A CONTENT-BASED APPROACH TO IMAGE DATABASE RETRIEVAL

K.Vijay, Senior Lecturer
MCA, K.S.R.College of Technology,
Tiruchengode.

Dr. R. Anitha, Director
MCA, K.S.R.College of Technology,
Tiruchengode.

Abstract

Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision to the image retrieval problem, that is, the problem of searching for digital images in large databases. It makes use of image features, such as color and texture, to index images with minimal human intervention. Content-based image retrieval can be used to locate medical images in large databases. Fundamentals of the key components of content-based image retrieval systems are introduced first to give an overview of this area. This is intended to disseminate the knowledge of the CBIR approach to the applications of medical image management.

Introduction

Initial image retrieval techniques were text-based that associated textual information, like filename, captions and keywords with every image in the repository. For image retrieval, keyword based matching was employed for finding the relevant images.

Soon it was realized that image retrieval based on the contents is a natural and an effective way of retrieving images. This led to the development of Content-Based Image Retrieval (CBIR). The Content-Based Image Retrieval (CBIR) is aimed at efficient retrieval of relevant images from large image databases based on automatically derived image features.

Hospitals take X-rays and MRI scans for patients producing hundreds of digital images each day. In order to facilitate easy access in the future, each image is registered in a medical image database based on the modality, region, and orientation of the image.

Large-scale image databases collect various images, including X-ray, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound (US), nuclear medical imaging, endoscopy, and microscopy. The most important aspect of image database management is how to effectively retrieve the desired images using a description of image content. This approach of searching images is known as content-based image retrieval (CBIR), which refers to the retrieval of images from a database using information directly derived from the content of images themselves, rather than from accompanying text or annotation

The main purpose of this paper is to disseminate the knowledge of the CBIR approach to the applications of medical image retrieval. This paper is organized as follows: First the problems and challenges of medical image retrieval and describes potential applications of medical CBIR. Next, the existing medical CBIR systems. Following that the key components of content-based image retrieval systems for medical imaging applications.

Image Database Retrieval

This section will discuss the problems of image retrieval using the conventional text-based method and addresses the challenges of the CBIR approach. Potential applications of the CBIR approach will also be discussed.

Challenges in Medical Image Retrieval

Before the emergence of content-based retrieval, medical images were annotated with text, allowing the images to be accessed by text-based searching. Through textual description, medical images can be managed based on the classification of imaging modalities, regions, and orientation. This hierarchical structure allows users to easily navigate and browse the database. Searching is mainly carried out through standard Boolean queries.

However, with the emergence of massive image databases, the traditional text-based search suffers from the following limitations:

• Manual annotations require too much time and are expensive to implement. As the number of images in a database grows, the difficulty in finding desired images increases.

• Manual annotations fail to deal with the discrepancy of subjective perception. The phrase, “an image says more than a thousand words,” implies that the textual description is not sufficient for depicting subjective perception.

• The contents of medical images are difficult to be concretely described in words. For example, irregular organic shapes cannot easily be expressed in textual form, but people may expect to search for images with similar contents based on the examples they provide.

These problems limit the feasibility of text-based search for medical image retrieval. In an attempt to overcome these difficulties, content-based retrieval has been proposed to automatically access images with minimal human intervention. However, due to the nature of medical images, content-based retrieval for medical images is still faced with challenges:

Medical Applications of Content-Based Image Retrieval

Content-based image retrieval has frequently been proposed for various applications. This section

will discuss three potential applications of medical CBIR.

PACS/Health Database Management

Content-based image retrieval has been proposed by the medical community for inclusion into Picture Archiving and Communication Systems (PACS). The idea of PACS is to integrate imaging modalities and interfaces with hospital and departmental information systems in order to manage the storage and distribution of images to radiologists, physicians, specialists, clinics, and imaging centers. A crucial point in PACS is to provide an efficient search function to access desired images. In addition to PACS, medical imaging databases that are unconnected to the PACS can also obtain benefits from CBIR technology.

Computer-Aided Diagnosis

Computer-aided diagnosis has been proposed to support clinical decision making. One clinical decision-making technique is case-based reasoning, which searches for previous, already- solved problems similar to the current one and tries to apply those solutions to the current problem. This technique has a strong need to search for previous medical images with similar pathological areas, scrutinize the histories of these cases which are valuable for supporting certain diagnoses, and then reason the current case.

Medical Research, Education, and Training

CBIR technology can benefit any work that requires the finding of images or collections of images with similar contents. In medical research, researchers can use CBIR to find images with similar pathological areas and investigate their association.

In medical education, lecturers can easily find images with particular pathological attributes, as those attributes can imply particular diseases. In addition, CBIR can be used to collect images for medical books, reports, papers, and CD-ROMs based on the educational atlas of medical cells, where typical specimens are collected according to the similarity of their features, and the most typical ones are selected from each group to compose a set of practical calibrators.

Existing Medical CBIR Systems

Although content-based image retrieval has frequently been proposed for use in medical image management, only a few content-based retrieval systems have been developed specifically for medical images. These research-oriented systems are usually constructed in research institutes and continue to be improved, developed, and evaluated over time. This section will introduce several major medical content-based retrieval systems.

