COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 107 (2012) 175–188

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Technique for preprocessing of digital mammogram

*Indra Kanta Maitra, Sanjay Nag, Samir Kumar Bandyopadhyay**

Dept. of Computer Science & Engg., University of Calcutta, 92 A.P.C. Road, Kolkata 700009, West Bengal, India

A R T I C L E I N F O

Article history:
Received 26 August 2010
Received in revised form
11 April 2011
Accepted 18 May 2011

Keywords:
Mammogram
Contrast limited adaptive histogram equalization (CLAHE)
Region of interest (ROI)
Seeded region growing (SRG)
Noise
Edge-shadowing effect and pectoral muscle

A B S T R A C T

Digital mammogram has emerged as the most popular screening technique for early detection of breast cancer and other abnormalities in human breast tissue. It provides us opportunities to develop algorithms for computer aided detection (CAD). In this paper we have proposed three distinct steps. The initial step involves contrast enhancement by using the contrast limited adaptive histogram equalization (CLAHE) technique. Then define the rectangle to isolate the pectoral muscle from the region of interest (ROI) and finally suppress the pectoral muscle using our proposed modified seeded region growing (SRG) algorithm. The proposed algorithms were extensively applied on all the 322 mammogram images in MIAS database resulting in complete pectoral muscle suppression in most of the images. Our proposed algorithm is compared with other segmentation methods showing superior results in comparison.

1. Introduction

Cancer is a group of diseases that cause cells in the body to change and grow out of control. Most types of cancer cells eventually form a lump or masses called a tumor, and are named after the part of the body where the tumor originates. Breast cancer begins in breast tissue, which is made up of glands for milk production, called lobules, and the ducts that connect lobules to the nipple. The remainder of the breast is made up of fatty, connective, and lymphatic tissue [1].

Breast cancer is a leading cause of cancer deaths among women. For women in US and other developed countries, it is the most frequently diagnosed cancer. About 2100 new cases of breast cancer and 800 deaths are registered each year in Norway [2]. In India, a death rate of one in eight women has been reported due to breast cancer.

Early and efficient detection is the most effective way to reduce mortality, and currently a screening programme based on mammography is considered one the best and popular method for detection of breast cancer. Mammography is a low-dose X-ray procedure that allows visualization of the internal structure of the breast. Mammography is highly accurate, but like most medical tests, it is not perfect. On average, mammography will detect about 80–90% of the breast cancers in women without symptoms [3].

An increasing number of countries have started mass screening programmes that have resulted in a large increase in the number of mammograms requiring interpretation [4]. In the interpretation process radiologists carefully search each image for any visual sign of abnormality. However, abnormalities are often embedded in and camouflaged by varying densities of breast tissue structures. Estimates indicate that between 10% and 30% of breast radiologists miss cancers

* *Corresponding author*. Tel.: +91 9830554236.
E-mail addresses: ikm.1975@yahoo.com (I.K. Maitra), sanjaynag75@gmail.com (S. Nag), skb1@vsnl.com (S.K. Bandyopadhyay).
0169-2607/$ – see front matter
doi:10.1016/j.cmpb.2011.05.007

during routine screening [4,5]. The images provided by different patients have different dynamics of intensity and present a weak contrast. Moreover the size of the significant details can be very small. Several research works have tried to develop computer aided diagnosis tools that could help the radiologists in the interpretation of the mammograms and could be useful for an accurate diagnosis [6–8].

Digital mammography is a technique for recording X-ray images in computer code instead of on X-ray film, as with conventional mammography. The first digital mammography [9] system received U.S. Food and Drug Administration (FDA) approval in 2000. Digital mammographies have advantage over conventional mammography as the images can be stored and retrieved electronically.

