Chapter 24
Feature Evaluation and Classification for Content-Based Medical Image Retrieval System

Ivica Dimitrovski
Ss. Cyril and Methodius University in Skopje, Macedonia

Suzana Loskovska
Ss. Cyril and Methodius University in Skopje, Macedonia

ABSTRACT

Image retrieval in general and content-based image retrieval (CBIR) in particular are well-known research fields in information management. A large number of methods have been proposed and investigated in both areas but satisfactory general solution has not still been developed. The aim of this research is to develop highly flexible web-based system for storage, organization and retrieval of medical images. The system besides text and metadata retrieval also supports querying by image to find visually similar images to presented query. Several algorithms and techniques were implemented in the system to support content-based retrieval. For efficient and reliable search machine learning techniques were included in the system.

INTRODUCTION

The number of digital images is rapidly increasing, prompting the necessity for efficient image storage and retrieval systems. The management and the indexing of these large image and information repositories are becoming increasingly complex. Therefore, tools for efficient archiving, browsing and searching images are required.

DOI: 10.4018/978-1-61520-777-0.ch024

A straightforward way of using the existing information retrieval tools for visual material, is to annotate records by keywords and then to use the text-based query for database retrieval. Several approaches were proposed to use keyword annotations for image indexing and retrieval (Datta, 2008). These approaches are not adequate, since annotating images by textual keywords is neither desirable nor possible in many cases. Therefore, new approaches of indexing, browsing and retrieval of images are required.

For very large image databases, manual description and annotation of every image is time-consuming and impractical. Rather than relaying on manual indexing and text description for every image, images can be represented by numerical features extracted directly from the image pixels. These features are stored in the database, as a signature together with the images and are used to measure similarity between the images in the retrieval process. This approach is known as Content-based Image Retrieval (CBIR).

The aim of CBIR systems is searching and finding similar multimedia items based on their content. Every CBIR system considers offline indexing phase and online content-based retrieval phase. The visual contents of the database images are extracted and described by multidimensional feature vectors in offline phase. The feature vectors of the database images form the feature database. In the second or online retrieval phase, the query-by-example (QbE) paradigm is commonly used. The user presents a sample image, and the system computes the features vector for the sample, compares it to those vectors for images stored in the database, and returns all images with similar features vectors. The query provided by the user can be a region, a sketch or group of images.

The quality of response depends on the image features and the distance or similarity measure used to compare features of different images. Regarding the features, different approaches are used but the most common for image content representation are color, shape and texture features.

Content-based image retrieval can be applied in various areas (Datta, 2008). The medicine is one of the most prospective application areas, because the growing number of digital image acquisition equipment in hospitals such as: X-ray, computed tomography (CT), magnetic resonance imaging (MR), positron emission tomography (PET), ultrasound, endoscopy and laparoscopy raises demands for new approaches of storing and accessing images. Therefore, the tasks of efficiently storing, processing and retrieving medical image data have become important research topic.

The medical community proposes inclusion of content-based image retrieval into picture archiving and communication systems (PACS) (Muller, 2003). The PACS integrates imaging modalities and interfaces with hospital and departmental information systems to manage storage and distribution of images to medical personal, researchers, clinics, and imaging centers. The important requirement of PACS is an efficient search function to access required images. A well known standard for medical images is Digital Imaging and Communications in Medicine (DICOM). File header for every image stored in DICOM format, contains information regarding image modality, acquisition device, and patient identification. At present, image search in medicine is carried out according to the alphanumerical order of these textual attributes. But users commonly have the visual content of medical images as starting information. The content of the images is a powerful query which can be used to search for images with similar content. Therefore, content-based approaches are expected to have a great impact on PACS, Radiology Information Systems (RIS), Pathology Information Systems and Hospital Information Systems (HIS). Even medical image databases that are not part of any PACS can benefit from CBIR technology, because CBIR technology supports any work that requires finding images or collections of images with similar contents. For example, medical researchers can use CBIR system to find images with similar pathological areas and investigate their association. Medical student will be provided by a system that enables efficient search for images with given pathological attributes. Additionally, CBIR can be used to collect images for medical books, reports, papers, and CD-ROMs where typical specimens are organized according to the similarity of their features.

This chapter describes the highly flexible web-based system for storing, organizing and retrieving medical images. The chapter is organized as fol-