The Digital Database for Breast Ultrasound Image

Jia-wei Tian[a, b] Ying Wang[b] Jian-hua Huang[a] Chun-ping Ning[b] Han-mei Wang[b] Yan Liu[a] Xiang-long Tang[a]

[a] School of Computer Science and Technology , Harbin Institute of Technology, Harbin, 150001, China
[b] Second Hospital affiliated to the Harbin Medical University, Harbin, 150086, China

Abstract[1]

The Digital Database for Breast Ultrasound Image (DDBUI) is a database of digitized screen sonography with associated ground truth and some other information. The DDBUI project is a collaborative effort involving the Harbin Institute of Technology and the Second Affiliated Hospital of Harbin Medical University. The purpose of this resource is to provide a set of sonography images that may be used by researchers to evaluate and compare the performance of computer-aided diagnosis (CAD) algorithms. It contained 285 cases, each of which has 3~8 images and some corresponding information (patient age, sex, family history), and image information included shape, boundary, envelope, echo pattern, microcalcification et al. Totally 1132 images have been collected in the database and the number is growing. This paper provided an overview of the completed database, including the background of the project, the contents of the database, the organization of the cases and the access to the database.

Keywords: digital database; ultrasound; breast; computer-aided diagnosis

1. Background

Breast cancer is still one of the most common cancers and is the leading cause of death among women in the world [1]. More and more researchers emphasized the early detection and diagnosis of breast cancer [2]. With recent development in ultrasound equipment, sonography is now a well-established tool in breast imaging, allowing identification of up to 27% of breast masses that are occult on mammography, especially in women younger than 50[3,4]. However, breast ultrasound images have some inherent limitations, such as: vague tumor boundaries, high amount of speckles, etc. Several studies indicated that CAD system can help radiologists in reading and interpreting sonography, and many CAD algorithms were studied during the past few years [5-7]. However, for an algorithm, different testing images may present different results, hence, it was very difficult to compare two different algorithms. There was one Digital Database for Screening Mammography (DDSM) [8] completed in the fall of 1999, which was used as a resource for analysis researches of mammographic image, but to our knowledge, no such a database has been used for the ultrasound researches.

We built this database for the further use of testing new CAD algorithms and systems. Provided the same data, the same performance measure and the same train and test method, the comparison of different algorithms can be more objective.

2. The contents of the DDBUI

The DDBUI contained information and images of breast with or without lesions. By far, the database contained 285 cases and 1132 images in total. All the images were acquired from the Second Affiliated Hospital of Harbin Medical University from 2002 to 2007. Similar to the DDSM, the DDBUI was organized into "classes" and "cases ".

2.1. Classes

There are three kinds of classes: normal, benign, and malignant. Each class is a collection of cases of the corresponding type.

Normal class is formed by cases from the patients who came to the hospital for routine examination and were confirmed as healthy for at least four years following up. There are cases different from ages and periods.

Benign class contains cases from screening exams in which something suspicious was found, but was determined to be benign by pathology (with surgical excision or US-guided percutaneous core-needle biopsy). Kinds of benign lesions were included in the database, such as fibroadenomas, introductal papillomas, lipomas, cysts, and granulomas and so on.

Malignant class contains cases which were confirmed to be malignant by pathology, also with surgical excision or US-guided percutaneous core-needle biopsy. Most of them are infiltrative ductal carcinomas and invasive lobular carcinomas; also there are simplex carcinomas, inflammatory carcinomas and medullar carcinomas.

2.2. Cases

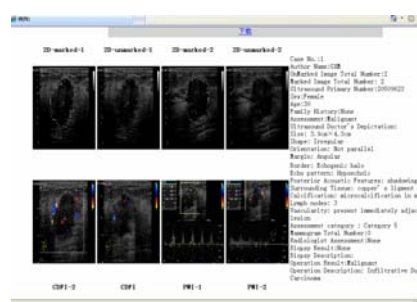In the DDBUI database, each case is stored in a separate directory.

Fig.1. Case 1 in the DDBUI

Figure 1 shows an example of the files in the directory for case 1. As you can see, there are two types of files: ".txt" and ".jpg".

The ".txt" files contain the whole information of the case and were written in ASCII format. It lists important information of the patient and the lesions, such as the number of the case, the name, age and sex of the patient, the patient's family history of breast cancer, the description and the pathology results of the lesions and some description of the axillary lymph nodes. Thereinto, the description of the lesions includes the shape, margin, size, orientation, echo pattern, posterior acoustic features, surrounding tissues, calcification, lood flow and so on. Remarkably, there's no information of lesions and lymph nodes in the "normal case".

All the images in the database were stored in a format using lossless JPEG. There are six kinds of images of the lesions, including two-dimensional (2D) Breast Ultrasound (BUS) images, three-dimensional (3D) BUS images, Power Doppler Images (PWI), Color Doppler Images (CWI), Mammographies and Pathologic images of the lesions. Most of

the images size is 450×450 pixels. To each 2D BUS image of the mass, there is one file called "marked" image and another named "unmarked", the difference between them is that the "marked" image has a marker along the margin of the mass marked manually by a doctor who has more than 10 years experience of BUS examination.

