IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Hershey PA 17033-1117, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITB7267

Chapter III

Content-Based Visual Information Retrieval

Oge Marques and Borko Furht
Florida Atlantic University, USA

This chapter provides a survey of the state-of-the-art in the field of Visual Information Retrieval (VIR) systems, particularly Content-Based Visual Information Retrieval (CBVIR) systems. It presents the main concepts and system design issues, reviews many research prototypes and commercial solutions currently available, and points out promising research directions in this area.

INTRODUCTION

The amount of audiovisual information available in digital format has grown exponentially in recent years. Gigabytes of new images, audio and video clips are generated and stored everyday, building up a huge, distributed, mostly unstructured repository of multimedia information, much of which can be accessed through the Internet.

Digitization, compression, and archiving of multimedia information have become popular, inexpensive and straightforward, and there is a broad range of available hardware and software to support these tasks. Subsequent retrieval of the stored information, however, might require considerable additional work in order to be effective and efficient.

There are basically three ways of retrieving previously stored multimedia data:

1. **Free browsing**: users browse through a collection of images, audio, and video files, and stop when they find the desired information.
2. **Text-based retrieval**: textual information (metadata) is added to the audiovisual files during the cataloguing stage. In the retrieval phase, this additional information is used to guide conventional, text-based query and search engines to find the desired data.
3. **Content-based retrieval**: users search the multimedia repository providing information about the actual contents of the image, audio, or video clip. A content-based search engine translates this information in some way as to query the database and retrieve the candidates that are more likely to satisfy the users' requests.

This chapter appears in the book, Distributed Multimedia Databases: Techniques and Applications by Timothy K. Shih.

The first two methods have serious limitations and scalability problems. Free browsing is only acceptable for the occasional user and cannot be extended to users who frequently need to retrieve specific multimedia information for professional applications. It is a tedious, inefficient, and time-consuming process and it becomes completely impractical for large databases.

Text-based retrieval has two big problems associated with the cataloguing phase:

a) the considerable amount of time and effort needed to manually annotate each individual image or clip; and
b) the imprecision associated with the subjective human perception of the contents being annotated.

These two problems are aggravated when the multimedia collection gets bigger and may be the cause of unrecoverable errors in later retrieval.

In order to overcome the inefficiencies and limitations of text-based retrieval of previously annotated multimedia data, many researchers, mostly from the Image Processing and Computer Vision community, started to investigate possible ways of retrieving multimedia information – particularly images and video clips – based solely on its contents. In other words, instead of being manually annotated using keywords, images and video clips would be indexed by their own visual content, such as color, texture, objects' shape and movement, among others.

Research in the field of Content-Based Visual Information Retrieval (CBVIR) started in the early 1990s and is likely to continue during the first decade of the 21st century. Many research groups in leading universities and companies are actively working in the area, and a fairly large number of prototypes and commercial products are already available. Current solutions are still far from reaching the ultimate goal, namely to enable users to retrieve the desired image or video clip among massive amounts of visual data in a fast, efficient, semantically meaningful, friendly, and location-independent manner.

FUNDAMENTALS OF CBVIR SYSTEMS

Preliminaries

Visual Information Retrieval (VIR) is a relatively new field of research in Computer Science and Engineering. As in conventional information retrieval, the purpose of a VIR system is to retrieve all the images (or image sequences) that are relevant to a user query while retrieving as few non-relevant images as possible. The emphasis is on the retrieval of *information* as opposed to the retrieval of *data*. Similarly to its text-based counterpart, a visual information retrieval system must be able to interpret the contents of the documents (images) in a collection and rank them according to a degree of relevance to the user query. The interpretation process involves extracting (semantic) information from the documents (images) and using this information to match the user's needs (Baeza-Yates and Ribeiro-Neto, 1999).

Progress in visual information retrieval has been fostered by many research fields (Figure 1), particularly: (text-based) information retrieval, image processing and computer vision, pattern recognition, multimedia database organization, multidimensional indexing, psychological modeling of user behavior, man-machine interaction, among many others.

VIR systems can be classified in two main generations, according to the attributes used to search and retrieve a desired image or video file (Del Bimbo, 1999):