Pectoral muscle deletion on a mammogram to aid in the early diagnosis of breast cancer

Pascal Vagssa¹*, Nafissatou Mallam Doudou², Tchoning Jolivo³, Olivier Videme⁴, Dina Taïwé Kolyang¹

¹LaRI Lab, University of Maroua, CAMEROON
²MINRESI/Agence Nationale de la Radio protection, Garoua, CAMEROON
³Department of Physics, University of Ngaoundéré, CAMEROON
⁴The National Advanced School of Engineering, University of Yaoundé, CAMEROON

*Corresponding Author: e-mail: pvagssa@gmail.com, Tel: + (237) 677 302 439

ORCID iD: http://orcid.org/0000-0002-2934-5155 (Vagssa), https://orcid.org/0000-0002-4472-8760 (Doudou), https://orcid.org/0000-0003-0367-3721 (Jolivo), https://orcid.org/0000-0002-0840-2405 (Videme)

Abstract

Mammograms are the images used by radiologists to diagnose breast cancer. In this diagnosis, the pectoral muscle appears in mammograms in oblique mediolateral views (MLO) of the right breast and another in the left breast appears in crano-caudal views which are marked with (CC). Considering that the pectoral muscle has the same density as the small, suspicious masses in the image, its presence in the image being processed could also require detection procedures. In this paper, we present a new general framework for pectoral muscle suppression which is the first work in the analysis of a mammography image. As a result, we proceed to four stages of image processing. The first step is to orient the image if necessary, then use a pre-processing which is to enhance the contrast of the image, and remove the digital lines of the image by morphological filters, apply a filter median. The third step involves segmenting all of the pectoral muscles, which involves threshold the entire image. The final step is to perform a pectoral muscle removal according to the orientation of the muscle in the image, which will be based on the development of the Hough transform for the recognition of borderline detections of the pectoral muscle. Some results obtained on the different images are discussed and compared with other methods (risk assessments). Evaluation of our method shows a significant improvement in performance in removing the pectoral muscle.

Keywords: Breast cancer, Mammogram, Pectoral muscle, Hough transform.

DOI: http://dx.doi.org/10.4314/ijest.v12i3.6

Cite this article as:

Received: January 19, 2020; Accepted: July 8, 2020; Final acceptance in revised form: August 25, 2020

1. Introduction

Breast cancer, a major public health problem, continues to be the leading cause of death in women around the world. More than 2,500 cases of breast cancer diagnosed each year (Jemal et al., 2011). The reduction in the death rate caused by this type of cancer as well as the promotion of the chances of recovery are only possible if the tumor has been detected and taken care of from the early stages of its onset (Jemal et al., 2011). The incidence of this serious pathology remains increasing in Cameroon, only one cancer in ten is diagnosed and the clinical characteristics and histology's of the latter are poorly described (Bouhnik et al., 1994). Mammography imaging is X-ray imaging of the breast, and using X-rays as the base technology makes mammography distinct
from other imaging techniques. Mammograms are usually high-resolution images with high bit-depth, which provides the possibility of discovering abnormalities masked by surrounding and overlapping breast tissue. It makes it possible to observe anomalies which may reflect malignant or benign lesions. This involves the transmission of radiation through the tissue and the projection of anatomical structures onto a film screen or image sensor. Two image projections of each breast are obtained: Cranio-Caudal view (CC) and Mediolateral Oblique view (MLO) (Figure 1). However, the number of images that the radiologist must interpret in a limited time is large and constitutes a difficulty for interpretation. Thus, diagnostic aid techniques are being developed to facilitate this interpretation. Our perspective is to set up a Diagnostic Assistance System (CAD) for breast cancer (Figure 2), the basis of which is the study of mammography images. The first step before the operation is the removal of the pectoral muscle from the images.

CAD systems should include multiple different possibilities for image enhancement, automatic segmentation, and registration, and, in the ideal scenario, automatic detection algorithms for various possible features that could indicate an abnormality. The main preprocessing technique in CAD applications is defining the proper region of interest which implies some sort of image segmentation and object detection. After the segmentation, the user of a CAD system can start all other automatic detection and diagnosis tools on the way to bring a reliable and quick diagnosis (Mario et al., 2015).

