Computers and Mathematics with Applications 64 (2012) 1100–1107

Contents lists available at SciVerse ScienceDirect

Computers and Mathematics with Applications

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

A pectoral muscle segmentation algorithm for digital mammograms using Otsu thresholding and multiple regression analysis

Chen-Chung Liu [a], Chung-Yen Tsai [b], Jui Liu [c], Chun-Yuan Yu [b], Shyr-Shen Yu [b,*]

[a] Department of Electronic Engineering, National Chin-Yi University of Technology, Taichung 411, Taiwan

[b] Department of Computer Science and Engineering, National Chung-Hsing University, Taichung 400, Taiwan

[c] Department of General Medicine, Royal North Shore Hospital, Sydney, Australia

ARTICLE INFO

Keywords:
Mammogram
Pectoral muscle
Otsu thresholding
Multiple regression analysis (MRA)

ABSTRACT

One of the issues when interpreting a mammogram is that the density of a pectoral muscle region is similar to the tumor cells. The appearance of pectoral muscle on medio-lateral oblique (MLO) views of mammograms will increase the false positives in computer aided detection (CAD) of breast cancer. For this reason, pectoral muscle has to be identified and segmented from the breast region in a mammogram before further analysis. The main goal of this paper is to propose an accurate and efficient algorithm of pectoral muscle extraction on MLO mammograms. The proposed algorithm is based on the positional characteristic of pectoral muscle in a breast region to combine the iterative Otsu thresholding scheme and the mathematic morphological processing to find a rough border of the pectoral muscle. The multiple regression analysis (MRA) is then employed on this rough border to obtain an accurate segmentation of the pectoral muscle. The presented algorithm is tested on the digital mammograms from the Mammogram Image Analysis Society (MIAS) database. The experimental results show that the pectoral muscle extracted by the presented algorithm approximately follows that extracted by an expert radiologist.

1. Introduction

Breast cancer is the most common cancer in women — which constitutes approximately 25% of all cancers in women, and every one in eight women develop breast cancer. Successful treatment relies on early detection, and screening mammography targeting high risk groups is a common practice in developed countries.

Extracting the breast region accurately from a mammogram is a kernel stage for mammography. It significantly influences the overall analysis accuracy and processing speed of the whole breast mass analysis. It is due to the density of a pectoral muscle region being similar to that of the mammographic parenchyma (tumor). Many studies on tumor detection on a mammogram have shown that the appearance of pectoral muscle in medio-lateral oblique (MLO) views of mammograms will increase the false positive in computer aided detection (CAD) of breast cancer [1–3]. Therefore, successful identification and segmentation of pectoral muscle from the breast region on a mammogram before further analysis should improve the accuracy when interpreting the mammogram.

Only a few studies have been presented in the literature to address the pectoral muscle detection. Nagi et al. [4] used morphological preprocessing and seeded region growing to detect the pectoral muscle. Yapa et al. [5] segment the pectoral muscle region by utilizing the combination of an improved fast-marching method and mathematical morphological

* Corresponding author.
E-mail addresses: ccl@ncut.edu.tw (C.-C. Liu), james1022jk@gmail.com (C.-Y. Tsai), titoakl@yahoo.com (J. Liu), t112@nkut.edu.tw (C.-Y. Yu), epyu@nchu.edu.tw (S.-S. Yu).

0898-1221/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2012.03.028

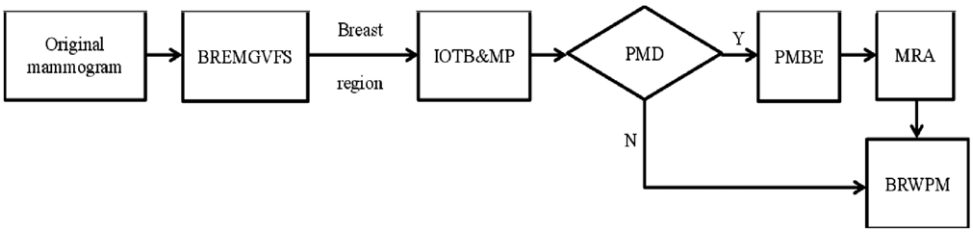

Fig. 1. The flow chart of the proposed pectoral muscle segmentation algorithm for mammograms.

