Chapter x
Classification and Retrieval of Images from Databases Using Rough Set Theory

Aboul Ella Hassanien
Cairo University, Egypt

Jafar M. Ali
Kuwait University, Kuwait

ABSTRACT

This chapter presents an efficient algorithm to classify and retrieve images from large databases in the context of rough set theory. Color and texture are two well-known low-level perceptible features to describe an image contents used in this chapter. The features are extracted, normalized, and then the rough set dependency rules are generated directly from the real value attribute vector. Then the rough set reduction technique is applied to find all reducts of the data which contains the minimal subset of attributes that are associated with a class label for classification. We test three different popular distance measures in this work and find that quadratic distance measures provide the most accurate and perceptually relevant retrievals. The retrieval performance is measured using recall-precision measure, as is standard in all retrieval systems.

INTRODUCTION

The growth of the size of data and number of existing databases far exceeds the ability of humans to analyze this data, which creates both a need and an opportunity to extract knowledge from databases. 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, &
Classification and Retrieval of Images from Databases Using Rough Set Theory

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 meta-data from computationally accessible low-level features poses tremendous scientific challenges (Laudon & Laudon, 2006; Mehta, Agrawal, & Rissanen, 1996; Mitra, Pal, & Mitra, 2002).

Content-based image classify and retrieval (CBICR) systems are needed to effectively and efficiently use the information that is intrinsically stored in these image databases. This image retrieval system has gained considerable attention, especially during the last decade. Image retrieval based on content is extremely useful in many applications (Smith, 1998; Molinier, Laaksonen, Ahola, & Häme, 2005; Yang & Laaksonen, 2005; Koskela, Laaksonen, & Oja, 2004; Viitaniemi & Laaksonen, 2006; Huang, Tan, & Loew, 2003; Smeulders, Worring, Santini, Gupta., & Jain, 2000; Ma & Manjunath, 1999; Carson, Thomas, Belongie, Hellerstein, & Malik, 1999) 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 query by example or similarity retrieval, selecting from a 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 classification (Hassanien & Dominik 2007) is an important data mining task which can be defined as a task of finding a function that maps items into one of several discrete classes. The most commonly used techniques in classification are neural network [Dominik et. al. 2004, Hassanien & Dominik 2007], genetic algorithms [Satchidananda et. al., 2008], decision trees [Yang et. al., 2003], fuzzy theory [Ashish G., Saroj K. Meher, & Uma B. Shankar 2008], multi-resolution wavelet [Uma et. al., 2007] and rough set theory [Hassanien & Ali, 2004]. Rough set concept was introduced by Polish logician, Professor Zdzislaw Pawlak in early eighties [Pawlak, Z. 1982]. This theory becomes very popular among scientists around the world and the rough set is now one of the most developing intelligent data analysis [Slowinski, 1995, Pawlak, 1995, Pawlak, 1991]. 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.

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 classification and retrieval. In CBICR systems, image processing algorithms are used to extract feature vectors that
Related Content

Interaction of Agent in E-Business: A Look at Different Sources
[www.igi-global.com/chapter/interaction-agent-business/19618?camid=4v1a](www.igi-global.com/chapter/interaction-agent-business/19618?camid=4v1a)

Agent Cognitive Capabilities and Orders of Social Emergence
[www.igi-global.com/chapter/agent-cognitive-capabilities-orders-social/19615?camid=4v1a](www.igi-global.com/chapter/agent-cognitive-capabilities-orders-social/19615?camid=4v1a)

Multi-Agent–Based Simulation of University Email Communities
[www.igi-global.com/chapter/multi-agent-based-simulation-university/49368?camid=4v1a](www.igi-global.com/chapter/multi-agent-based-simulation-university/49368?camid=4v1a)

Norm and Social Compliance: A Computational Study
[www.igi-global.com/article/norm-social-compliance/39032?camid=4v1a](www.igi-global.com/article/norm-social-compliance/39032?camid=4v1a)