In computer-aided diagnosis systems for breast mammography, the pectoral muscle region can easily cause a high false-positive rate and misdiagnosis due to its similar texture and low contrast with breast parenchyma. Pectoral muscle region segmentation is a crucial pre-processing step to identify lesions, and accurate segmentation in poor-contrast mammograms is still a challenging task. In order to tackle this problem, a novel method is proposed to automatically segment pectoral muscle region in this paper. It takes place as follows:
- guidance if the need arises; mammograms in mediolateral oblique views (MLO) have two orientations (left and right), to simplify the task for the algorithm set up we have opted for the left orientation in this work;
- image preprocessing: which enhances the contrast of the image, removing digital lines from the image by morphological filters;
- closure suppression (it is based on the development of the Hough transform for the recognition of limit detections of the pectoral muscle)

The purpose of this article is to provide a technique for removing the pectoral muscle on mammography. Indeed, the development of a technique like the one we are proposing would reduce errors when analyzing mammography images. Thus, first of all, we propose a pre-processing step which has the role of enhancing the contrast of the image. We go to the threshold of the latter, and we end with the removal of the pectoral muscle. In this work, we have adopted Hough's algorithm, we exploit its simplicity in implementation and its importance in pattern recognition in imagery.

The rest of the document is organized as follows: In section 2, we review the literature. This section presents the different techniques implemented for the pre-processing and for the removal of the pectoral muscle on mammograms. Section 3 describes the materials and methods offered. It presents the different stages of the pre-processing and the technique used for the removal of the pectoral muscle on the mammogram. This deletion will first allow you to better locate and observe the development of a potential anomaly. Section 4 presents the results of the developed study method. It explains the contours and specificity of our work. Finally, the last section shows the conclusion and future direction of our research.

2. Literature review

Many approaches have been developed for pectoral muscle segmentation on mammography. According to recent research in this area by (Mustra et al., 2016), the most commonly used approaches contain straight line estimation, region growth, threshold for edge detection and polynomial fit. The mini MIAS database is the most widely used database, and only a few approaches evaluate all mammograms in this database (Suckling et al., 1994).

Therefore, linear estimation is a very intuitive approach. It first estimates a straight line, then refines the line boundary with a change in gray intensity or gradient in the region of interest (ROI). But in many cases, the authors have assumed the boundaries of the pectoral muscle as straight lines. Thus (Karssemeijer, 1998) was one of the first authors to use a straight line to identify the pectoral muscle. (Kwok et al., 2001) presented a straight line method to estimate the edge of the pectoral muscle and refined the edge detected by cliff detection and surface smoothing. About 94% of the images were considered to be segmented in an acceptable manner. (Kinoshita et al., 2008) proposed a straight line method based on radon domain information. In the random domain, several suitable straight lines were detected as candidates to represent the pectoral muscle boundary. The average FP and FN were 8.99 and 9.13%, respectively.

In this sense, the Hough transform has been widely used (Ferrari et al., 2004a; Yam et al., 2001; Weidong et al., 2007). Ferrari et al. (2004a) and Yam et al. (2001) through this transform find a peak in Hough space, except that Yam et al. (2001) refined the linear approximation in a pectoral boundary curve using a dynamic programming method. Kwok et al. (2004) used an algorithm and a linear approximation to fit the edges of the pectoral muscle and the Hough transform was used to identify the lines. This method is heuristic dynamic programming in which the linear approximation uses algorithmic search. It is also claimed to be more robust to noise and bright spots in the image, thus giving less broken edges (Karssemeijer, 1998). Mustra et al. (2009) presented a method of segmentation and suppression of the pectoral muscle on mammograms. They used the wavelet transformation for proper sub-band selection and obtained 85% acceptable pectoral muscle segmentation results. Vaidehi and Subashini (2013) used the straight-line method to demarcate the border between the pectoral muscle and the breast, for this purpose for a total of 120 mammography images used for the study. The accuracy obtained was 85% and 90.2% true positive rate achieved. Chen et al. (2015) proposed a shape-based detection method in extracting the pectoral muscle boundary on a mammogram. This method produced an acceptable rate of 97.2% in the MIAS database. Recently, Yoon et al. (2016) proposed a threshold method with morphological operations and a random sample consensus algorithm (RANSAC). They obtained an acceptable rate of 92.2% with the MIAS database.

