COMPUTER-AIDED DIAGNOSIS OF BREAST CANCER VIA AN INDEXED ATLAS AND CONTENT-BASED RETRIEVAL OF MAMMOGRAMS

H. Bosnak, R. Grewal, J. Howland, B. Liu, and R. M. Rangayyan

Department of Electrical and Computer Engineering, University of Calgary
Calgary, Alberta, CANADA T2N 1N4

ranga@ucalgary.ca

Abstract: This paper describes the development of an indexed atlas of digitized mammograms to facilitate content-based retrieval and comparative analysis of mammograms. The atlas includes all available mammograms of each patient, as well as the associated clinical information. A graphical user interface is being developed to facilitate efficient access to the atlas and content-based retrieval of images and case files. At the current stage of development, the atlas is focused on qualitative and objective representation of masses in mammograms. The procedures for the retrieval and analysis of masses include the use of mammographic image content, in the form of content-based numerical feature values and text-based descriptors of radiological features. The procedures for content-based representation, retrieval, and analysis of masses are described, along with preliminary results.

Introduction

Breast cancer is the most commonly occurring cancer in Canadian women [1]. It is in the early stages of the disease when treatment methods are most effective. "Screen Test: Alberta Program for the Early Detection of Breast Cancer" provides breast cancer screening programs through mammography for women through screening centers in Calgary and Edmonton, and to women in over 100 rural communities via mobile units [2]. Mammographic images are difficult images to interpret; the large numbers of screening mammograms being acquired increase the complexity of their management and interpretation. It is desired that computer-aided diagnosis (CAD) technology [3] be introduced to provide advanced tools for radiologists. CAD technology can assist in the analysis of mammograms using image processing techniques. The results of image processing can be used in the form of quantitative or objective indices to organize and search databases of images, providing methods for comparative analysis of mammograms and teaching tools. Our research team at the University of Calgary is in the process of developing a CAD system including procedures for content-based retrieval and comparative analysis of mammograms [4]. An indexed atlas of digital mammograms [5, 6], being developed as a part of this project, provides an effective and efficient tool for comparative analysis of mammograms, as well as a teaching tool for radiologists. The design of the indexed atlas and its components is described in the following sections, along with illustrations of the associated graphical user interface (GUI) and preliminary results.

Computer-aided Diagnosis of Breast Cancer

Objective methods for the analysis of mammographic features are needed to assist radiologists in the evaluation of ambiguous features [3 – 5]. Digital mammographic imaging and image analysis systems that can detect diagnostic features on a mammogram and give visual prompts to a radiologist are now available [3, 7, 8]. An indexed database or atlas of mammograms could be a useful resource for radiologists, both for diagnosing difficult cases and for teaching purposes [4 – 6]. The atlas could contain digital mammographic images, associated patient history, physical findings, and the results of clinical investigation. Such information could facilitate comparative analysis of mammographic features of an ambiguous case with those of similar biopsy-proven cases.

Analysis of Masses: At the initial stages, the design of the indexed atlas has been focused on the analysis of masses in mammograms. The procedure for indexing the masses in the atlas includes the use of image content, in the form of numerical feature values, image attributes, and other image descriptors, to assist in the identification of relevant images or image-related information in response to a query. Image content must be represented in the atlas in a radiologically and diagnostically meaningful way in order for content-

ISSN: 1727-1983

based image retrieval system to be possible. The indices used to access the atlas, designed so as to facilitate content-based retrieval and comparative analysis of mammograms, include features related to the most significant indicators of breast cancer, such as shape factors, edge sharpness, and texture of masses [4].

Sahiner et al. [9] found that the combined use of sets of morphological and texture features can provide better classification accuracy than either set of features on its own. Although it has been found that the derivation of accurate contours of masses is difficult, it has also been observed that several shape factors can provide significantly higher individual classification accuracies than texture features [4, 9]. Thus, it can be said that shape plays an important role in distinguishing between malignant tumors and benign masses. Texture features, however, cannot be ignored. Therefore, efforts have been focused on quantifying textural content in the mass region of interest (ROI) and the mass margins [4, 9].

