DESIGN OF PECTORAL MUSCLE SEGMENTATION ALGORITHM IN MAMMOGRAPHY BASED ON MORPHOLOGICAL OPERATIONS

JUAN SEBASTIAN DIAZ PATIÑO

INDUSTRIAL UNIVERSITY OF SANTANDER
FACULTY OF PHYSICAL-MECHANICAL ENGINEERING
SCHOOL OF ELECTRICAL, ELECTRONIC AND TELECOMUNICATIONS ENGINEERING
BUCARAMANGA
2021

DESIGN OF PECTORAL MUSCLE SEGMENTATION ALGORITHM IN MAMMOGRAPHY BASED ON MORPHOLOGICAL OPERATIONS

JUAN SEBASTIAN DIAZ PATIÑO

A thesis to get the degree of Electronic Engineer

Director:
SAID DAVID PERTUZ ARROYO
PhD in Computer Science

INDUSTRIAL UNIVERSITY OF SANTANDER
FACULTY OF PHYSICAL-MECHANICAL ENGINEERING
SCHOOL OF ELECTRICAL, ELECTRONIC AND TELECOMUNICATIONS ENGINEERING
BUCARAMANGA
2021

AGRADECIMIENTOS

Agradezco a mi familia por su inmenso esfuerzo para apoyar cada paso de mi proceso de formación. A mi director de proyecto quien de la mejor manera mostro el camino para llevar a cabo este trabajo y a cada una de las personas que estuvieron presente para motivarme.

CONTENTS

pag.

LIST OF TABLES

pag.

LIST OF FIGURES

pag.

RESUMEN

TÍTULO: DISEÑO DE ALGORITMO DE SEGMENTACIÓN MUSCULAR PECTORAL EN MAMOGRAFÍA BASADO EN OPERACIONES MORFOLÓGICAS *

AUTOR: JUAN SEBASTIAN DIAZ PATIÑO**

PALABRAS CLAVE: CAD, ROI, SEGMENTACIÓN, MAMOGRAFIA.

DESCRIPCIÓN: Sistemas de detección y diagnóstico asistidos por computadora han demostrado ser útiles en el diagnóstico precoz del cáncer de mama. En el caso de proyecciones de mamografía oblicua medio lateral, la visibilidad del músculo pectoral afecta el rendimiento del diagnóstico. La segmentación de esta estructura mamaria es entonces una de las principales tarea para un diagnóstico eficaz. Basado en la hipótesis de que el músculo y el tejido mamario denso tienen los mismos niveles de radio densidad, en este trabajo propusimos una segmentación método basado en operaciones morfológicas. Se realizaron experimentos en 210 imágenes del conjunto de datos INBreast disponible públicamente donde las segmentaciones obtenidas se comparan con referencias anotadas manualmente obtenidas de lectores experimentados. Los resultados obtenidos sugieren que el método propuesto funciona a la par con el estado de los métodos, al tiempo que reduce el número de segmentaciones fallidas.

---

*Trabajo de grado

** Facultad de Ingeniería Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director: Said David Pertuz Arroyo, PhD en Ciencias de la Computación

ABSTRACT

TITLE: DESIGN OF PECTORAL MUSCLE SEGMENTATION ALGORITHM IN MAMMOGRAPHY BASED ON MORPHOLOGICAL OPERATIONS [*]

AUTHOR: JUAN SEBASTIAN DIAZ PATIÑO[**]

KEY WORDS: CAD, ROI, SEGMENTATION, MAMMOGRAPHY.

DESCRIPTION: Computer-aided diagnostic and detection systems have proven to be helpful in the early diagnosis of breast cancer. In the case of medio-lateral oblique mammography projections, the visibility of pectoral muscle affects the performance of the diagnosis. Segmentation of this breast structure is then a major task for an effective diagnosis. Based on the hypothesis that the muscle and the dense breast tissue have the same radio-density levels, in this work we proposed a segmentation method based on morphological operations. Experiments were performed on 210 images of the publicly available INBreast dataset where the obtained segmentations are compared with manually annotated references obtained from experienced readers. Obtained results suggest that the proposed method performs on par with state of the methods, while reducing the number of failed segmentations.