Several other works have been carried out in the same context, we find Alam and Islam (2014) which detected the pectoral muscle using the region growth method and obtained an accuracy of about 90.3%. Raba et al. (2005) presented a segmentation method using region growing algorithm. Breast orientation was used to initial seed and a size restriction was applied to avoid a wrong growing. After the detection process, a morphological operation was used to refine the boundary. Segmentation methods based on active contours are also widely used (Daniel et al., 1987). They have been used in the detection of the borders of the pectoral muscle on mammograms, for example we have the work of Wirth et al. (2004) where he segments the regions of the breast using the active contours. To prevent false segmentation of mammograms that have a part of a well-contrast pectoral muscle region, polynomial fitting of the seed points is proposed to predict the boundary in the poor-contrast part. Xu et al. (2009) presented an optimal threshold method in combination with Hough transform and poly-line fitting. After finding an optimal threshold, they extracted points from the initial pectoral muscle mask and performed poly-line fitting in Hough space.

Of all the methods for modeling pectoral muscle on mammography by detection of the limit, the Hough transform has often been used (Ferrari et al., 2004a; Sultana et al., 2010). Many of these works give acceptable results but not a perfectly precise and often
poor performance in the segmentation of the pectoral muscle mammography, so we look disappointed to improve with the method we have proposed.

3. Material and method

Figure 3 shows the different steps of our method of removing the pectoral muscle on a mammogram.

![Diagram of the proposed method]

**Figure 3**: Diagram of the proposed method

3.1 Orientation

In order to simplify our method of deletion, first, we are looking for orientation (Figure 4). Thus each mammogram is detected according to the method introduced by Shah (2015). This method initially divides the image into equal parts (left part, right part) and calculates the sum of each part. So if the sum of the left part is greater than the sum of the right part, then the orientation of the breast is left, else breast orientation is right. Left orientation is achieved by flipping the image using the MATLAB fliplr function.

![Result of the orientation for the image (mdb001.pgm)]

**Figure 4**: Result of the orientation for the image (mdb001.pgm)
3.2 Pre-processing

Typically, grayscale images are observed by image acquisition methods that exhibit noise from electronic devices. We will therefore, achieve an improvement by subjecting it to a pre-treatment (noise reduction and artifact removal). To improve the visual quality of the image, the effects of noise (interference) must be eliminated by subjecting it to a treatment called filtering. We thus use a 2D median filter (Figure 5). After filtering, we convert our gray level mammography image to a binary image using a given threshold. This results in a binary black and white mask. Then, an opening operation is used to remove the abnormal spots in the breast area (the white area). Note that artifacts such as labels present on mammograms can be removed after performing the image threshold. To keep the large component connected, a component labeling process is then applied to isolate the different classes in order to extract the breast region and remove artifacts. We thus keep the part whose area is greater than a certain threshold. The threshold image is still called the initial breast mask and the one without artifacts is called the final breast mask (Figure 5).

3.3 Removal of the pectoral muscle

The algorithm applied for the removal of the pectoral muscle uses the Hough transformation illustrated in the following steps:

Step 1: First, we recover the oriented and preprocessed image.

Step 2: Find the region of interest containing the Hough line. To do this, we create a region of interest. Our region of interest is defined by the upper left quarter of the preprocessed image.

Step 3: Contour detection. Apply the Canny filter, the objective of which is to highlight all the contours of the region of interest (Figure 6).

Step 4: We perform linear aperture filtering to bring out the boundaries of the image.

Step 5: Line detection and creation of the Hough mask (Figure 6).

Here we use the following Hough’s algorithm adapted to the context of the study.

Note that at this level, our region of interest is 512 x 512 pixels in size and the representation of a line by the Hough transformation is defined by

\[ \rho = x \cos \theta + y \sin \theta \]  

(Ferrari et al., 2004b).

(1) Start

(2) Quantify the parameter space (region of interest) with a maximum and a minimum for the 2 parameters. For a number of points of 128, the parameter \( \theta \) is included in the interval \([\theta_{\min}, \theta_{\max}]\) and \( \rho \) included in the interval \([\rho_{\min}, \rho_{\max}]\) with \( \theta_{\min} = 25\pi / 180; \theta_{\max} = 40\pi / 180; \rho_{\min} = 1 \) and \( \rho_{\max} = (\sum \text{size (Image)}^2)^{\frac{1}{2}} \)

(3) Initialize an accumulator A (dim ((\( \rho \)), dim ((\( \theta \)))) to 0. For each pixel (x, y) of the image, calculate \( \rho = x \cos \theta + y \sin \theta \)

(4) Find the maximum (s) of matrix A

(5) Creation of the Hough mask.

