# Content-Based Image Retrieval

**Timo R. Bretschneider**
*Nanyang Technological University, Singapore*

**Odej Kao**
*University of Paderborn, Germany*

## INTRODUCTION

Sensing and processing multimedia information is one of the basic traits of human beings: The audiovisual system registers and transports surrounding images and sounds. This complex re-cording system, complemented by the senses of touch, taste, and smell, enables perception and provides humans with data for analysing and interpreting the environment. Imitating this perception and the simulation of the processing was and still is one of the major leitmotifs of multimedia technology developments. The goal is to find a representation for every type of knowledge, which makes the reception and processing of information as easy as possible. The need to process given information, deliver it, and explain it to a certain audience exists in nearly all areas of day-to-day life: commerce, science, education, and entertainment (Smeulders, Worring, Santini, Gupta, & Jain, 2000).

The development of digital technologies and applications allowed the production of huge amounts of multimedia data. This information has to be systematically collected, registered, organised, and classified. Furthermore, search procedures, methods to formulate queries, and ways to visualise the results have to be provided. In early years, this task was tended to by existing database management systems (DBMS) with multimedia extensions. The basis for representing and modelling multimedia data is so-called binary large objects, which store images, video, and audio sequences without any formatting and analysis done by the system. Often, however, only a reference to the object is handled within the DBMS. For the utilisation of the stored multimedia data, user-defined functions (e.g., content analysis) access the actual data and integrate their results in the existing database. Hence, content-based retrieval becomes possible. A survey of existing retrieval systems was presented, for example, by Naphade & Huang (2002).

This article provides an overview of the complex relations and interactions among the different aspects of a content-based retrieval system, whereby the scope is purposely limited to images. The main issues of the data description, similarity expression, and access are addressed and illustrated for an actual system.

## BACKGROUND

The concept of content-based retrieval is datacentric per se; that is, the design of a system has to reflect the characteristics of the data. Hence, neither an optimal solution that can span all kinds of multimedia data exists nor is addressing the variety of data characteristics within one type even possible. However, there are parallels that lay the foundation, which then require tailor-made adaptation and specialisation. This section provides the general groundwork by pointing out the different types of the so-called metainformation, which describes the raw data:

- Technical information refers to the details of the recording, conversion, and saving process (i.e., format and name of the stored media).
- Extracted attributes are those that have been deduced by analysing the media content. They are usually called *features* and emphasise a certain aspect of the media. Simple features describe, for instance, statistical values of the contained information, while complex features and their weighted combinations attempt to describe the entire media content.
- Knowledge-based information links the objects, people, scenarios, and so forth, detected in the media to entities in the real world.
- World-oriented information encompasses information on the producer of the media, the date, location, and so forth. Manually added keywords are especially in this group, which makes a primitive description and characterisation of the content possible.

As can be seen by this classification, technical and world-oriented information can be modelled straightforwardly in traditional database structures. Organising and searching can be done by using existing database functions. The utilisation of the extracted attributes and knowledge-based information is more complex in nature. Although most of the currently avail-able DBMSs can be extended with multimedia add-ins, in many cases these are not sufficient, because they cannot describe the stored data to the required degree of retrieval accuracy. However, only these two latter types of metainformation lift the system to an abstract level that allows the full exploitation of the content.

For an in-depth overview of content-based image retrieval techniques and systems, refer to Deb and Zhang (2004), Kalipsiz (2000), Smeulders et al. (2000), Vasconcelos and Kunt (2001), and Xiang and Huang (2000).

## MAIN THRUST

The goal of multimedia retrieval is the selection of one or more images whose metainformation meets certain requirements or is similar to a given sample media instance. Searching the metainformation is usually based on a full-text search among the assigned keywords. Furthermore, content references, such as colour distributions in an image, or more complex information, such as wavelet coefficients, can be used. To solve the issue of having the desired search characteristic in the first place, most systems prefer to use a query with an example media item. The systems use this media as a starting point for the search and processed it in the same manner as the other media objects, when they were inserted in the database. The content is then analysed with the selected procedures, and the media is mapped to a vector consisting of (semi-) automatically extracted features. Hereafter, the raw data is only needed for display purposes, and all further processing focuses on analysing and comparing the representative vectors. The result of this comparison is a similarity ranking. The following interfaces can be used to specify a query in a multimedia database:

- Browsing: Beginning with a predefined data set, the user can navigate in any desired direction by using a browser until a suitable media sample is found. This approach is often used when no suitable starting media is available.
- Search with keywords: Technical and world-oriented data are represented by alphanumerical fields. These can be searched for a given keyword. Choosing these keywords is extraordinarily difficult for abstract structures such as textures, partially due to the subjectivity of the human perception.
- Similarity search: The similarity search is based on comparing features extracted from the raw data. Most of these features do not exhibit immediate references to the image, making them highly abstract for users without special knowledge (Assfalg, Del Bimbo, & Pala, 2002; Brunelli & Mich, 2000). Depending on the availability and characteristic of the query medium, one differentiates between query by pictorial example, query by painting, selection from standards, and image montage.

All approaches have their individual advantages as well as disadvantages, and a suitable selection depends on the domain. For example, a fingerprint database is best realised by using the query-by-pictorial-example technique, but selection from standards is a suitable candidate for a comic strip data-base with a limited number of characters. However, the similarity search, in particular the query-by-pictorial-example approach, is one of the most powerful methods because it provides the greatest degree of flexibility. Thus, it determines the focus hereafter.

Many different methods for feature extraction were developed and can be classified by various criteria. Based on the point in time in which the features are extracted, a-priori and dynamically extracted features are distinguished. Although the first group was extracted during insertion of the corresponding media object in the database, the latter kind is generated at query time. The advantage of the dynamic feature extraction is that the user can define relevant elements in the sample image, and the remaining parts of the query image do not distract the actual search objective. Note that both approaches can be combined.

Regardless of the chosen approach, the actual features have to be extracted from the considered data. Examples for this step are histogram-based methods, calculation of statistical colour information (Mojsilovic, Hu, & Soljanin, 2002), contour descriptors (Berretti, Del Bimbo, & Pala, 2000), texture analysis (Gevers, 2002), and wavelet coefficient selection (Albuz, Kocalar, & Khokhar, 2001). The gained information — possibly from different algorithms — is combined in a so-called feature vector that is, by orders of magnitude, smaller than the raw data. This reduction in volume enables not only a suitable handling within the DBMS but also a higher level of abstraction. Therefore, it can often be utilised directly by semantic-based approaches (Djeraba, 2003; Fan, Luo, & Elmagarmid, 2004; Lu, Zhang, Liu, & Hu, 2003) and data-mining techniques (Datcu, Daschiel, & Pelizzari, 2003; Li & Narayanan, 2004).

The similarity of two multimedia objects in the content-based retrieval process is determined by comparing the representing feature vectors. Over the years, a large variety of metrics and similarity functions was developed for this pur-pose, whereby the best-known methods compute a multi-dimensional distance between the vectors: The smaller the distance, the higher the similarity of the corresponding media objects. Through the introduction of weights for the individual positions within the feature vectors, it is possible to emphasise and/or suppress desirable and undesirable query characteristics, respectively. In particular, the approach can help to particularise the query in iterative retrieval systems; that is, if the users select suitable and unsuitable retrievals, which are used by the system for adaptation in the next iteration (Jing, Li,

## Related Content