---

[*] Bachelor Thesis

[**] Faculty of Physical-Mechanical Engineering. School of Electrical, Electronic and Telecomunications Engineering. Director: Said David Pertuz Arroyo, PhD in Computer Science

INTRODUCTION

Cancer is the second leading cause of death in women worldwide[1]. Breast cancer is the most common invasive cancer in women. Mammograms are used as the main screening and diagnostic tool, as they are an inexpensive, efficient, and minimally invasive alternative for obtaining images for subsequent diagnosis. To facilitate the radiologist's work in identifying those suspicious areas that require a more in-depth analysis, computer assistance systems or CAD systems are increasingly used, which can be utilized partially or totally throughout the detection and diagnosis process.

As these systems have shown the potential to improve risk assessment and diagnosis[2], many efforts focus on making them more reliable[3]. To take advantage of the full potential of these systems, it is necessary to perform some preliminary tasks, such as background and breast detection. These tasks are relatively easy for

[1] CANCER FACT SHEETS, [Website] Available from internet: https://gco.iarc.fr/today/home.

[2] P. A. Habas, *et al.* Probabilistic Framework for Reliability Analysis of Information-Theoretic CAD Systems in Mammography, International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, 2006, pp. 6113-6116.

[3] ANDREADIS, Ioannis I. *et al.* A CADx Scheme for mammography empowered with topological information from clustered microcalcifications' atlases. IEEE Journal of Biomedical and Health Informatics, pp. 166–173.

craniocaudal (CC) mammograms since the breast region can easily be separated from the background. However, this is not the case with the medio-lateral oblique (MLO) view, because part of the pectoral muscle is captured in this view and must be segmented. This segmentation is usually performed by hand by the radiologist on duty, but due to the difficulty and experience required with these procedures, the performance of CAD systems can be affected. For this reason, it is increasingly imperative to leave these detection and segmentation tasks to algorithms that, in an automated way, can perform the segmentation reliability at the same level as an expert radiologist.

The segmentation of the pectoral muscle is a difficult task due to its lack of regularity and also because it has intensities of gray and a texture similar to that of breast tissue, which does not allow simple and standardized strategies to be applied for its detection. This problem has generally been tackled using classical digital image processing techniques. In[4], the real limits of the breast are estimated by means of a borderless active thresholding and contouring process. After this, the limits of the pectoral muscle are estimated by means of a canny border segmentation algorithm. In[5], they use a strategy that combines the use of geometric figures with the region growing technique to perform segmentation of the pectoral muscle. In[6] a pre-processing and segmentation of the pectoral muscle is applied on the base to morphological operations, in which basic operations are applied to eliminate artifacts and reduce noise in the image, Subsequently, a watershed algorithm is used for the

[4]RAMPUN, Andrik, *et al.* Fully automated breast boundary and pectoral muscle segmentation in mammograms. Artificial Intelligence in Medicine, pp. 28–41.

[5]TAGHANAKI, Saeid Asgari, *et al.* Geometry-Based Pectoral Muscle Segmentation from MLO Mammogram Views. IEEE Transactions on Biomedical Engineering, pp. 2662–2671.

[6] TAIFI, Khaddouj, *et al.* Automatic breast pectoral muscle segmentation on digital mammograms using morphological watersheds. Proceedings - 2017 14th International Conference on Computer Graphics, Imaging and Visualization, CGiV 2017, pp. 126–131.

segmentation. In[7] is to use an algorithm to get the optimal threshold based on the average gray level.

The two main aspects to comment regarding the performance of all the aforementioned methods are the performance measures used for the evaluation and adjustments of their respective parameters and the datasets used for the experiments. In the first aspect, we have found little that performance measures are rarely reported in a standardized way, which makes it difficult to compare different methods, in addition to the fact that the measures used in many cases provide little information. The other aspect is the databases where the experiments are carried out. Many of the methods use datasets like the publicly available (mini)-MIAS and the DDSM datasets [8] that consist of scanned screen-film mammograms in an era of full-field digital mammography.