The couples \((\rho, \theta)\) selected to characterize the lines of the pectoral muscle.

(6) End

Step 6: Apply the Hough mask to the region of interest (Figure 6)

Step 7: The region of interest is recovered by doing a simple multiplication between the filtered image and the final mask filtered by a binary dilation. The resulting outline is superimposed in white on the filtered image (Figure 6).

4. Results and discussion

4.1 Database

The approach proposed in this study was applied to all 322 images in the MIAS database; these images are in PGM format and corresponding to the left and right breast, freely available online for scientific purposes (Rangayyan et al., 2000). Each pixel is described as an 8-bit word. Each image is the size of 1024 x 1024. The 322 images are divided into 207 normal images, 38 images containing masses, and 169 images containing other anomalies.

4.2 Pre-processing

We applied the pre-processing procedure used here on the 322 images of the MIAS database. Figure 5 presents 5 images with different noise levels, notably labels and artifacts. The images resulting from the pre-processing process are presented (Figure 5) and are based on the steps of the process used (filtering (2D median filter), binarization of the image using a given threshold, these two operations give us a binary mask in black and white Then the operations (opening on the image to remove the abnormal points in the breast region, the threshold to remove the artifacts and labels), and at the end, we will recover the preprocessed image).
In the previous figure, we can observe three of the main types of noise that are removed from our original image: the clear rectangular labels (mdb002.pgm, mdb012.pgm, mdb112.pgm, and mdb007.pgm), the weak labels intensity (mdb002.pgm, mdb007.pgm, and mdb119.pgm), intensity sweep lines and artifacts (mdb012.pgm, mdb112.pgm, mdb119.pgm). We can observe that all these different noises have been perfectly eliminated. In addition, the contrasts have been improved.

4.2 Removal of the pectoral muscle

In this part, we will see the results of the actual deletion. Thus, removal is of good quality if the pectoral muscle is correctly removed and the removal of artifacts is satisfactory (Figure 6). On the other hand, removal is not of good quality if part of the muscle has remained after removal or the removal has affected part of the breast (Figure 6). Note that in this study the muscle appears in a mediolateral view and represented as a triangle in an upper corner of the image. Thus our method realizes the spots (looking for the orientation of the image, in this case, we use the left orientation then reorient the opposite case, identify the dividing line between the muscle and the breast tissue and finally apply the Hough mask to remove the muscle to extract the region of interest). Our algorithm, shown in Figure 4, recovers the region of interest from the preprocessing step before applying the Hough algorithm, in order to have a mask (Hough mask). However, there is a problem at this step which is due to the size of the region of interest which is sometimes difficult to apply on some images. The results are presented in Figure 6, our method tested on all the images of the MIAS database provides a visibly acceptable result. Note that the removal procedure has a lower success rate than the pre-processing procedure. However, we can manually improve the results of the erroneous images, so we will simply have to play on the size of the region of interest selected in order to facilitate the task of our algorithm used.
### Table 1: Comparison of the proposed method with some methods in the literature

<table>
<thead>
<tr>
<th>Image Id</th>
<th>Original image With Artifact</th>
<th>Breast Region after removing artefact</th>
<th>Hough mask</th>
<th>Breast Region after removal of Pectoral region</th>
</tr>
</thead>
<tbody>
<tr>
<td>mdb002.pgm</td>
<td><img src="image1" alt="Original image" /></td>
<td><img src="image2" alt="Breast Region" /></td>
<td><img src="image3" alt="Hough mask" /></td>
<td><img src="image4" alt="Breast Region" /></td>
</tr>
<tr>
<td>mdb012.pgm</td>
<td><img src="image1" alt="Original image" /></td>
<td><img src="image2" alt="Breast Region" /></td>
<td><img src="image3" alt="Hough mask" /></td>
<td><img src="image4" alt="Breast Region" /></td>
</tr>
<tr>
<td>Mdb112.pgm</td>
<td><img src="image1" alt="Original image" /></td>
<td><img src="image2" alt="Breast Region" /></td>
<td><img src="image3" alt="Hough mask" /></td>
<td><img src="image4" alt="Breast Region" /></td>
</tr>
<tr>
<td>Mdb007.pgm</td>
<td><img src="image1" alt="Original image" /></td>
<td><img src="image2" alt="Breast Region" /></td>
<td><img src="image3" alt="Hough mask" /></td>
<td><img src="image4" alt="Breast Region" /></td>
</tr>
<tr>
<td>mdb119.pgm</td>
<td><img src="image1" alt="Original image" /></td>
<td><img src="image2" alt="Breast Region" /></td>
<td><img src="image3" alt="Hough mask" /></td>
<td><img src="image4" alt="Breast Region" /></td>
</tr>
</tbody>
</table>