3. The Organization of the DDBUI

3.1. Case Gathering

All the cases were selected from the clinical patients who came to the Second Affiliated Hospital of Harbin Medical University for BUS examination since 2002. The mean age was 43.7 years, range 0.5~81years. All the patients were examined by a high-resolution Vivid7 sonography system (GE Healthcare, Milwaukee, WI) equipped with 7.5~14MHz liner transducer. Conventional ultrasonography was performed with the patients in the supine position with double breasts exposed fully, or lateral decubitus position if the lesions were in the lateral quadrant of the breast. Both breasts were scanned directly and compared each other. Images of each quadrant of the normal breast were get and stored in the hard disk. The lesions were scanned and stored along different directions and angles, and the occipitofrontal, transverse and up-down diameters were measured. At least 3 images were get each case, varied from three to ten, the size of the lesions were 0.8~6.5cm. The images were quantified into 256 gray levels. The average size was 450×450 pixels. The examinations were performed by doctors who have more than 6 year's experience of BUS images. The nature of the lesions was all confirmed by pathology (either with surgical excision or with US-guided percutaneous core-needle biopsy).

3.2. Processing

Before inputting into the database, some privacy information of the patient was removed from the record. And the 2-D sonograms with at least one mass were marked manually by a doctor who has more than 10 years experiences of BUS examination along the mass margin to provide a result of segmentation for the further CAD process.

3.3. Organization of the images

Images in the database including the sonograms, mammography and pathology images. We collected these images from the continuous following up of the patient. Thereinto, the sonogram includes 2D BUS images, 3D BUS images, PWI and CWI. And the marked and unmarked 2D BUS images were put into the database separately. They were organized in the following module (shown fig.2.)

Fig.2. Module of the images and Retrieval

3.4. Organization of the information

The information of the patient was get from interrogation and was put into the database in the module shown in fig3.
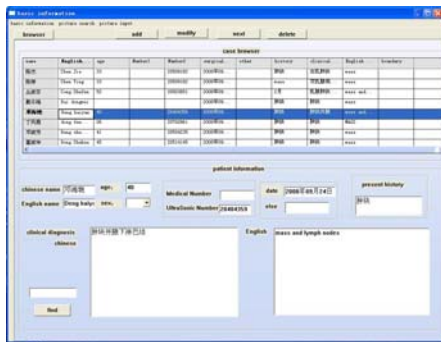
Fig.3. Basic Information Module

Five doctors who have more than 10 years of BUS examination were invited to collect the information of the image. Based on the static images and active films of the mass, the doctors described the mass according to the Breast Imaging Reporting and Data System for Ultrasound (BI-RADS-US) [9]. The description included 2-D ultrasound image (boundary, margin, size, orientation, texture, calcification, capsule, position, surrounding tissues, lymph nodes et al), CDFI (vascularity, the number, position and trend of the vessels) and PWI (velocity and resistance index of the arteries and veins). The information was stored into the database by the module for Ultrasound information. (Shown in Fig 4.)

Fig.4. The module for Ultrasound Information

Only the administrator has the right to add, delete or modify the information in the modules.

4. Retrieval

There is a retrieval module for users to search special cases in the database quickly. Several retrieval words were provided in the database, for example: number of the case, number of the pathology, calcification, time of the examination, pathology results, vascularity et al. The number of the cases associated with the retrieval words were given out quickly. The readers also can do some complex search by using "and" "or" "not". The retrieval module was also shown in fig.2.

5. Access to the DDBUI resource

The web page for the DDBUI resource is http://pr-ai.hit.edu.cn/snacl/index.html. The visitor need to register to get an ID before allowed into the database, and please credit the DDBUI project as the source of the data, also acknowledge to the Harbin Institute of Technology and the Second Affiliated Hospital of Harbin Medical University at the end of your paper if you use data from DDBUI in publication.

6. Reference

[1] A. Jemal, T. Murray, E. Ward, A. Samuels, R. C. Tiwari, A. Ghafoor, E. J. Feuer, and M. J. Thun, "Cancer Statistics", *CA Cancer J Clin*, vol.55, no.1, pp. 10-30, 2005.

[2] E. Paci, "Mammography and beyond: developing technologies for the early detection of breast cancer", *Breast Cancer Res*, vol.4, no.3, pp. 123 - 125, 2002.

[3] Hong AS, Rosen EL, Soo MS, Baker JA."BI-RADS for sonography. Positive and negative predictive values of sonographic features". *AJR Am J Roentgenol*, vol.184, pp.1260–1265, 2005.

[4] Hille H, Vetter M, Hackeloer BJ. "Re-evaluating the role of breast ultrasound in current diagnostics of malignant breast lesions", *Ultraschall Med* ,vol. 25, pp: 411–417,2004.

[5] Oliver A, Lladó X, Freixenet J, Martí J."False positive reduction in mammographic mass detection using local binary patterns",*Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, vol.10, pp:286-93,2007.

[6] Ball JE, Bruce LM. "Digital mammographic computer aided diagnosis (CAD) using adaptive level set segmentation",*Conf Proc IEEE Eng Med Biol Soc*, pp:4973-4978, 2007.

[7] Regentova E, Zhang L, Zheng J, Veni G. "Microcalcification detection based on wavelet domain hidden markov tree model: study for inclusion to computer aided diagnostic prompting system", *Med Phys*, vol.34, no. 6, pp:2206-19,2007.

[8] http://marathon.csee.usf.edu/mammography/Database.html

[9] American College of Radiology. BI-RADS: ultrasound, 1st ed. In: Breast imaging reporting and data system: BI-RADS atlas, 4th ed. Reston, VA: American College of Radiology, 2003