The main objective of this work is to design a method for the detection and segmentation of the pectoral muscle based on morphological operations that are at the performance level of methods in the state of the art. For comparison purposes, we adopt the performance measures proposed in[9] since they provide and standardized procedure and the use of a publicly available database that allows easy validation and, in turn, good reproducibility in the results.

[7] KOLAHDOUZAN, F. (2014). Isfahan University of Medical Sciences Isfahan University of Technology. May 2006.

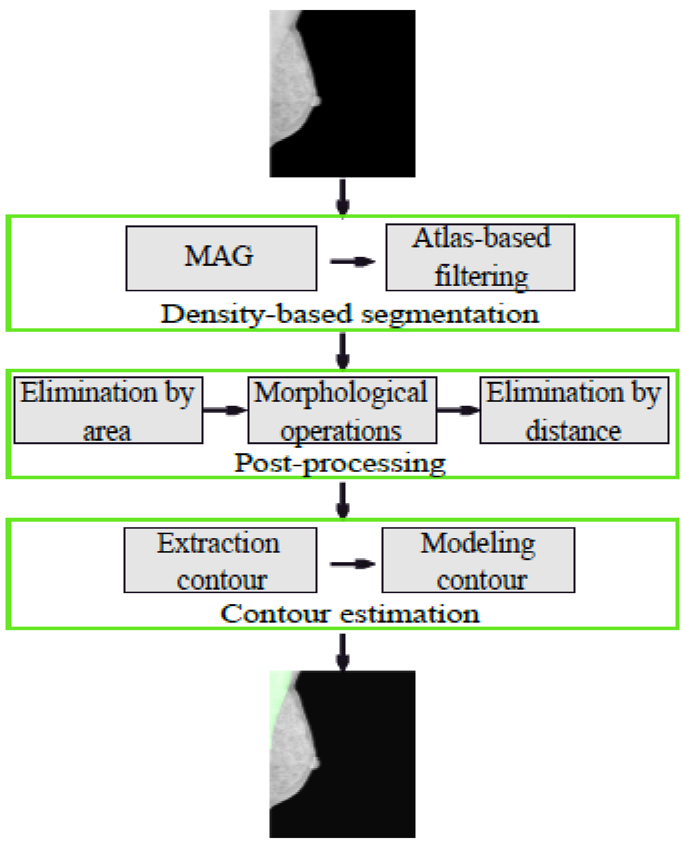
[8] MUSTRA, Mario; GRGIC, Mislav and RANGAYYAN, Rangaraj M. Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms. Medical and Biological Engineering and Computing, pp. 1003–1024.

[9] AFRICANO, Gerson, *et al*. A New Benchmark and Method for the Evaluation of Chest Wall Detection in Digital Mammography. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2020-July, pp. 1132–1135.

1. METHODS

Figure 1 shows a general description of the proposed segmentation method divided into three steps. The first step is the density-based pre-segmentation and atlas-based filtering. Second, post-processing is applied where small, spurious, and disconnected objects resulting from the previous process are filtered through morphological operations. Finally, contour estimation refers to the processing and contour modeling to obtain the approximation of the pectoral muscle.

Figure 1. Overview of the pectoral muscle segmentation process.



1.1 DENSITY-BASED SEGMENTATION

The first step is to use the method developed in[10], called morphological area gradient (MAG) designed to segment dense tissue in the full-field digital mammography (FFDM). This is followed by a filtering step based on a probabilistic map[11], as described below.

1.1.2 Mag segmentation. The first step is based on the hypothesis that muscle tissue has levels of gray intensity and texture, similar to fibroglandular tissue in the breast area. Based on this, we use the MAG method since it allows the segmentation of radio-dense tissue, see figure 3a.

