# Chapter IV
# Content-Based Image Classification and Retrieval:
## A Rule-Based System Using Rough Sets Framework

**Jafar M. Ali**
*Kuwait University, Kuwait*

## ABSTRACT

*Advances in data storage and image acquisition technologies have enabled the creation of large image datasets. Thus, it is necessary to develop appropriate information systems to efficiently manage these datasets. Image classification and retrieval is one of the most important services that must be supported by such systems. The most common approach used is content-based image retrieval (CBIR) systems. This paper presents a new application of rough sets to feature reduction, classification, and retrieval for image databases in the framework of content-based image retrieval systems. The suggested approach combines image texture features with color features to form a powerful discriminating feature vector for each image. Texture features are extracted, represented, and normalized in an attribute vector, followed by a generation of rough set dependency rules from the real value attribute vector. The rough set reduction technique is applied to find all reducts with the minimal subset of attributes associated with a class label for classification.*

## INTRODUCTION

There is a pressing need for efficient information management and mining of the huge quantities of image data that are routinely being used in databases (Cios, Pedrycz, & Swiniarski, 1998; Laudon, & Laudon, 2006; Starzyk, Dale, & Sturtz, 2000). These data are potentially an extremely valuable source of information, but their value is limited unless they can be effectively explored and retrieved, and it is becoming increasingly clear that in order to be efficient, data mining must be based on semantics. However, the extraction of semantically rich metadata from computationally accessible low-level features poses tremendous scientific challenges (Laudon & Laudon; Mehta, Agrawal, & Rissanen, 1996; Mitra, Pal, & Mitra, 2002). Content-based image retrieval (CBIR) systems are needed to effectively and efficiently use the information that is intrinsically stored in these image databases. The image retrieval system has gained considerable attention, especially during the last decade. Image retrieval based on content is extremely useful in many applications (Carson, Thomas, Belongie, Hellerstein, & Malik, 1999; Huang, Tan, & Loew, 2003; Koskela, Laaksonen, & Oja, 2004; Ma & Manjunath, 1999; Molinier, Laaksonen, Ahola, & Häme, 2005; Smeulders, Worring, Santini, Gupta, & Jain, 2000; Smith, 1998; Viitaniemi & Laaksonen, 2006; Yang & Laaksonen, 2005) such as crime prevention, the military, intellectual property, architectural and engineering design, fashion and interior design, journalism and advertising, medical diagnosis, geographic information and remote sensing systems, cultural heritage, education and training, home entertainment, and Web searching. In a typical CBIR system, queries are normally formulated either by example or similarity retrieval, selecting from color, shape, skeleton, and texture features or a combination of two or more features. The system then compares the query with a database representing the stored images. The output from a CBIR system is usually a ranked list of images in order of their similarity to the query.

Image searching (Graham, 2004) is one of the most important services that need to be supported by such systems. In general, two different approaches have been applied to allow searching on image collections: one based on image textual metadata and another based on image content information. The first retrieval approach is based on attaching textual metadata to each image and uses traditional database query techniques to retrieve them by keyword. However, these systems require a previous annotation of the database images, which is a very laborious and time-consuming task. Furthermore, the annotation process is usually inefficient because users, generally, do not make the annotation in a systematic way. In fact, different users tend to use different words to describe the same image characteristic. The lack of systematization in the annotation process decreases the performance of the keyword-based image search. These shortcomings have been addressed by so-called content-based image retrieval. In CBIR systems, image processing algorithms are used to extract feature vectors that represent image properties such as color, texture, and shape (Hassanien & Ali, 2004; Ma & Manjunath, 1999; Viitaniemi & Laaksonen, 2006). In this approach, it is possible to retrieve images similar to one chosen by the user (i.e., query by example). One of the main advantages of this approach is the possibility of an automatic retrieval process, contrasting with the effort needed to annotate images.

The work introduced in this article is based on the second retrieval approach. Image similarity is typically defined using a metric on a feature space. Numerous similarity metrics have been proposed so far. The search results are combined with existing textual information and collections of other features via intelligent decision-support systems. In this article, we use a new similarity function based on the rough set theory (Grzymala-Busse, Pawlak, Slowinski, & Ziarko, 1999; Kent, 1994; Pawlak, 1982, 1991; Pawlak, Grzymala-Busse, Slowinski, & Ziarko, 1995). This theory has become very popular among scientists around the world. Rough sets data analysis was used for the discovery of data dependencies, data reduction, approximate set classification, and rule induction from databases. The generated rules represent the underlying semantic content of the images in the database. A classification mechanism is developed by which the images are classified according to the generated rules.