Currently Available Commercial CAD Systems: Currently, there are two established CAD systems that have been approved for clinical use by the Food and Drug Administration (FDA) in the USA, that provide an interpretive aid and may be used as a second reader. One such system is R2 Technology's ImageChecker system [7]. This system is a CAD tool that uses pattern recognition software to interpret digital mammographic images, and brings to the attention of the radiologist possible ROIs of diagnostic importance (such as calcifications and masses). R2 Technology claims the ImageChecker system provides earlier detection of up to 23.4% of cancers currently detected by screening mammography, and that it has the lowest false-marker rate among comparable systems [7]. It was also found that an 8% to 19% increase in cancer detection was actually observable [7]. Of 55 cases analyzed, a false-marker rate of 1.9 per image was observed.

Another CAD system on the market is iCAD's Second Look series, which includes systems designed to aid in the early detection of breast cancer. iCAD claims that its products can detect 25% of breast cancer cases an average of 14 months earlier than screening mammography alone [8].

The CAD systems mentioned above do not appear to facilitate comparative analysis of mammograms or content-based retrieval. This limitation has prompted us to develop an indexed atlas of mammograms and related patient information to facilitate content-based retrieval and analysis of mammograms [4-6].

Performance Analysis of CAD Systems: CAD techniques and systems have been shown to increase the sensitivity of detection of breast cancer. However, while CAD techniques have been effective in detecting masses and calcifications, they have been found to perform poorly in the detection of architectural distortion, which represents distortion of the architecture of the breast without the presence of a definite mass [10]. Because subtle mammographic abnormalities, such as architectural distortion, are the most frequent source of screening errors and false-negative findings in cases of interval cancer, new systems for CAD should target the detection of such abnormalities [10].

The performance of a commercial CAD system in the detection of masses and calcifications in screening mammography was investigated by Burhenne et al. [11], and a sensitivity of 75% was obtained. The performance of a commercial CAD system to mark invasive lobular carcinoma of the breast was investigated by Evans et al. [12], and the system was found to identify correctly 17 of 20 cases of architectural distortion. Birdwell et al. [13] investigated the performance of a CAD system to mark cancers overlooked by radiologists; the system detected five of the six cases of architectural distortion presented. Baker et al. [14] found that two commercial CAD systems could detect fewer than 50% of the cases of architectural distortion presented. It has also been observed that CAD systems can produce a high number of false-positive marks. Thus, while CAD technology has the potential to aid greatly in the early detection of breast cancer and other breast abnormalities, these findings emphasize the need for further research in the area of image processing algorithms to detect and analyze subtle signs such as architectural distortion, as well as reduce false-positive findings.

Databases of Mammograms

Several databases for research in mammographic image analysis have been developed; see Alto et al. [5] for a detailed discussion on databases of mammograms. A database should not only facilitate the retrieval of selected image files, but should also promote organized storage and access of associated data. The most widely used databases in mammography research circles include the Mammographic Image Analysis Society (MIAS) database [15] and the University of South Florida Digital Database for Screening Mammography (DDSM) [16].

The MIAS Database: MIAS is an organization of research groups in the UK, interested in working on the analysis and understanding of mammograms. The MIAS database contains 161 cases obtained from screening programs in the UK, and has an image–labeling scheme that may be used to find specific types of cases. The database provides the location and approximate size of each mass. The database contains an unusually large number of cases with spiculated benign masses. A major limitation of the MIAS database is that it includes only one mammogram per breast.