To apply the MAG method for the detection of the pectoral muscle, it is first necessary to segment the background of the mammogram. This task can be performed in digital mammograms at the threshold value found from the image histogram. After this, the optimal threshold i for segmentation is as:

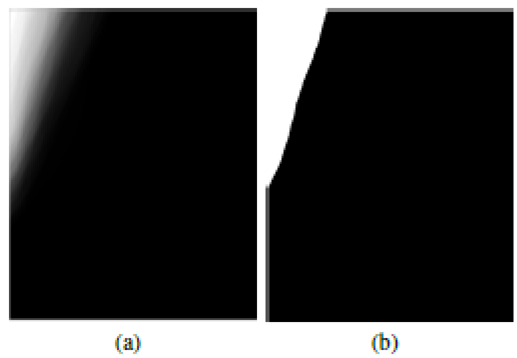
$$i = argmin \frac{1}{M} \nabla\{A(n)\} \qquad (\mathbf{1})$$

Where $A(n)$ is the morphological area, which is defined as $|\, I\,(x, y) \geq i_n|$, where $I\,(x, y)$ the gray intensity value located at position (x, y), m is the total number of pixels in the foreground of the chest, and $i_n \in \{i_1, i_2, .., i_N\}$ is the nth level of gray in the range $\min\{I(x, y)\}, \max\{I(x, y)\}$. The total number of gray levels is N and $\nabla\{.\}$ is the first difference operator. After calculating the gray value i, we obtain an initialization binary mask $M_0(x, y)$ by thresholding image $I(x, y)$ at level i_n.

[10] G. F. Torres *et al*. Morphological area gradient: System-independent dense tissue segmentation in mammography images. In Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2019, pp. 4855–4858.

[11] P. A. Habas, *et al.* Probabilistic Framework for Reliability Analysis of Information-Theoretic CAD Systems in Mammography, International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, 2006, pp. 6113-6116.

Figure 2. a) Atlas or probabilistic map in grayscale. b) Atlas after the thresholding process.

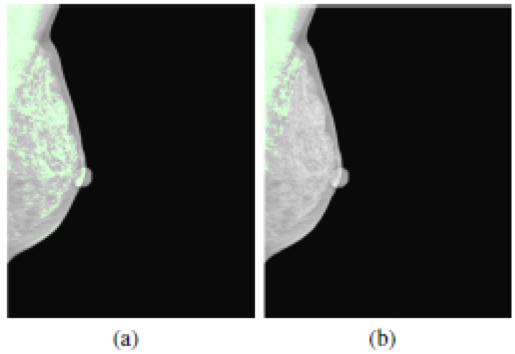


1.1.2 Atlas-based filtering. The binary mask M_0 resulting from the previous process includes dense tissue from both the pectoral muscle and the fibroglandular breast tissue. In order to obtain only the pectoral muscle, an atlas-based process elimates the fibroglandular tissue by means of a probability map. The atlas is a grayscale image $A(x,y)$, as shown in Figure 2a, where the areas of higher intensity represent a higher probability of the pixel to be part of the pectoral muscle. This atlas is obtained by building a probability map from the ground-truth segmentation of a training set of images. The filtering process consists of two steps. First, a reference mask $M_0(x,y)$ is obtained by thresholding atlas $A(x,y)$ at probability level p see Figure 2b. Subsequently, the segmented pectoral muscle is obtained by the intersection of masksM_0 and M_A:

$$M_A(x,y) = \begin{cases} 1 \; si \; A(x,y) > p \\ 0 \quad otro \; valor \end{cases} \quad (2)$$

$$M(x,y) = M_A(x,y) \cap M_0(x,y) \quad (3)$$

Figure 3. Results of the segmentation algorithm color green reflects the segmented tissue a) Results of the MAG method b) Results of the atlas filtering process.

