Chapter XIV
Content-Based Retrieval for Mammograms

Chia-Hung Wei
Ching Yun University, Taiwan

Chang-Tsun Li
University of Warwick, UK

Yue Li
University of Warwick, UK

ABSTRACT

As distributed mammogram databases at hospitals and breast screening centers are connected together through PACS, a mammogram retrieval system is needed to help medical professionals locate the mammograms they want to aid in medical diagnosis. This chapter presents a complete content-based mammogram retrieval system, seeking images that are pathologically similar to a given example. In the mammogram retrieval system, the pathological characteristics that have been defined in Breast Imaging Reporting and Data System (BI-RADS™) are used as criteria to measure the similarity of the mammograms. A detailed description of those mammographic features is provided in this chapter. Since the user's subjective perception should be taken into account in the image retrieval task, a relevance feedback function is also developed to learn individual users' knowledge to improve the system performance.

1. INTRODUCTION

In hospitals and medical institutes, a large number of medical images are being produced in ever increasing quantities and used for diagnostics and therapy. The need for effective methods to manage and retrieve those image resources has been actively pursued in the medical community. The design of Picture Archiving and Communication Systems (PACS) is to integrate imaging modalities and inter-

faces with hospital and departmental information systems in order to manage the storage and distribution of images to radiologists, physicians, specialists, clinics, and imaging centers (Huang, 2003). A crucial requirement in PACS is to provide an efficient search function to access desired images. Image search in the digital imaging and communication in medicine (DICOM) protocol of PACS is currently carried out according to the alphanumerical order of the textual attributes of images (Lehmann et al., 2003). However, the information which users are interested in is the visual content of medical images rather than that residing in alphanumerical format. Traditional search requires images to be annotated with text, allowing the images to be accessed by text-based retrieval. As the size of the medical image database grows, it becomes impractical to manually annotate all contents and attributes of the images. The content of images is an informative and direct query which can be used to search for other images containing similar content. As content-based access approaches are expected to have a great impact on PACS and health database management, content-based image retrieval has been proposed for inclusion in PACS (Lehmann et al., 2003). In a PACS environment, a content-based image search system can support the usual comparisons made on images by physicians, answering similarity queries over the images stored in the distributed databases (Muller et al., 2004). The importance of searching for similar images comes from the fact that physicians usually try to recall similar cases by seeking images that are pathologically similar to a given image (Traina et al., 2005). As medical images are digitally represented in a multitude of formats based on their modality and the scanning device used, image retrieval systems have to be developed for their specific image types. 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. Those systems include ASSERT for High-Resolution Computed Tomography (HRCT) of lungs (Shyu et al., 1999), CasImage for a variety of images from CT, MRI, and radiographs, to color photos (Muller et al., 2004), IRMA for various imaging modalities (Lehmann et al., 2004), and NHANES II for cervical and lumbar spine X-ray images (Antani et al., 2004).

Breast cancer is the most common cancer among women and affects approximately one million women worldwide. In the UK, for example, breast cancer accounts for 30 per cent of all female cancers and approximately 1 in 9 women in the UK will suffer from breast cancer sometime during their life (Beaver & Witham, 2007). Mammography has been a reliable method for the detection of breast cancer (Highnam & Brady, 1999) and women are usually asked to take mammograms regularly, and as a result many digital mammograms are produced in hospitals and breast screening centers. As distributed mammogram databases at hospitals and breast screening centers connect together through PACS, a mammogram retrieval system is needed to help medical professionals locate the mammograms they want in aid of medical diagnosis and case-based reasoning (Wei et al., 2006), thereby reducing false positives and false negatives in medical screening.

The main purpose of this chapter is to disseminate the knowledge of the content-based retrieval approach to mammogram database indexing and to attract greater interest from various research communities to rapidly advance research in medical image retrieval. The rest of the chapter is organized as follows: The second section reviews the background information and addresses the challenges of content-based mammogram retrieval. The third section presents the proposed framework for content-based mammogram retrieval. The fourth section provides more details on feature extraction methods for mass and calcification presentation. The fifth section discusses potential research issues in the future research agenda. The last section concludes this chapter.