Content-Based Image Retrieval in Picture Archiving and Communications Systems

Hairong Qi and Wesley E. Snyder

We propose the concept of content-based image retrieval (CBIR) and demonstrate its potential use in picture archival and communication system (PACS). We address the importance of image retrieval in PACS and highlight the drawbacks existing in traditional textual-based retrieval. We use a digital mammogram database as our testing data to illustrate the idea of CBIR, where retrieval is carried out based on object shape, size, and brightness histogram. With a user-supplied query image, the system can find images with similar characteristics from the archive, and return them along with the corresponding ancillary data, which may provide a valuable reference for radiologists in a new case study. Furthermore, CBIR can perform like a consultant in emergencies when radiologists are not available. We also show that content-based retrieval is a more natural approach to man-machine communication.
Copyright © 1999 by W.B. Saunders Company

MOTIVATION

WHILE MORE WORK in picture archiving and communications systems (PACS) design has been focused on image transmission, display, and enhancement, we address the importance of image retrieval. Currently, the most popularly used retrieval methods are based on textual information like keywords. Keyword is not a property that relates to the content of the image directly, it is only a language that humans use to characterize or describe the properties of the image. It is hard to find a complete, accurate, and unambiguous set of words that is able to describe all the image properties for all the users, since different users may have different views on how an image should be described.

The limitation of the traditional keyword-based approach has led to the concept of content-based image retrieval (CBIR)[1,2]—retrieve images by their contents, such as texture, color, shape, etc. CBIR represents a more natural approach to man-machine communication. Upon user request, which is usually a query image, the system can find those images that possess similar characteristics and return the corresponding ancillary data. CBIR cannot only assist the radiologists in making a high-quality and more efficient diagnosis of the new case, it can also perform as a consultant in emergency when radiologists are not available.

In addition, CBIR can help locate all of the similar pathologies[3] and store them on-line, greatly reducing the need to fetch them from optical disk on spot, which often takes 4 to 15 minutes, compared with the less than 1 minute on-line locating.[4] CBIR also puts more guarantee in a proper understanding of the images while saving the surgeon's trip to the radiology department for consultation.

APPROACH

A CBIR system consists of two components: index creation and retrieval (Fig 1). We take digital mammography as an example.

When a mammogram is first input into PACS, the index creation component derives the shape information from the suspicious lesions, which is segmented based on the local maxima of the color histogram. The shape information of each lesion is characterized by the length of its two principal components (square root of eigenvalue of the object's scatter matrix); and the histogram shape of data projected on these components. Figure 2 shows the corresponding results by analyzing a testing mammogram. A circular or oval shape will have its projected data histogram match Gaussian very well. By comparing the eigenvalues, circular can be distinguished from oval. As for projection histograms that do not match Gaussian, or have more than one local maximum, an irregular shape or stellate shape is indicated. The feature vector for each image then has three components: length of the first principal component, length of the second principal component, and the degree of Gaussian matching.

When doing retrieval, the user provides a query image, which goes through the index creation

From the Center for Advanced Computing and Communication, North Carolina State University, Raleigh, NC.
Supported in part by Army Research Office through Contract No. DAAH04-93-D-0003D.
Address reprint requests to H. Qi, Center for Advanced Computing and Communication, Box 7914, North Carolina State University, Raleigh, NC 27695-7914.
Copyright © 1999 by W.B. Saunders Company
0897-1889/99/1202-1023$10.00/0

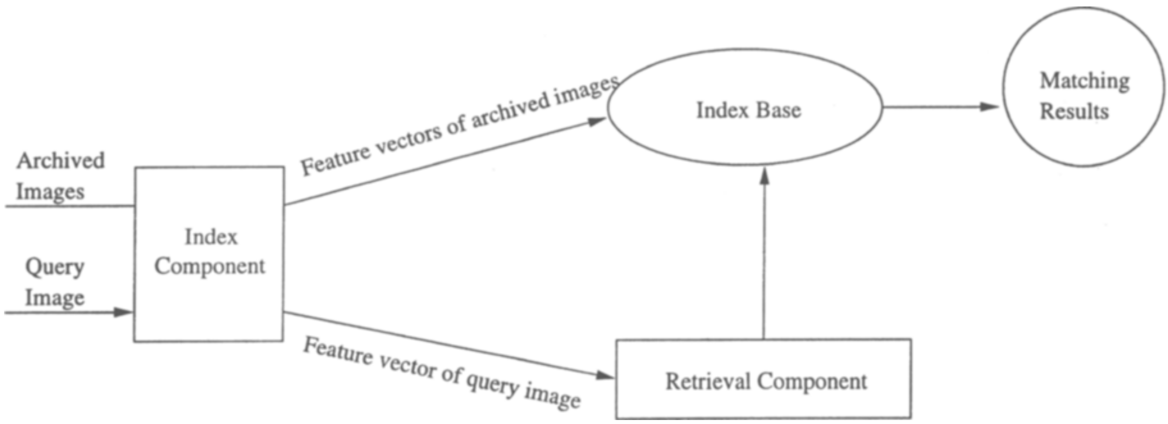

Fig 1. Indexing and retrieving procedure.