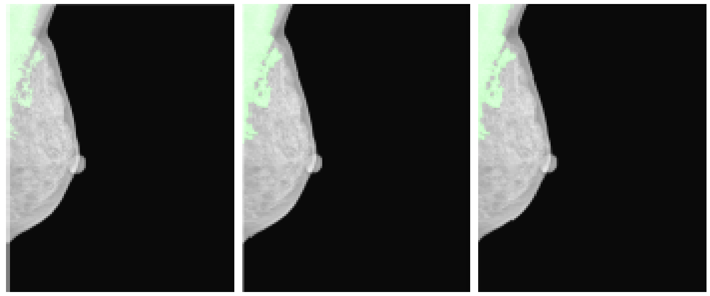


1.2 POST-PROCESSING

After the segmentation process, mask M often has small, spurious, unconnected objects remnants of the filtering process or small areas outside the pectoral muscle region. The presence of these objects can lead to a decrease in the performance of the algorithm. We solve this problem using morphological operations. First, we apply an area-based morphological filtering to remove all objects with areas below α see figure 4a. This area-based filtering is followed by two additional morphological operations[12]: a dilation with a diamond-shaped structuring element of radius $r_1 = 1.6\ \mathrm{mm}^2$, followed by an erosion with a circular structuring element radius $r_1 = 0.4\ \mathrm{mm}^2$ see figure 4b. Finally, a distance filtering is performed based on the average distance of the object towards the coordinates (1,1) of mask M, those objects whose distance is greater than ε are eliminated see figure 4c.

Figure 4. Images resulting from the post-processing process color green reflects the segmented tissue a) Elimination result by area b) Results of morphological operations c) Elimination result by area.

[12] J. Serra, Image Analysis and Mathematical Morphology. New York: Academic, 1982, vol. 1.



(a) (b) (c)

1.3 CONTOUR ESTIMATION

Finally, the contour of the mask resulting from the previous processes is extracted and modeled.

1.3.1 Contour extraction. The 2D outline of the mask resulting from the previous processes is extracted. The process consists of drawing the boundaries between the foreground and background pixels, the boundaries of the holes or objects within the mask are extracted. This is done to increase the amount of data and therefore the resolution of the next process.

1.3.2 Contour modeling. Once the 2D contour is obtained, it is modeled using a second or third-degree polynomial since they present a closer approximation to the segmentation performed by the expert radiologist. This polynomial is obtained using RANSAC as described in[13]. To decide what degree of a polynomial will be used for each mask, area error AEr as defined in the referenced method[14] is used as an

[13]MOLINARA, M.; MARROCCO, C. and TORTORELLA, F. Automatic segmentation of the pectoral muscle in mediolateral oblique mammograms. Proceedings of CBMS 2013 - 26th IEEE International Symposium on Computer-Based Medical Systems, pp. 506–509.

[14]AFRICANO, Gerson, *et al*. A New Benchmark and Method for the Evaluation of Chest Wall Detection in Digital Mammography. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2020-July, pp. 1132–1135.

error measure. The procedure is as follows: two contours are modeled with polynomials of second and third order, we model two contours as shown in figures 5a and 5b, the AEr of each one is calculated concerning the original mask and the final model is the one that gives the lowest error. In this way, the segmentation and segmentation of the pectoral muscle are obtained, seen in figures 6a and 6b.

Figure 5. Estimated pectoral muscle contour a) third degree b) second degree.

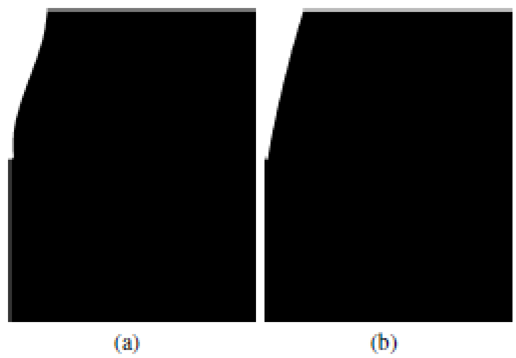
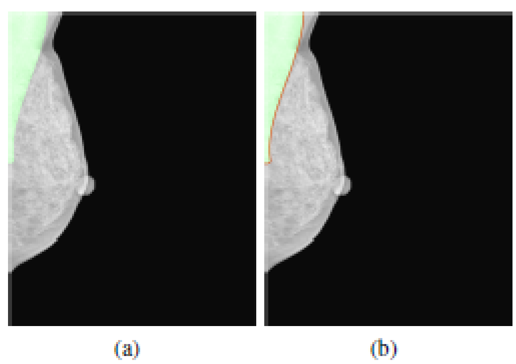


Figure 6. Estimated pectoral muscle a) pectoral muscle estimated in mammogram b) estimated contour.