operators such as area morphology, alternating sequential filter, openings and closings. Kwok et al. [6] detected the pectoral muscle edge in a restricted neighborhood derived by estimating the pectoral muscle edge as a straight line segment, and then refined the detected edge by surface smoothing. However, the hypothesis of a straight line segment for the representation of the pectoral muscle edge is not always correct.

Multiple regression analysis (MRA) is a statistical method used to model the relationships between several independent variables and a dependent variable by fitting an equation to the observed data [7,8]. Numerous experiments have shown that multiple regression models can be used to make accurate predictions. Multiple regression procedures are very widely used in the social and natural sciences today [9]. It is also a suitable technique for us to modify the extracted pectoral muscle border to increasing the precision of the proposed pectoral muscle segmentation algorithm.

The proposed pectoral muscle segmentation algorithm bases on the positional characteristic of pectoral muscle in a breast region to combine the iterative Otsu thresholding scheme and the mathematic morphological processing to find a rough border of the pectoral muscle. Once the pectoral muscle border is extracted, the multiple regression analysis is then employed on the rough border to obtain an accurate segmentation of the pectoral muscle. The presented algorithm is tested on the digital mammograms from the Mammogram Image Analysis Society (MIAS) database. The experimental results show that (i) the pectoral muscle extracted by the presented algorithm approximately follows that extracted by an expert radiologist, (ii) the proposed scheme is adaptive to large variations in appearance of the pectoral muscle; it remains effective when the pectoral border is obscured by superimposed muscle tissue or artifacts, (iii) the proposed scheme will decrease the false positives in computer aided detection (CAD) of breast cancer. The remainder of this paper is organized in the following order: Section 2 introduces the presented extraction algorithm of pectoral muscle region from mammograms. Section 3 presents the experimental results. The conclusions of this paper are presented in Section 4.

2. Proposed pectoral muscle segmentation algorithm

In order to construct an accurate pectoral muscle segmentation algorithm for digital mammograms, several schemes are used in this paper. The overall pectoral muscle segmentation algorithm for digital mammograms is shown in Fig. 1. There are mainly three stages involved in the algorithm: (i) breast region extracted by using modified gradient vector flow (MGVF) snake, (ii) rough pectoral muscle region detection by using the iterative Otsu thresholding scheme, (iii) pectoral border modification by utilizing multiple regression analysis. The details of these stages used in the presented pectoral muscle segmentation algorithm are described in the following subsections.

2.1. Breast region extracted by using modified gradient vector flow (MGVF) snake [10]

To segment a digital mammogram into the breast region and the background is an essential step for pectoral muscle segmentation. By extracting the breast region, the precision of the pectoral muscle segmentation is increased and the time consumption is shortened. The breast region extraction algorithm must be fully automated and give correct results for all digital mammograms. The breast region segmentation by using an MGVF snake is employed in this paper, a method that was introduced in our previous paper: "*a breast extraction scheme for digital mammograms using gradient vector flow snake*" [10]. The proposed breast region extracting algorithm integrated the median filtering step, the scale down step, the binarization processing step, the morphological erosion processing step and novel gradient adjusting step. The median filter step was used to filter out the noise in a mammogram, the scale down step was used to resize down the mammogram size to speed up the breast region extraction. The binarization processing step and the morphological erosion processing step were used to find a rough breast border. The novel gradient adjusting step was applied to get a modified edge map; and then the gradient vector flow snake (GVF snake) was employed to obtain an accurate breast border from the rough breast border. This breast extraction scheme can output the corresponding accurate breast region from an input digit mammogram, and the output breast region will be used in the following stage to detect a rough border of the pectoral muscle.