speed up the search process. Furthermore, it supports the use of different image descriptors (metric and non-metric; color, texture, and shape descriptors; with 1D or 2D feature vectors) [2], which can be easily combined to yield improved

effectiveness. We will show that this search component can be tailored for particular image collections by a trained designer, who carries out a clearly defined set of pilot experiments to select the appropriate descriptors.

Image descriptors are typically domain and usage-dependent. Further, a given image can be associated with very many descriptors. However, standard *CBIR* methods only support a fixed set of descriptors. *CBISC*, instead, allows progressive extension of the descriptor base.

Figure 1 presents a screen shot showing the *CBISC Configuration Tool* developed to support *CBISC* designers in the configuration process. Basically, this process concerns the description/definition of both the image descriptors that will be used to retrieve images by content, and the image database to which the *CBISC* is related. The XML file generated in this process is used during *CBISC* execution.

Categories and Subject Descriptors

H.3.7 [**Information Systems**]: Information Storage and Retrieval—*Digital Libraries*

General Terms

Design, Experimentation

Keywords

Content-Based Image Retrieval, OAI

1. REFERENCES

[1] P. Ciaccia, M. Patella, and P. Zezula. M-tree: An Efficient Access Method for Similarity Search in Metric Spaces. In *23th VLDB Conference*, pages 426–435, 1997.

[2] R. da S. Torres, A. X. Falcão, and L. da F. Costa. A Graph-based Approach for Multiscale Shape Analysis. *Pattern Recognition*, 2004. To appear.

[3] OAI. Open Archives Initiative, 2003. http://www.openarchives.org (as of December 2003).

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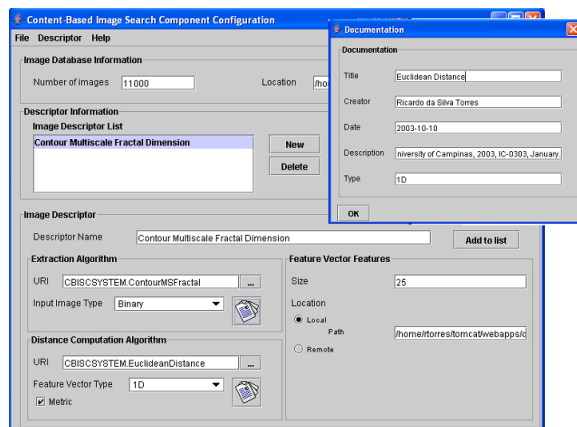*Online Information Review*, 27(5):301–310, 2003.