1.4 PERFORMANCE MEASURES

The performance measures used to evaluate the performance of this method were those proposed in[15], which are: Dice similarity coefficient DSC, area error AEr and mean contour distance MCD. DSC is defined as:

$$\mathrm{D}SC = 2\frac{|M \cap M_1|}{|M| + |M_1|} \qquad (4)$$

Here M is the segmentation resulting from the method and M_1 is the reference segmentation performed by the expert radiologist and $|\,.\,|$ is the cardinal operator in set theory.

$$AEr{=}s|M{\cap}M$$_1 |+|$$M{\cap}M$$_1 |. (5)

The AEr is a measurement that represents how much area of the pectoral muscle has been over-segmented or under-segmented, it is a measure that is measured in mm^2.

$$MCD = 0{,}01 \sum \|p_i - p_i^*\| \qquad (6)$$

Where $p_i^* \in C_0$ is the closest point to a $p_i \in C$ being C_0 and C the contour of the algorithm and reference algorithm respectively.

2. EXPERIMENTS AND RESULTS

[15]Ibid.

In this section we evaluate the proposed method, for reference purposes, we compare with the state-of-the-art method presented in[16].

2.1 DATABASE

INbreast is a publicly available *dataset* with 410 FFDM images including CC and MLO views[17]. This *dataset* contains manually annotated information related to the pectoral muscle and suspicious regions within the breast. From the full data set, MLO view images were extracted, which is a total of 210 mammograms. We call this new subset of images INbreast full, which was divided into training and test sets, following the recommendations set forth in the reference method[18]. The INbreast training contains 60% of the INbreast full images and they were chosen through a broiling process. All images were acquired with MammoNovation Siemens FFDM with 0.07 mm / pixel, in 14-bit format.

2.2 PARAMETERS

The presented method has a total of four parameters. Firstly, the resolution at which the method named s works, secondly the probability level p used to threshold the atlas in the filtering process, and lastly the area α and distance ε from removed used in post-processing. Method parameters were adjusted from a grid search. It should be noted that this adjustment was made in the "training" data set and all values remain fixed in the rest of the experiments. The ranges of these parameters and their

---

[16] Ibid.

[17] MOREIRA, Ines C., *et al*. INbreast: Toward a Full-field Digital Mammographic Database. Academic Radiology, pp. 236–248.

[18] AFRICANO, Gerson, *et al*. A New Benchmark and Method for the Evaluation of Chest Wall Detection in Digital Mammography. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2020-July, pp. 1132–1135.

units can be seen in table 1. The value of the parameters obtained are the following $s = 0.2$, $p = 0.083$, $\alpha = 1700\ \mathrm{mm}^2$ and $\varepsilon = 50\ \mathrm{mm}$.

Table 1**.** Ranges, units and resolution of the parameters used.

Parameter	**Unit**	**Range**	**Resolution**
s	mm/pixels	[0.1−1]	0.1
p	probability levels	[0.08−0.09]	0.01
α	mm^2	[1000−2000]	100
ε	mm	[40−60]	2

2.3 RESULTS

The results obtained are shown in table 2. The Wilcoxon rank-sum test was applied to determine if the results obtained in the proposed method were statistically significant compared to the reference method, it was used because the measurements did not pass the normality test (chi-square test, $p < 0.01$) table 3 shows a comparison between the number of images with null segmentation in proposed method and reference method, understanding like null segmentation images where only it is background black see figure 7b. In this aspect, the method proposed yields a significant improvement with respect to the reference method. This can be observed in figure 7 where figure 7a is the mask binary resulting from the segmentation algorithm in both methods. The figure 7b and figure 7c are the end result of each respective method.

Table 2. Performance of the proposed method compared to the referenced method.