The DDSM Database: The objective of the DDSM database [16] is to facilitate research in the development of computer algorithms to aid in screening for breast cancer. The DDSM database contains approximately 2,500 studies. Each study includes two images of each breast, along with some associated patient information and image information. Images containing suspicious areas also have associated pixel-level information about the locations and types of the suspicious regions. Regions containing masses are identified approximately with a contour. DDSM has a search capability designed to allow the user to identify cases that meet specified criteria such as normal, cancer, or benign disease; ACR (American College of Radiology) breast density rating; and ACR abnormality keyword description [17].

Major limitations of the MIAS and DDSM databases are that they include only one set of mammograms per subject, and that they do not provide for content-based retrieval and comparative analysis of mammograms. Furthermore, the clinical information provided is limited. Similar limitations apply to the commercial CAD systems mentioned above.

Indexed Atlas and Content-based Retrieval

Currently available CAD systems and mammographic databases do not retain any information obtained from the application of image processing algorithms to the cases. We are attempting to address this limitation by developing an indexed atlas of mammograms and related patient information along with a content-based retrieval system for CAD of breast cancer. The system has been designed to provide the user with an effective method for retrieving mammograhic images and related patient files for comparative analysis. The database is organized by image content using an indexing system for storage and retrieval of the specific images and clinical information in the database. Specifically, breast masses are indexed and analyzed using shape factors, texture measures, and edge-sharpness measures [4–6]. The representation of breast masses and tumors includes the design of 21 content-based descriptors to represent the image features with minimal loss of information [4, 5]. These features are related to radiologically established attributes of breast masses. Measures related to shape, texture, and the sharpness of mass margins have been shown to be effective in differentiating between benign masses and malignant tumors in pattern classification studies, and therefore, are being used in the indexed atlas [4, 5].

High-level Design of the Indexed Atlas: Figure 1 shows the high-level design of the indexed atlas. As can be seen in the figure, there are two main components of the indexed atlas, both of which interface with outside components. The database component contains the mammographic images and their related patient files, and is stored on a terabyte server. The database on the terabyte server must facilitate effective and efficient access by an end-user in order to be a useful CAD entity. Such access is achieved by the implementation of a GUI that provides two search methods to the user – text-based and content-based – and implements search algorithms for the retrieval of results that match the user's query.

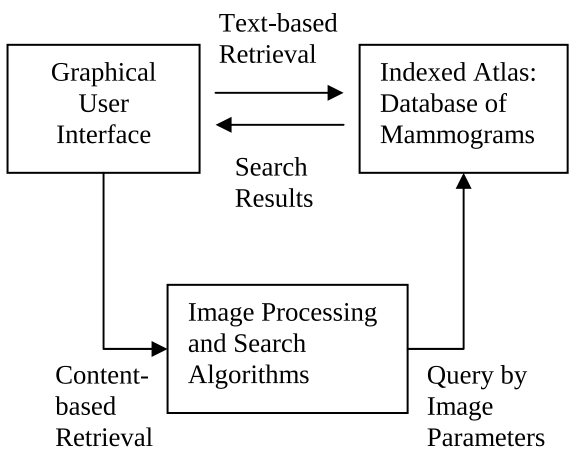

Figure1: High-level design of the indexed atlas of digitized mammograms.

Detailed Design of the Indexed Atlas:

Database: Cases used in the building of the atlas were selected from Screen Test: Alberta Program for the Early Detection of Breast Cancer [2]. The images and cases were not selected by the ease or difficulty of diagnosis, and therefore, they are not necessarily examples of classical diagnoses typically found in most hard-copy atlases [5]. All available images for each woman are included, along with the corresponding patient information, physical findings, and related biopsy or pathology reports in the case of an abnormality. Some cases consist of up to six exams spanning a period of up to 10 years with as many as 30 images in total, thus providing a good overview of the progression of breast tissue changes due to the normal aging process or the development of an abnormality. The time element offers potential for the analysis of mammograms taken prior to the detection of cancer (interval cancer or screen-detected cancer), and provides the opportunity to test CAD methods for their

ISSN: 1727-1983 © 2005 IFMBE

ability to detect signs of cancer earlier than with mammography on its own.

