Content–Based Image Retrieval Query Paradigms

Colin C. Venters
University of Manchester, UK

Richard J. Hartley
Manchester Metropolitan University, UK

William T. Hewitt
University of Manchester, UK

INTRODUCTION

The proliferation in volume of digital image data has exacerbated the general image retrieval problem, creating a need for efficient storage and flexible retrieval of vast amounts of image data (Chang, 1989). Whilst there have been significant technological advances with image data capture and storage, developments in effective image retrieval have not kept pace. Research in image retrieval has been divided into two areas: concept-based image retrieval and content-based image retrieval. The former focuses on the use of classification schemes or indexing terms to retrieve images while the latter focuses on the visual features of the image, such as colour, shape, texture, and spatial relationships.

The field of content-based image retrieval has been a thriving area of research and development for over a decade. Its origins lie in the subject areas of artificial intelligence, computer vision, image processing, information retrieval, pattern recognition, and signal processing. Advancement of the approach has been attributed to the early experiments conducted by Kato, Kurita, and Shimogaki (1991) into the automatic retrieval of images by colour and shape feature (Eakins & Graham, 1999). The approach focuses on the semiautomatic or automatic extraction, indexing, and retrieval of images by their visual attributes and characteristics. The process involves a direct matching operation between a query image and a database of stored images, where a feature vector or global signature has been computed for the unique visual properties of an image. Similarity is calculated by comparing the vector or signature of the query image against those of images stored in a database. The visual distance D between images can be measured by functions such as Euclidean distance, L_1 and L_2 distances, weighted Euclidean distance, city-block distance, and Minkowsky distance (Figures 1 to 6).

Figure 1. Euclidean distance

$$d(s_1, s_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Figure 2. Weighted Euclidean distance

$$D_H(I_Q, I_D) = (H(I_Q) - H(I_D))^t A(H(I_Q) - H(I_D))$$

Figure 3. L_1 distance

$$D_H(I_Q, I_D) = \sum_{j=1}^{n} \left| H(I_Q, j) - H(I_D, j) \right|$$

Figure 4. L_2 distance

$$D_H(I_Q, I_D) = \left(\sum_{j-1}^{n} \left(H(I_Q, j) - H(I_D, j) \right)^2 \right)$$

Figure 5. City-Block distance

$$d(s_1, s_2) = \|x_2 - x_1\| + \|y_2 - y_1\|$$

Figure 6. Minkowsky distance

$$d(s_1, s_2) = \sum_{k-1}^{n} |x_k - y_k|$$

The result of this process is a quantified similarity score that measures the visual distance between the two images, represented by the feature vectors, in a 2-D feature space (Del Bimbo, 1999). The most common image-matching feature characteristics are colour, shape, and texture.

Research in the field of content-based image retrieval has investigated only a small number of the problems inherent in the development of visual information management systems, with a considerable amount of attention being directed at the development and advancement of retrieval algorithms (Venters & Cooper, 2000). However, many research topics in this field, which are considered fundamental to the advancement of content-based image retrieval as a viable retrieval tool, have been largely ignored. One such area is the user interface.

BACKGROUND

The user interface provides the bridge between the end user and the system and is crucial to effective human-computer interaction. Retrieval is commonly obtained through an interactive session (Del Bimbo, 1999). To initiate a search, the user provides or creates the visual representation of their information need and then selects the features, range of model parameters, and similarity measure that are important.Smeulders, Worring, Santini, Gupta, and Jain (2000) defined an abstract query space to represent user interaction in a content-based image retrieval system (Figure 7).

Figure 7. Abstract query space (Smeulders et al., 2000)

$$Q = \{I_Q, F_Q, S_Q, Z_Q\}$$

To start a query session, an instantiation of $Q = \{I_Q, F_Q, S_Q, Z_Q\}$ is created. Where I_Q is the selection of images from the image archive I, F_Q is the selection of features derived from the images in I_Q, S_Q is the similarity function, and Z_Q is the goal-dependent semantics. The query space forms the basis of user interaction in a content-based image retrieval system, specifying queries and displaying results. Similarly, Rui and Huang (2001) state that an image object O can be modelled as a function of the image data D, features F, and representations R (Figure 8).

Figure 8. Image object model (Rui & Huang, 2001)

$$O = O(D, F, R)$$

Where D is the raw image data, $F = \{f_i\}$, $i = 1, ..., I$ is a set of visual features associated with the image object, such as colour, texture, or shape, and $R_i\{r_{ij}\}$, $j = 1, ..., J_i$ is a set of representations for a given feature, such as colour histogram or colour moments. Each representation of r_{ij} is a vector consisting of multiple components where K_{ij} is the length of the vector r_{ij}, i.e., $r_{ij} = \lfloor r_{ij_1}, ..., r_{ij_k}, ..., r_{ij}k_{ij} \rfloor$. In general, the majority of content-based image retrieval systems exemplify these interaction models although research activity has only explored specific elements. The models emphasise that user interaction is a complex interplay between the user and the system, suggesting that only highly skilled and specialised users would be able to utilise the system effectively to obtain relevant retrieval results.

VISUAL QUERY FORMULATION

Query formulation is a core activity in the process of information retrieval and a number of paradigms have been proposed for specifying a query Q_n in an image database I. Several database query languages have been proposed to support image query expression in both first- and second-generation image retrieval systems and are based on either the relational or tabular data models. The most widespread languages are extensions of the Structured Query Language (SQL) proposed by Codd (1970). With this approach the user formulates the query and the system retrieves a set of results. Examples include ISQL (Assmann, Venema, & Hohne, 1986), PROBE (Orenstein & Manola, 1988), Spatial SQL (Egenhofer, 1991), and $\sum QL$ (Chang & Jungert, 2001). A major criticism levelled at these languages is the potential complexity of the query statement required to express a simple query (Catarci, Costabile, Levialdi, & Batin, 1997).

As an alternative to the relational-based query language approach, Zloof (1977) proposed Query by Example (QBE). This is based on a tabular data model and employs a tabular query structure to express queries. The user specifies an example output of the query by directly constructing entries into relational skeletons instead of writing lengthy queries, and the system identifies the goal by generalizing an example. QBE is a revision of Codd's relational calculus language SQL. Several examples have been based on this approach, including Aggregate by Example (Klug, 1981), Time by Example (Tansel, Arkun, & Ozsoyoglu, 1989), Generalized Query by Example (Jacobs & Walczak, 1983), Office by Example (Whang et al., 1987), and Natural Forms Query Language (NFQL; Embley, 1989). The languages share a commonality in the method of query expression and differ only in their expansion or revision of the QBE paradigm to serve different application areas. Two languages specifically developed for image databases were Query by Pictorial Example (Chang & Fu, 1980) and PICQUERY (Joseph & Cardenas, 1988).

Content-based image retrieval requires the user to provide or create a visual example or representation of