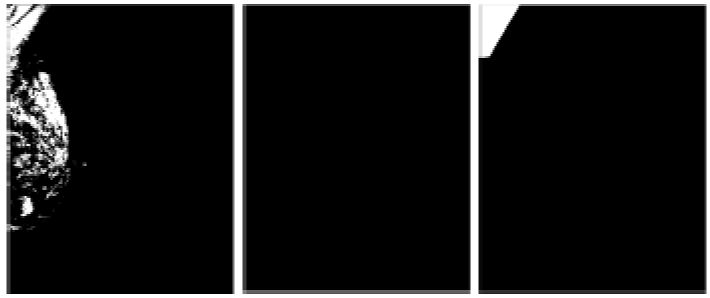
Measure	**Method proposed**	**Method referenced**
INbreast-training		
$\boldsymbol{MCD}$ **(**mm)	4,04	3,66

AEr (mm^2)	738	734
DSC	0,905	0,906
INbreast-test		
$\boldsymbol{MCD}$ **(**mm)	513	4,96
AEr (mm^2)	970	1169
DSC	0,900	0,878
INbreast-full		
$\boldsymbol{MCD}$ **(**mm)	4,18	4,30
AEr (mm^2)	694	829
DSC	0,910	0,888

Table 3. Number of images with null segmentation comparing between the proposed method and reference method.

Images with null segmentation	**Method proposed**	**Method referenced**
INbreast-training	1	24
INbreast-test	1	12
INbreast-full	1	25

Figure 7. Comparison between the final segmentation referenced method and the proposed method a) segmentation performed with the MAG method b) Result obtained by the referenced method c) Result obtained by the proposed method.



(a) (b) (c)

3. CONCLUSSIONS

In this work, a new method of segmentation and segmentation of the pectoral muscle was proposed using as a basis the method presented in the referenced method, finding a superior performance in two of the three performance measures, although the results of the indicated method were not statistically representative. However, the advantage was the reduction of images with null segmentation as seen in table 3, which is a consequence of the post-processing that allows the recovery of certain FFDM that in the first segmentation steps show a pectoral muscle almost null or null see figure 7.

BIBLIOGRAPHY

AFRICANO, Gerson, et al. A New Benchmark and Method for the Evaluation of Chest Wall Detection in Digital Mammography. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2020-July, pp. 1132–1135.

ANDREADIS, Ioannis I. et al. A CADx Scheme for mammography empowered with topological information from clustered microcalcifications' atlases. IEEE Journal of Biomedical and Health Informatics, pp. 166–173.

CANCER FACT SHEETS, [Website] Available from internet: https://gco.iarc.fr/today/home.

TORRES G. F. et al. Morphological area gradient: System-independent dense tissue segmentation in mammography images. In Annual International

SERRA, J, Image Analysis and Mathematical Morphology. New York: Academic, 1982, vol. 1.

KOLAHDOUZAN, F. (2014). Isfahan University of Medical Sciences Isfahan University of Technology. May 2006.

MOLINARA, M.; MARROCCO, C. and TORTORELLA, F. Automatic segmentation of the pectoral muscle in mediolateral oblique mammograms. Proceedings of CBMS 2013 - 26th IEEE International Symposium on Computer-Based Medical Systems, pp. 506–509.

MOREIRA, Ines C., et al. INbreast: Toward a Full-field Digital Mammographic Database. Academic Radiology, pp. 236–248.

MUSTRA, Mario; GRGIC, Mislav and RANGAYYAN, Rangaraj M. Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms. Medical and Biological Engineering and Computing, pp. 1003–1024.

SERRA, J., Image Analysis and Mathematical Morphology. New York: Academic, 1982, vol. 1.

P. A. Habas, et al. Probabilistic Framework for Reliability Analysis of Information-Theoretic CAD Systems in Mammography, International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, 2006, pp. 6113-6116.

RAMPUN, Andrik, et al. Fully automated breast boundary and pectoral muscle segmentation in mammograms. Artificial Intelligence in Medicine, pp. 28–41.

TAGHANAKI, Saeid Asgari, et al. Geometry-Based Pectoral Muscle Segmentation from MLO Mammogram Views. IEEE Transactions on Biomedical Engineering, pp. 2662–2671.

TAIFI, Khaddouj, et al. Automatic breast pectoral muscle segmentation on digital mammograms using morphological watersheds. Proceedings - 2017 14th International Conference on Computer Graphics, Imaging and Visualization, CGiV 2017, pp. 126–131.