Computer Vision and Image Understanding 114 (2010) 33–43

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Automated assessment of breast tissue density in digital mammograms

T.S. Subashini *, V. Ramalingam, S. Palanivel

Department of Computer Science and Engineering, Annamalai University, Annamalainagar 608 002, India

ARTICLE INFO

Article history:
Received 9 September 2008
Accepted 11 September 2009
Available online 24 September 2009

Keywords:
Mammograms
Breast tissue density
Segmentation
Pectoral muscles
Artifact removal
Statistical features
Support vector machines

ABSTRACT

Mammographic density is known to be an important indicator of breast cancer risk. Classification of mammographic density based on statistical features has been investigated previously. However, in those approaches the entire breast including the pectoral muscle has been processed to extract features. In this approach the region of interest is restricted to the breast tissue alone eliminating the artifacts, background and the pectoral muscle. The mammogram images used in this study are from the Mini-MIAS digital database. Here, we describe the development of an automatic breast tissue classification methodology, which can be summarized in a number of distinct steps: (1) preprocessing, (2) feature extraction, and (3) classification. Gray level thresholding and connected component labeling is used to eliminate the artifacts and pectoral muscles from the region of interest. Statistical features are extracted from this region which signify the important texture features of breast tissue. These features are fed to the support vector machine (SVM) classifier to classify it into any of the three classes namely fatty, glandular and dense tissue.The classifier accuracy obtained is 95.44%.

1. Introduction

Breast density, a measure of the extent of radio dense fibro glandular tissue in the breast, has the potential to be used as a predictor of breast cancer risk, it is a measure of how well tissue can be seen on mammogram [1]. Some tissue, such as the milk gland, is dense and appears white on a mammogram. This density makes it hard for doctors to see tumors, which also appear white. Fatty tissue is less dense and appears clear on the mammogram, allowing better tumor detection. In 1976, Wolfe published an article that demonstrated a relationship between breast density and breast cancer risk. He showed women with dense breasts have been shown to have a four- to six-fold increased risk of developing breast cancer [2]. Dense tissue in more than 50% of the breast could account for approximately one-third of breast cancers [3]. Mammographic density has also been associated with breast cancer tumor characteristics, including tumor size, lymph node status, and lymphatic or vascular invasion in screen-detected cancers [4]. A threefold increased risk of second breast cancers has also been observed in women diagnosed with ductal carcinoma in situ who have highly dense breasts [5,6].

Mammography has major problems due to high breast density which obscures the mammographic image. The main drawback of mammography today is that it is hard to differentiate between normal, dense tissue and cancerous tissue when looking for small tumors surrounded by glandular tissue. A woman's breasts are naturally denser, or more glandular when young, which makes it difficult for the radiologist to analyze the mammogram image. Technology to detect breast cancer is changing rapidly, with recent entrants to the field like digital mammography and computer-aided detection. Enhancing the image by manipulation of fine differences in intensity by means of image processing algorithms forms the basis of any computer aided detection system. In this work breast tissue is classified based on the intensity level of histogram of a mammogram using SVM. Statistical features of a mammogram are extracted using simple image processing techniques. This technique uses texture models to capture the mammographic appearance within the breast.

1.1. Paper outline

Section 2 provides background information about breast cancer and Section 3 describes mammography and computer assisted screening mammography. A literature review of related computer methods applied in mammography is provided in Section 4. Section 5 gives the description of the proposed method. Artifact removal is discussed in Section 6 and pectoral muscle extraction technique is presented in Section 7. Section 8 covers statistical feature extraction and Section 9 describes the support vector machine (SVM) classifier. Breast tissue classification using SVM is illustrated in Sections 10 and 12 concludes the paper.

2. Breast cancer overview

Breast cancer is the one of the commonest malignancies afflicting women. Despite all medical and technological advances, breast

* Corresponding author. Fax: +91 4144 238275.
E-mail addresses: rtramsuba@yahoo.com (T.S. Subashini), aucsevr@yahoo.com (V. Ramalingam), spal_yughu@yahoo.com (S. Palanivel).

1077-3142/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2009.09.009

cancer cases as been on the rise in the last 50 years or so. It is alarming but it is true that there is more women affected by breast cancer now than ever before. It is currently estimated that one in 14 of all female children born will develop breast cancer in their lifetime. Globally, every 3 min a woman is diagnosed with breast cancer in the world, amounting to 1 million cases annually. The incidence could go up by 50% to 1.5 million by 2020, says the World Cancer Report. The incidence of breast cancer is rising in every country of the world especially in developing countries such as India. Breast cancer is the second leading cause of cancer deaths in women today (after lung cancer) and the most common cancer among women, excluding nonmelanoma skin cancers. According to the WHO, an estimated 1.2 million people worldwide were diagnosed with breast cancer in 2004. Estimates indicated that in another 46,400 women, ductal DCIS, a noninvasive breast cancer, were diagnosed. The incidence of breast cancer increased by approximately 4% during the 1980s but leveled off to 100.6 cases per 100,000 women in the 1990s. According to a study by International Agency for Research on Cancer (IARC), there will be approximately 250,000 new cases of breast cancer in India by 2015 [7–9]. The report shows that around 1.7 million breast cancers were diagnosed worldwide in 2007 and 465,000 (approx.) women died due to breast cancer in 2007. What is more alarming, is that 75–80% of patients are in advanced stages of the disease at the time of diagnosis [10].

