Chapter VI
Image Classification and
Retrieval with Mining
Technologies

Yu-Jin Zhang
Tsinghua University, Beijing, China

ABSTRACT

Mining techniques can play an important role in automatic image classification and content-based retrieval. A novel method for image classification based on feature element through association rule mining is presented in this chapter. The effectiveness of this method comes from two sides. The visual meanings of images can be well captured by discrete feature elements. The associations between the description features and the image contents can be properly discovered with mining technology. Experiments with real images show that the new approach provides not only lower classification and retrieval error but also higher computation efficiency.

INTRODUCTION

Along with the progress of imaging modality and the wide utility of digital image (include video) in various fields, many potential content producers have emerged, and many image databases have been built. In addition, the growth of Internet and storage capability not only increasingly makes images a widespread information format in World Wide Web (WWW), but also dramatically expands the number of images on WWW and makes the search of required images more complex and time consuming. To efficiently search images on WWW, effective image search engines need to be developed.

Since images require large amounts of storage space and processing time, how to quickly and efficiently access and manage these large, both in the sense of information contents and data volume, databases has become an urgent problem to solve. The research solution for this problem, using content-based
Image retrieval (CBIR) techniques, is initiated in the last decade (Kato, 1992). An international standard for multimedia content descriptions, MPEG-7, is also formed in 2001 (MPEG). With the advantages of comprehensive descriptions of image contents and consistence to human visual perception, research in this direction is considered as one of the hottest research points in the new century (Castelli, 2002; Zhang, 2003; Deb 2004; Zhang 2007).

Among the many research topics in CBIR, automatic image classification (categorization) plays an important role both for Web image searching and retrieving (classification and retrieval are closely related), as it is time consuming for users to browse through and treat the huge data on Web. A successful image classification will significantly enhance the performance of the content-based image retrieval system by filtering out images from irrelevant classes during matching. Classification has been used to provide access of large image collections with more efficient manner because the classification can reduce search space by filtering out the images in unrelated category (Hirata, 2000).

The heterogeneous nature of Web images makes their classification a challenge task. A functional classification scheme should take the contents of images in consideration. Web mining is a tool suitable for helping image classification and retrieval on the Web. It consists of (Scime, 2005):

1. **Pre-processing:** It is one of the most important steps in Web mining. It includes data purging, user recognition, dialog recognition, and event recognition.
2. **Pattern discovering (Mining algorithm):** It uses statistical analysis, association rule, clustering, and classification.
3. **Pattern analysis:** It transforms the rules, patterns and statistical values into knowledge. By using this knowledge, valuable patterns (interesting rules, patterns) can be obtained.

Traditional mining techniques often generate huge amounts of numeric data that could be difficult to interpret and use. Visual mining transforms raw data into visualization and makes it easier to understand the meaning of data and make suitable decisions, in addition to opening the world of visual tools to a much broader audience (Soukup, 2002).

In this chapter, a novel method for image classification based on feature element through association rule mining is presented. The feature elements can capture well the visual meanings of images according to the subjective perception of human beings. In addition, feature elements are discrete entities, and are suitable for working with rule-based classification models. Different from traditional image classification methods, the proposed classification approach based on feature element does not compute the distance between two vectors in the feature space. This approach just tries to find associations between the feature elements and class attributes of the image. Techniques for mining the association rules are adapted and the mined rules are applied to image classifications. Experiments with real images show that the new approach not only reduces the classification errors but also diminishes the time complexity.

The contents of following sections are:

1. In background section, some concepts and definitions about feature-based image retrieval, association rules and rule mining, and classification based on association are introduced.
2. In main thrust section, some techniques and results on feature elements and extraction, database used for testing, feature element based image classification, image classification comparison, feature element based image retrieval, image retrieval comparison, and association feedback are presented in details.