component and has its feature vector computed. Then the retrieval component computes the vector distance between the query image and images in the archive. The matching images are those with small distance values. The user can choose how small they want the distance to be, that is, how close the old pathologies are to the query one.

EXPERIMENTAL RESULTS

The testing images are downloaded from the digital mammography database maintained by Mallinckrodt Institute of Radiology of Washington University.[5] Figure 3 shows two query images and the corresponding matching results if using shape as the matching criteria.

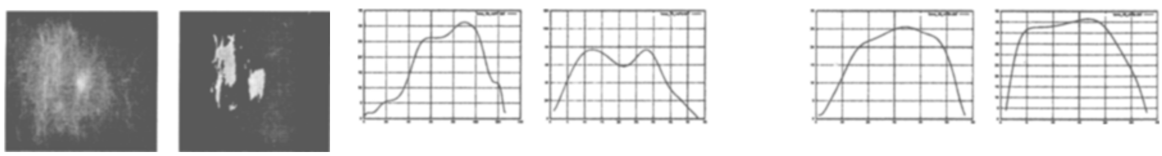

Fig 2. Process of feature vector deviation. Left to right: original mammogram, the segmentation; histograms for data projected on 2 principal components of the left segment; histograms for those of the right segment.

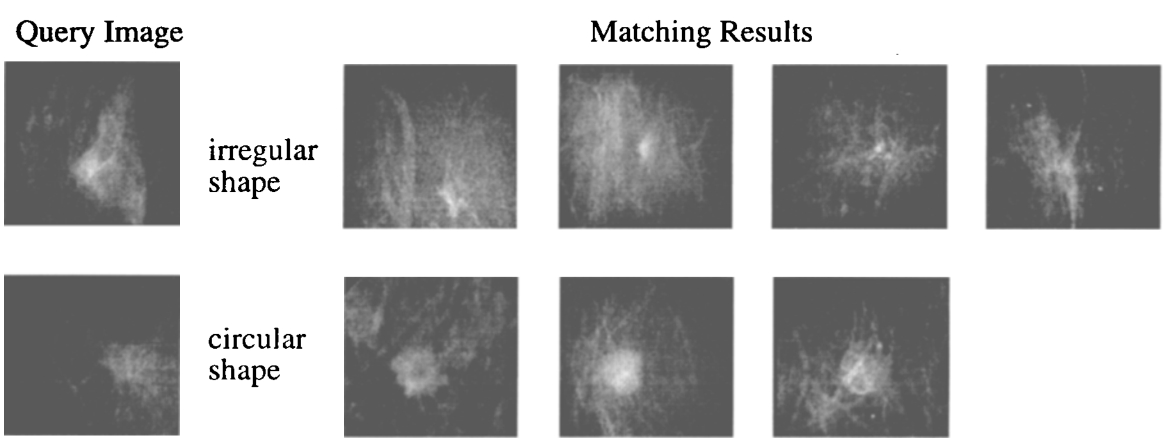

Fig 3. Several matching images retrieved based on the similarity of lesion shape.

ACKNOWLEDGMENT

We thank the researchers in Mallinckrodt Institute of Radiology of Washington University for the work done on the digital mammography database they maintained and shared from the web.

REFERENCES

1. Special Issue on Content Based Image Retrieval. IEEE PAMI, vol 18, no 8, August, 1996

2. Special Issue on Content Based Image Retrieval. IEEE Computer, vol 18, no 9, September 1995

3. Leotta DF, Kim Y: Requirements for picture archiving and communications. IEEE Eng Med Biol vol 12, no 1, 62-69, March, 1993

4. Pomerantz SM, Siegel EL, Pickar E, et al: PACS in the operating room: Experience at the Baltimore VA Medical Center. Proceedings of the Fourth International Conference on Image Management and Communications, Oahu Island, HI, August 20-24, 1995, pp 238-242

5. Mallinckrodt Institute of Radiology of Washington University: Digital mammography database. http://www.erl.wustl.edu/mammo/digital2.html, 1997