The database is encoded in the form of two spreadsheets containing all of the 181 cases in the atlas at present. The first spreadsheet includes the case number, diagnosis, type of abnormality found (if any), and the dates of the examinations. The second spreadsheet contains the case number and the numerical values of 21 content-based indices computed via the application of image processing algorithms. The features include shape factors known as compactness (*C*), fractional concavity (F_{cc}), spiculation index (*SI*), and Fourier-descriptor-based factor (*FF*); 14 measures of texture labeled as F1 to F14; and three measures of edge sharpness known as acutance, coefficient of variation, and contrast [4, 5]. Thus, the images with masses in the database are indexed in terms of quantitative measures of image content. The mammographic images stored in the database are also linked with text files containing related clinical and other information. This information includes patient information (made anonymous), diagnostic information, and the results of image processing. In this manner, the indexed atlas facilitates the retrieval of images or cases based on image content, and will allow for comparative analysis of mammograms.

GUI: Access to the indexed database is provided through a GUI. The GUI includes adequate detail to allow for full searching capability, and is easy to use. In order to utilize the full potential of the indexed atlas, search procedures with both content-based indices and text-based indices have been implemented using two tabs on the search screen, as shown in Figure 2: one for content-based search and one for text-based search.

When a user selects text-based search, he or she can enter a text descriptor of the item(s) desired. The GUI then calls the search software to retrieve the appropriate information.

Figure 3 shows the process for a content-based search. When a content-based search is initiated, a pop-up window containing check boxes for all of the 21 content-based indices for masses appears (see Figure 3a). The user may select one, some, or all of the content-based indices to perform a search. The next step pops up a window containing scroll bars for the selected content-based indices (see Figure 3b). The user can then specify appropriate values for the indices and then click the "Perform Search" button.

In both text-based and content-based searches, the GUI calls the search software based on the user input, and returns a list of the matching cases in a pop-up table, as shown in Figure 4. In the case of content-based retrieval, the Euclidean distance is computed between a vector representing the selected features that specify the query and a similar vector composed for each case to be evaluated, and the images are sorted in increasing order of the distance. A specified number of cases at the upper end of the list are presented in the output. Figure 5 shows how the user is able to view the mammographic images and related patient information of any of the retrieved cases by double clicking on the case identifier in the table.

Figure 2: GUI screen in the indexed atlas for text-based and content-based searching.

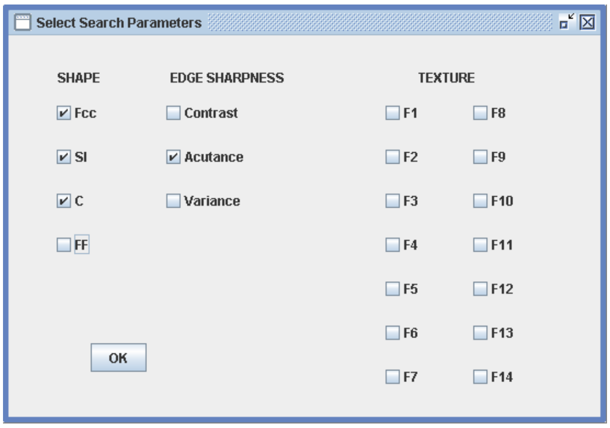

Figure 3a: First GUI screen for content-based search.

ISSN: 1727-1983

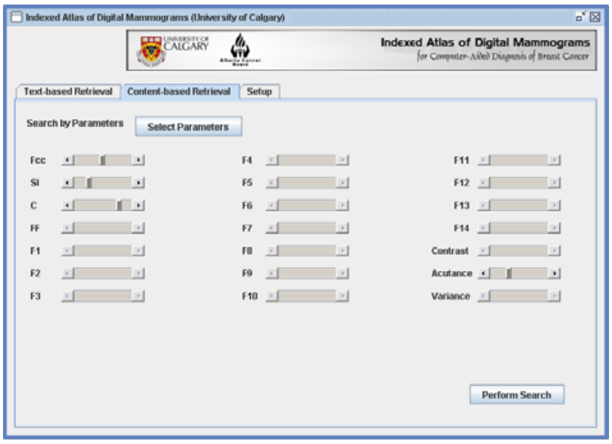

Figure 3b: Second GUI screen for content-based search.

