Semantics Retrieval by Content and Context of Image Regions*

Wei Wang, Yuqing Song and Aidong Zhang
Department of Computer Science and Engineering
State University of New York at Buffalo
Buffalo, NY 14260 USA

Abstract

We propose a novel approach for semantics retrieval from images in multimedia databases. In our approach, we use color-texture classification to generate the codebook which is used to segment images into regions. The content of a region describes the lower-level features of the region, including color and texture. The context of regions in an image describes their relationships in the image. The content and context of image regions provide a way for semantics retrieval. On top of semantics retrieval, high-level (semantics-based) querying and query-by-example are supported. The experimental results demonstrate that our approach outperforms the traditional CBIR approaches.

1 Introduction

Although content-based image retrieval (CBIR) techniques based on low-level features such as color, texture, and shape have been extensively explored, their effectiveness and efficiency are not satisfactory. The ultimate goal of image retrieval is to provide the users with the facility to manage large image databases in an automatic, flexible and efficient way. Therefore, image retrieval systems should be armed to support high-level (semantics-based) querying and browsing of images. The basic elements to carry semantic information are the image regions which correspond to semantic objects, if the image segmentation is effective. After the regions are obtained, proper representation of their content and context remains a challenge. The extraction of content and context of image regions is a necessary step for semantics retrieval.

Many methods have been proposed for region-based image retrieval. Under keyblock model [4], images are partitioned into equal-sized blocks and features are extracted by training blocks to form the codebook. Images are indexed using the codebook. Image retrieval is performed based on the indexed images. The equal-sized blocks ignore the boundary of regions, therefore they cannot represent the objects correctly. Another image querying system developed by Smith et al [7] first decomposes an image into regions with characterizations pre-defined in a finite pattern library. With every pattern labeled by a symbol, images are represented by region strings. Region strings are converted to descriptor matrices of composite region templates(CRT), which reflect the relative ordering of symbols. The CRT library is solely dependent on color feature. If texture or shape features are added to distinguish patterns, the size of library will increase dramatically which makes the retrieval process inefficient. Li et al proposed IRM [3] which allows matching a region of one image to several regions of another image. That is, the region matching between two images is a many-to-many relationship. As a result, the similarity between two images is defined as the weighted sum of distances in the feature space between all regions from different images. All the approaches introduced above could not integrate the semantic descriptions into the regions (or blocks), therefore cannot support the high-level querying of images.

Semantics retrieval from images can be performed by annotating the images. In [9, 8], the monotonic tree is used as a hierarchical representation of image structures. Microstructure called *structural elements* are classified and clustered using methods based on minimum spanning tree. The images are rendered with semantic annotation keywords. The system performs well on scenery images. However, it does not take the context of image regions into consideration.

In this paper, we propose a novel approach for semantics retrieval from images based on the content and context of image regions. Our method consists of three levels. At pixel level, color-texture classification is used to form the semantic codebook. At region level, the semantic codebook is used to segment the images into regions. At image level, the content and context of image regions are defined and represented to support the semantics retrieval from images. The content of a region describes the lower-level features of the region, including color and texture. The context of regions in an image describes their relationships in the image. The three levels are illustrated in Figure 1.

The remainder of the paper is organized as follows. From Section 2 to Section 4, each step of our approach is elaborated. In Section 5 experimental results will be pre-

*This research was supported by NSF Digital Government Grant EIA-9983430.