II. BACKGROUND

A. CBIR Systems

Content-based image retrieval system (CBIR) provides different ways of searching the database, enabling searches based on features such as texture, shape, and color. Such feature-based searches can be combined with searches for textual information. The CBIR systems are based on a similarity search that ranks the images in the database based on a computable measure for their similarity to a chosen image. Similarity searches often involve user interaction, whereby the user provides feedback on the relevance of the search results by selecting a different feature, or modifying the weight of certain features. Fig. 1 shows a typical architecture of a content-based image retrieval system. Two main functionalities (which take a time-consuming task) are supported: data insertion and query processing. The data insertion subsystem is responsible for extracting appropriate features from images and storing them into the image database (see dashed modules and arrows). This process is usually performed off-line [27]. The query processing, in turn, is organized as follows: the interface allows a user to specify a query by means of a query pattern and to visualize the retrieved similar images. The query-processing module extracts a feature vector from a query pattern and applies a metric (such as the Euclidean distance) to evaluate the similarity between the query image and the database images. Next, it ranks the database images in a decreasing order of similarity to the query image and forwards the most similar images to the interface module.

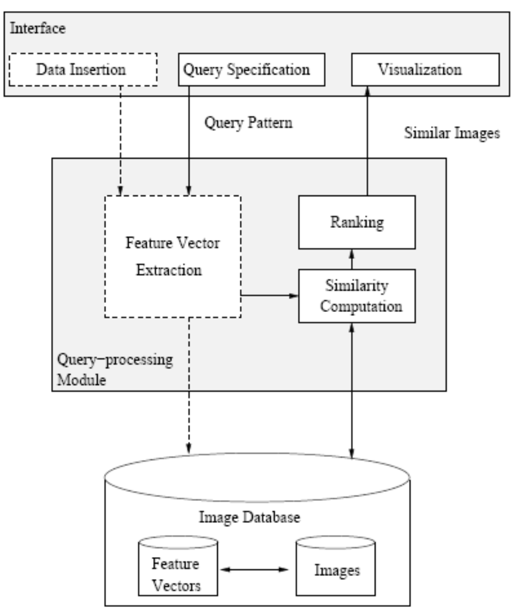

Figure 1. Typical architecture of a content-based image retrieval system [27].

B. Feature Extraction

Image features affect every aspect of a CBIR system, and so it is important to carefully choose the right image features for any CBIR system. Feature extraction and representation is the fundamental process behind the content-based image retrieval systems. Most of the CBIR systems explore low-level image features such as color, texture, shape, motion, and so on because they can be computed automatically.

1) Texture Feature

There is no widely accepted definition of texture. However, this image property can be characterized by the existence of basic primitives, whose spatial distribution creates some visual patterns defined in terms of granularity, directionality, and repetitiveness. There exist different approaches to extract and represent textures. They can be classified into space-based, frequency-based models, and texture signatures [9]. Some of methods such Co-occurrence matrix is one of the most traditional techniques for encoding texture information [10].

2) Shape Feature

In pattern recognition and related areas, shape is an important characteristic to identify and distinguish objects [11, 12, 13, 14]. Standard mathematical distances are used to calculate the degree of similarity between two shapes. They can be classified into three different features such as: Simple Geometric Features, Spatial-Domain Features, and Transform-Domain Features [15]. In the first group, contour-based features can be simple scalar values extracted from the boundary itself. The Second group used the spatial information to extract the shape feature. For example, Chain Codes technique [16], the direction vectors between successive boundaries pixels are encoded. The last group of feature is based on transformation-domain such as Fourier transform and Wavelet transform.

3) Color Feature

Color is one of the most prominent perceptual features. Most commercial CBIR systems include color as one of the features [17]. Color histograms [19] are amongst the most traditional techniques for describing the low-level properties of an image. Although, several color description techniques have been proposed [18], [19], [20], [21] they can be grouped into two classes based on whether or not they encode information related to the color spatial distribution.

C. Similarity Measurement

This functionality is very take time to execute the visual vector feature of query image and database image, depend on the number of image data base. In general, the type of similarity measure to be considered depends on the technique used for feature extraction. Swain and Ballard proposed the use of an L_1 norm, and argued that it had a useful interpretation as a measure of intersection when used with histograms[25]. Alternatively, L_2 norms have also been applied [22], [23], and Lq norms are an obvious generalization. Any of these metrics may be generalized by preprocessing the feature vector with a linear transformation. This leads to the special cases of L_1 norms of the cumulative histogram [23], or weighted L_2 norms [22].

D. Distributed processing

Distributed processing is a phrase used to refer to a variety of computer systems that use more than one computer (or processor) to run an application. This includes parallel processing in which a single computer uses more than one CPU to execute programs. More often, however, distributed processing refers to local-area networks (LANs) designed so that a single program can run simultaneously at various sites. Most distributed processing systems contain sophisticated software that detects idle CPUs on the network and parcels out programs to utilize them. Another form of distributed processing involves distributed databases. These are the type of databases in which the data is stored across two or more computer systems. The database system keeps track of where the data is stored so that the distributed nature of the database is not apparent to users.