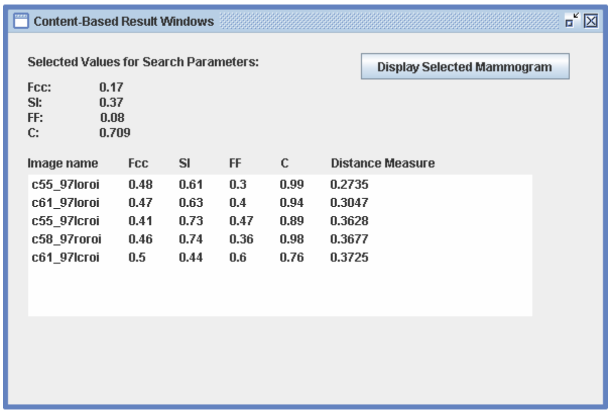

Figure 4: Pop-up table with the results of a content-based search.

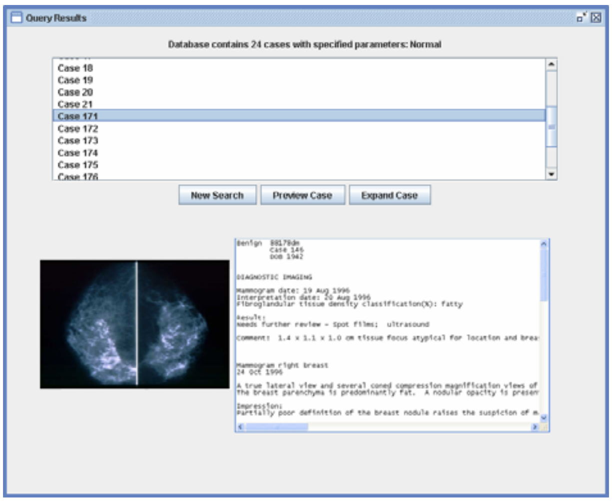

Figure 5: Viewing of a selected case including images and clinical information.

Future Work

Future work will be directed toward the realization of the full potential of the indexed atlas by expanding the capabilities of the associated CAD system. The full potential of the atlas will be realized when all of the image processing algorithms used to compute the content-based indices for masses (and other measures to characterize various diagnostic features such as calcifications, bilateral asymmetry, and architectural distortion) are compiled into one concurrent program, and comparative analysis of mammograms is fully implemented. At that stage, the user will be able to input a mammogram; the programs will detect the diagnostic features present in the image, compute content-based indices of the features, and retrieve cases from the database that match the input mammogram.

Conclusion

The potential of CAD of breast cancer as well as the limitations of the currently available CAD systems and databases were the motivation behind the development of an indexed atlas of digital mammograms and their related patient files. The GUI developed as a part of this project will allow radiologists and researchers to access the atlas in an effective and efficient manner for diagnostic and teaching purposes. To date, the database of mammograms and the associated clinical information have been prepared, and the GUI implemented; these components together result in an indexed atlas of digital mammograms and the related patient files. The atlas will serve as a tool for content-based retrieval and comparative analysis of digital mammograms, and should facilitate early detection of breast cancer.

Acknowledgment

This work was supported by grants from the Alberta Heritage Foundation for Medical Research (AHFMR) and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

[1] CANADIAN CANCER SOCIETY (2005): http:// www. cancer.ca. Accessed on March 3, 2005.
[2] ALBERTA CANCER BOARD (2004): "Screen Test: Alberta Program for the Early Detection of Breast Cancer," 2001/03 Biennial Report. (Edmonton, Alberta, Canada.)
[3] YAFFE, M. J. (2001): Editor – 'Digital Mammography', Proceedings of the 5th International Workshop on Digital Mammography, Toronto, ON, Canada. (Medical Physics Publishing, Madison, WI.)