CasImage

• The CasImage system, which has been integrated into a PACS environment, contains a teaching and reference database, and the medGIFT retrieval system, which is adapted from the open-source GIFT (GNU Image Finding Tool). Combinations of textual labels and visual features are used for medical image retrieval.

IRMA (Image Retrieval in Medical Applications)

• The IRMA system is implemented as a platform for content-based image retrieval in medical applications

• This system splits the image retrieval process into seven consecutive steps, including categorization, registration, feature extraction, feature selection, indexing, identification, and retrieval.

Content-Based Retrieval Systems

Content-based retrieval uses the contents of images to represent and access the images. A typical content-based retrieval system is divided into off-line feature extraction and online image retrieval. A conceptual framework for content-based image retrieval is illustrated in Figure 1. In offline feature extraction, the contents of the images in the database are extracted and described with a multi-dimensional feature vector, also called descriptor.

The feature vectors of the image constitute a feature dataset stored in the database. In online image retrieval, the user can submit a query example to the retrieval system in search of desired images. The system represents this example with a feature vector. Finally, the system ranks the search results and then returns the results that are most similar to the query examples. If the user is not satisfied with the search results, the user can provide relevance feedback to the retrieval system, which contains a mechanism to learn the user's information needs. The following sections will clearly introduce each component in the system.

Feature Extraction

Representation of images needs to consider which features are most useful for representing the contents of images and which approaches can effectively code the attributes of the images. Feature extraction of the image in the database is typically conducted off-line so computation complexity is not a significant issue. This section will introduce two features-texture and color-which are used most often to extract the features of an image.

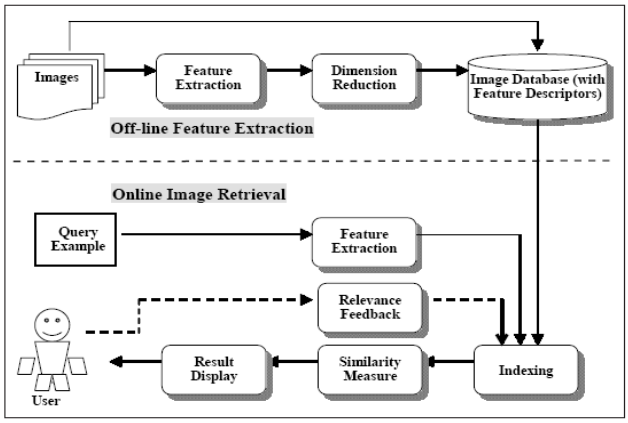

Fig.1 A conceptual framework for content-based image retrieval

Content Comparison Techniques

The sections below describe common methods for extracting content from images so that they can be easily compared. The methods outlined are not specific to any particular application domain.

Color

Retrieving images based on color similarity is achieved by computing a color histogram for each image that identifies the proportion of pixels within an image holding specific values (that humans express as colors). Current research is attempting to segment color proportion by region and by spatial relationship among several color regions. Examining images based on the colors they contain is one of the most widely used techniques because it does not depend on image size or orientation. Color searches will usually involve comparing color histograms, though this is not the only technique in practice.

Texture

Texture measures look for visual patterns in images and how they are spatially defined. Textures are represented by texels which are then placed into a number of sets, depending on how many textures are detected in the image. These sets not only define the texture, but also where in the image the texture is located. Texture is a difficult concept to represent. The identification of specific textures in an image is achieved primarily by modeling texture as a two-dimensional gray level variation. The relative brightness of pairs of pixels is computed such that degree of contrast, regularity, coarseness and directionality may be estimated.

Shape

Shape does not refer to the shape of an image but to the shape of a particular region that is being sought out. Shapes will often be determined first applying segmentation or edge detection to an image. In some cases accurate shape detection will require human intervention because methods like segmentation are very difficult to completely automate.

Conclusion

The goal of medical image databases is to provide an effective means for organizing, searching, and indexing large collections of medical images. This requires intelligent systems that have the ability to recognize, capture, and understand the complex content of medical images. Content-based retrieval is a promising approach to achieve these tasks and has developed a number of techniques used in medical images. Despite recent developments, medical content-based image retrieval still has a long way to go and more efforts are expected to be devoted to this area. Ultimately, a well-organized image database, accompanied by an intelligent retrieval mechanism, can support clinical treatment, and provide a basis for better medical research and education.

Reference

[1] El-Naqa, I., Yang, Y., Galatsanos, N. P., Nishikawa, R. M., & Wernick, M. N. A similarity learning approach to content-based image retrieval: Application to digital mammography. IEEE Transactions on Medical Imaging, 23 (10), 1233-1244.

[2] Brodley, C. E., Kak, A, Dy, J., Shyu, C. R., Aisen, A., & Broderick, L. (1999). Content-based retrieval from medical image database: A synergy of human interaction, machine learning, and computer vision. In Proceedings of the Sixteenth National Conference on Artificial Intelligence (pp. 760-767).

[3] Muller, H., Michous, N., Bandon, D., & Geissbuhler, A. (2004a). A review of content-based image retrieval systems in medical applications-Clinical benefits and future directions. International Journal of MedicalInformatics, 73(1), 1-23.