**Figure 6:** Removal pectoral muscle result

### 4.3. Comparison of results with existing methods

In this work, our results have been grouped into two categories (acceptable and unacceptable). An acceptable image if the pectoral muscle is correctly or the majority of the pectoral muscle is removed and the removal of artifacts is satisfactory, and an image is not acceptable if most of the muscle is not removed or not removed at all. Table 1 presents a comparison of the proposed method with some methods in the literature which use the same performance evaluation methods. From this table, it can be seen that this method gives good results compared to other methods in the literature. This result indicates that our algorithm defines more pixels as being a pectoral muscle region when they are not. In contrast, the lowest unacceptable frame rate is achieved using our proposed approach. Also, out of all 322 images, the approach gets better quality results, including smaller errors.
5. Conclusion

In this article, we present a method that removes the pectoral muscle on digital mammograms using the Hough transformation. Mammography images acquired from MIAS databases have been preprocessed to reduce noise and remove artifacts so that the removal process can work accurately. Next, we obtained a mask using the Hough algorithm which we applied to the resulting region of the pretreatment step to get our mammogram without the pectoral muscle. The deletion accuracy rate is approximately 93.8%. However, the algorithm could not effectively remove some images. Nonetheless, it is expected that the proposed method will be complemented by the use of a metaheuristic to optimize image removal. In the future, the result of the proposed method should be useful for cancer diagnostic support systems on mammograms.

Acknowledgement

We like to thank the editor and the anonymous reviewers for their valuable remarks that helped us in better improving the content and presentation of the paper.

References


### Table 1. Comparison between proposed work and previous works

<table>
<thead>
<tr>
<th>Methods</th>
<th>Images</th>
<th>Acceptable</th>
<th>Unacceptable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mustra et al. (2009)</td>
<td>40</td>
<td>85%</td>
<td>15%</td>
</tr>
<tr>
<td>Vaidehi and Subashini (2013)</td>
<td>120</td>
<td>85%</td>
<td>15%</td>
</tr>
<tr>
<td>Alam and Islam (2014)</td>
<td>322</td>
<td>90.3%</td>
<td>9.7%</td>
</tr>
<tr>
<td>Yoon et al. (2016)</td>
<td>322</td>
<td>92.2%</td>
<td>7.8%</td>
</tr>
<tr>
<td><strong>Our method</strong></td>
<td>322</td>
<td>93.8%</td>
<td>6.2%</td>
</tr>
</tbody>
</table>

Biographical notes

**Pascal Vagssa** obtained his Master in EEA (Electronics, Electrotechnics, and Automatic) at the University of Ngaoundéré in 2013. Still the same year he obtained diploma as a Design Engineer in Telecommunications at the National Advanced School of Maroua. In 2016 he obtained Teacher Diploma from the Secondary Technical Teaching Schools in Electronics at ENSET Douala/Cameroon. Currently, he is involved in the processing of biomedical images.

**Nafissatou Mallam Doudou** received master in radiology and medical imagery at the University of Ngaoundéré/Cameroon, 2014. Presently she is researcher at National Agency for Radiation Protection. She is interested in biomedical image processing.

**Tchoning Jolivo** received masters in EEA (Electronics, Electrotechnics, and Automatic) at the University of Ngaoundéré in 2013. Presently, Tchoning is interested in image processing and is a researcher at the *Department of Physics, University of Ngaoundéré*.

**Olivier Videme** is a Professor at The National Advanced School of Engineering of Yaoundé/Cameroon, Cameroon. Polytechnician, obtained his Doctorate/PhD since 2007. He is interested in Telecommunications.

**Dina Taïvé Kolyang** is Full professor at Cameroonian universities since 2016. Coordinator of the Computer Science Research Laboratory (LaRI) of the University of University/Cameroon. His areas of research are Ontologies, data mining, formal methods.