Imaging techniques play an important role by helping to perform digital mammogram, especially of abnormal areas that cannot be physically felt but can be seen on a conventional mammogram [10]. Before any image-processing algorithm can be applied on mammogram, preprocessing steps are very important in order to limit the search for abnormalities without undue influence from background of the mammogram. Breast segmentation consists of breast border contour extraction, pectoral muscle extraction, nipple identification etc. On images obtained directly from the digital mammography devices segmentation process is much easier. Previous work from many authors used mammography image databases including this paper, especially MiniMIAS [11] and DDSM [12]. In this paper we have proposed three steps of preprocessing of raw digital mammogram to isolate pectoral muscle from digital mammogram.

In the second section of this paper we have discussed different image segmentation techniques and their role in image analysis. In the third section we have discussed region growing algorithm and more specifically seeded region growing algorithm. The fourth section we have reviewed several works performed on seeded region growing algorithm by different authors. The fifth section contains the core discussion of our proposed work. This section has three subsections that describe the enhancement of digital mammogram (CLAHE) followed by defining the rectangle for pectoral muscle isolation. The final subsection we perform suppression of pectoral region. The sixth section we define the dataset we have used for evaluation. The section seven and eight cites the execution evaluation and result analysis. We have performed failure assessment in section nine and we have compared our results with the results obtained by other authors in section ten. Finally we have concluded our discussion in section eleven.

2. Image segmentation

The paper is based on the image segmentation method, which refers to the major step in image processing, the inputs are images and, outputs are the attributes extracted from those images. Segmentation divides image into its constituent regions or objects. The level to which segmentation is carried out depends upon the problem being solved i.e. segmentation should stop when the objects of interest in an application have been isolated. Image segmentation refers to the decomposition of a scene into its components. Image segmentation techniques can be broadly classified as into five main classes threshold based, Cluster based, Edge based, Region based, Watershed based segmentation [13].

Segmentation plays an important role in image analysis. The goal of segmentation is to isolate the regions of interest depending on the problem and its characteristics. Many applications of image analysis need to obtain the regions of interest before the analysis can start. Therefore, the need of an efficient segmentation method has always been there. A gray level image consists of two main features, namely region and edge.

Segmentation algorithms for gray images are generally based on two basic properties of image intensity values, discontinuity and similarity. In the first category, the approach is to partition an image based on abrupt changes in intensity, such as edges in an image. The principle approaches in the second category are based on partitioning image into regions that is similar according to a set of predefined criteria. Thresholding, region growing, region splitting and merging are examples of the methods in this category.

3. Region growing

For the segmentation of intensity images like digital mammograms, there are four main approaches [14,15], namely, threshold techniques, boundary-based methods, region-based methods, and hybrid techniques which combine boundary and region criteria.

Threshold techniques [16] are based on the postulate that all pixels whose value (gray level, color value, or other) lies within a certain range belong to one class. Such methods neglect all of the spatial information of the image and do not cope well with noise or blurring at boundaries. Boundary-based methods [17] use the postulate that the pixel values change rapidly at the boundary between two regions. The complement of the boundary-based approach is to work with the regions [18]. Region-based methods rely on the postulate that neighbouring pixels within the one region have similar value. This leads to the class of algorithms known as region growing of which the “split and merge” technique [19] is probably the best known. The general procedure is to compare one pixel to its neighbour(s). If a criterion of homogeneity is satisfied, the pixel is said to belong to the same class as one or more of its neighbours. The fourth type is the hybrid techniques, which combine boundary and region criteria. This class includes morphological watershed segmentation [20] and variable-order surface fitting [14]. The watershed method is generally applied to the gradient of the image.

We use a method known as “seeded region growing” (SRG), which is closer to that of the watershed [21] with some necessary change in proposed technique, which is based on the conventional region-growing postulate of similarity of pixels within the regions. For seeded region growing (SRG), seed or a set of seeds can be automatically or manually selected. Their automated selection can be based on finding pixels that are of interest, e.g. the brightest pixel in an image can serve as a seed pixel. They can also be determined from the peaks found in an image histogram. On the other hand, seeds can also be selected manually for every object present in the image.