If the relationship between tissue density and breast cancer risk is to be studied, a more accurate and objective method of assessing tissue density is needed. In this work an attempt has been made to automate classification of breast density using statistical features.

3. Digital mammograms and computer aided diagnosis (CAD)

In standard mammography, images are recorded on film using an X-ray cassette. The film is viewed by the radiologist using a "light box" and then stored in a jacket in the facility's archives. With digital mammography, the breast image is captured using a special electronic X-ray detector, which converts the image into a digital picture for review on a computer monitor. The digital mammogram is then stored on a computer. With digital mammography, the magnification, orientation, brightness, and contrast of the image may be altered after the exam is completed to help the radiologist more clearly see certain areas. A landmark trial, conducted by the American College of Radiology Imaging Network (ACRIN) in conjunction with the Center for Statistical Sciences at Brown Medical School, showed no difference between digital and film mammography in detecting breast cancer for the general population of women [11]. However the trial shows that digital mammography detected more cancers that is up to 28% more than screen film mammography in women under 50 years of age and women with dense breasts.

Benefits of digital mammograms are quicker mammograms, since there is no need to wait for film images to be developed, images can be viewed instantly by the technologist and radiologist, images can be easily transferred electronically with no loss of image quality and digital mammography allows radiologists to use computer software to manipulate the images in order to optimize their ability to evaluate the breast tissue that might be missed on traditional film mammograms.

Computer aided detection (CAD) increases the detection of early breast cancers, especially those in women with dense breast tissue. With the sharp surge in digital mammography system implementations, CAD is riding the curve, too. According to many radiologists this "second set of eyes" that helps to confirm radiological findings in mammograms, makes sense from both a cost perspective and the potential it offers to detect more breast lesions. CAD is easy to integrate into that process because it is so accessible and available on demand either at the location where the images are taken or at remote offices and home offices that have the necessary high-speed internet connections and high-resolution monitors as well as other equipments.

4. Previous work

Radiologists mainly estimate breast density by visual judgement of the imaged breast. Thus automatic tissue classification methods try to imitate such visual judgment, learning from the radiologist experience. In the literature different approaches for classifying breast tissue based only on the use of histogram information have been proposed [12]. BPNN and histogram features were used in [13] for classifying breast densities. A first approach to qualify mammograms according to the radiographic density was the Wolf scheme [14–16] aimed at finding a correlation between density and cancer risk, but the technique lacked objectivity due to intra and inter observer variations. Recently, researchers have studied intensity-histogram features and applied threshold techniques and fractal characteristics to analyze radiographic density in digital images [17–19].

[20] shows that texture information described by multi-scale histogram based on multi-class DAG-SVM classifier is useful in classifying breast densities.

A semiautomatic computer measure based on interactive thresholding and the percentage of the segmented dense tissue over the segmented breast area has been proposed in [21]. In [22] measures were based on skewness and fractal dimension. Texture-based discrimination between fatty and dense breast types applying granulometric techniques and Laws texture masks has been investigated in [23]. Another method based on fractal dimension is proposed in [22]. Spatial gray level dependency matrices were constructed and features we estimated based on these matrices in[24] to classify breast tissue. In [25] SFS + kNN (sequential forward selection) classifier and morphological and textural features were used for classification. Based on co-occurrence matrices [26,27] segmented mammograms into density regions. Textons are used in [28] to classify breast tissue. However in many of the above approaches the entire breast including the pectoral muscle has been proposed to extract features. The inclusion of the pectoral muscle can affect the results of intensity based image processing methods in the detection of breast densities. In our approach an attempt has been made to restrict the region of interest the breast tissue alone by eliminating the artifacts, background and the pectoral muscle. Statistical features are extracted from this well focussed region which signifies the important texture features of the breast tissue. Numerous techniques have been proposed for breast density pattern classification. Bayesian classifier was used in classifying the breast tissue in [22,21,29]. kNN was used in [26,27,30,31]. In [32,33], rule-based classifiers were used to classify breast tissue density. In the proposed method statistical feature that were extracted are fed to the support vector machine classifier to classify it into any of the three classes namely fatty, glandular and dense tissue.

When mammograms are analyzed by computer, the pectoral muscle should preferably be excluded from processing intended for the breast tissue. In the literature different approaches for automatic pectoral muscle segmentation have been proposed. Segmentation of the breast, and the pectoral muscle are often prerequisites for automatic assessment of breast density. In the work of others, the pectoral boundary has usually been approximated by a straight line as the first step [34,30,35,36]. The straight-line approximation has been determined by a number of techniques, including region growing [34], the Hough transform [30,35,37] and local adaptive thresholding followed by line fitting [36]. Once obtained, the straight-line approximation can be refined to follow the slightly curved outline of the pectoral muscle [36,37] as a further step.