 ISSN: 1727-1983

[4] ALTO, H., RANGAYYAN, R. M., AND DESAUTELS, J. E. L. (2005): 'Content-based retrieval and analysis of mammographic masses', *J. Electron. Imag.*, **14** (2), 023016, pp. 1-17.

[5] ALTO, H., RANGAYYAN, R. M., PARANJAPE, R. B., DESAUTELS, J. E. L., AND BRYANT, H. (2003): 'An indexed atlas of digital mammograms for computer-aided diagnosis of breast cancer', *Annales des Télécommunications*, **58** (5-6), pp. 820-835.

[6] LAU, H., MOK, K., SO, D., TSE, C., AND RANGAYYAN, R. M. (2004): 'An indexed atlas for content-based retrieval and analysis of mammograms', Proc. of 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, September 2004, San Francisco, CA, pp. 3167-3170.

[7] R2 TECHNOLOGY (2005): Sunnyvale, CA. http:// www. r2tech.com. Accessed on February 3, 2005.

[8] *i*CAD (2005): Beavercreek, OH. http:// www. icadmed.com. Accessed on February 3, 2005.

[9] SAHINER, B. S., CHAN, H.-P., PETRICK, N., HELVIE, M. A., AND HADJIISKI, L. M. (2001): 'Improvement of mammographic mass characterization using spiculation measures and morphological features', *Med. Phys.*, **28** (7), pp. 1455-1465.

[10] AYRES, F. J. AND RANGAYYAN, R. M. (2005): 'Characterization of architectural distortion in mammograms using orientation fields', *IEEE Engineering in Medicine and Biology*, January, **24** (1), pp. 59 – 67.

[11] BURHENNE, L. J. W., WOOD, S. A., D'ORSI, C. J., FEIG, S. A., KOPANS, D. B., O'SHAUGHNESSY, L. F., SICKLES, E. A., TABAR, L., VYBORNY, C. J., AND CASTELLINO, R. A. (2000): 'Potential contribution of computer-aided detection to the sensitivity of screening mammography', *Radiol.*, **215**, pp. 554-562.

[12] EVANS, W. P., BURHENNE, L. J. W., LAURIE, L., O'SHAUGHNESSY, K. F., AND CASTELLINO, R. A. (2002): 'Invasive lobular carcinoma of the breast: Mammographic characteristics and computer-aided detection', *Radiol.*, **225** (1), pp. 182-189.

[13] BIRDWELL, R. L., IKEDA, D. M., O'SHAUGHNESSY, K. F., AND SICKLES, E. A. (2001): 'Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection', *Radiol.*, **219** (1), pp. 192-202.

[14] BAKER, J. A., ROSEN, E. L., LO, J. Y., GIMENEZ, E. I., WALSH, R., AND SOO, M. S. (2003): 'Computer-aided detection (CAD) in screening mammography: Sensitivity of commercial CAD systems for detecting architectural distortion', *Am. J. Roentgenol.*, **181**, pp. 1083-1088.

[15] THE MAMMOGRAPHIC IMAGE ANALYSIS SOCIETY (2005): http://www.wiau.man.ac.uk/services/ MIAS/MIASweb.html. Accessed on February 1, 2005.

[16] UNIVERSITY OF SOUTH FLORIDA DIGITAL MAMMOGRAPHY HOME PAGE (2005): http:// marathon.csee.usf.edu/Mammography/Database.html. Accessed on February 1, 2005.

[17] AMERICAN COLLEGE OF RADIOLOGY – ACR (1998): 'Illustrated breast imaging reporting and data system (BI-RADS)'. 3rd edition. (American College of Radiology, Reston, VA.)