

Automatic Segmentation of Regions of Interest in Breast Thermographic Images

Adrian J. Villalobos-Montiel¹, Mario I. Chacon-Murguia^{1(✉)},
Jorge D. Calderon-Contreras¹, and Leticia Ortega-Maynez²

¹ Visual Perception Applications on Robotics Lab,
Chihuahua Institute of Technology, Chihuahua, Mexico
{avillalobos, mchacon, jdcalderon}@itchihuahua.edu.mx

² Universidad Autonoma de Ciudad Juarez, Ciudad Juarez, Mexico
lortega@uacj.mx

Abstract. Breast thermography is a promising technique allowing breast cancer detection with the aid of infrared technology. However, the automatic segmentation of the regions of interest (ROI) to be analyzed is a difficult task and, thus, it is not commonly performed. In this paper we propose an automated technique for ROI extraction. The algorithm uses Canny edge detection with automatic threshold and symmetry inspection in order to correct tilt errors and to define a symmetry axis in the image. Furthermore, the Hough transform is used to provide circles giving a favorable approximation to the ROI. A region-based active contours technique and an edge smoothing technique are also used to improve the previous segmentation, taking in consideration the energies of the images. Experimental results indicate that the ROI were extracted more accurately and with higher precision than a state-of-the-art method.

Keywords: Breast cancer · Thermography · Breast segmentation · Active contours

1 Introduction

Breast cancer is a disease in which a group of malignant cells located in the breast grow uncontrollably and invade surrounding tissues or spread (metastasize) to distant areas of the body. It figures among the leading causes of death worldwide. The World Health Organization reported that in 2012, there were 521,000 deaths due to this disease [1]. Early cancer detection methods not only increase the chances of survival, but also allow treatments that will avoid disfiguring surgery.

Breast thermography is a promising noninvasive technique specializing in the early detection of breast cancer through thermal images. Due to the chemical and physiological principles, in precancerous, cancerous and surrounding tissues, there is a greater blood supply and an increased cellular activity which causes an increase in temperature equal to or greater than 0.5 °C compared to other tissues. Thermographic images are captured with infrared (IR) cameras, and can be digitally analyzed in order to detect any abnormalities within the breast tissue [2]. Breast thermography is a safe technique because it emits no radiation, and it is more comfortable and less expensive compared to other screening methods. With this technique, it is possible to detect small (<0.8 cm)

non-palpable lesions, that cannot be detected by physical examination or with a mammography [3]. Adding a thermography study to mammography provides a 95 % of accuracy in the diagnosis, whereas mammography by itself has only an 80 % of accuracy in the early stages of cancer [4]. Additionally, breast thermography allows tumors to be detected even 10 years earlier than with mammography [5].

Segmentation of the ROI allows delimitation of the data to be analyzed by the computer aided diagnosis system (CAD). The ROI must include all the breast tissue and the near ganglion groups since cancerous cells usually appear in the glands that produce milk and the ducts that carry it to the nipples [6]. Most authors perform semi-automatic or manual ROI extraction because it is a hard task due to the inherent characteristics of each breast that make them amorphous, and to the lack of clear limits in this kind of images [6].

In reference [6], a study of several algorithms found in the literature that perform ROI extraction is presented. These algorithms are usually based on a combination of some of the following methods: borders detection, parabolic Hough transform, active contours, anisotropic filters and interpolation methods [6]. However, most algorithms fail to perform a correct segmentation in images where the borders are not clear due to camouflage. So the main change on segmenting the ROI, is usually detecting the lower bounds of the breasts.

In this study we present an algorithm that performs the ROI segmentation from breast thermographic images. The proposed technique can perform this task correctly even when there is a camouflage problem. The algorithm also performs a tilt correction through symmetry analysis. Figure 1 summarizes the entire process in a block diagram.

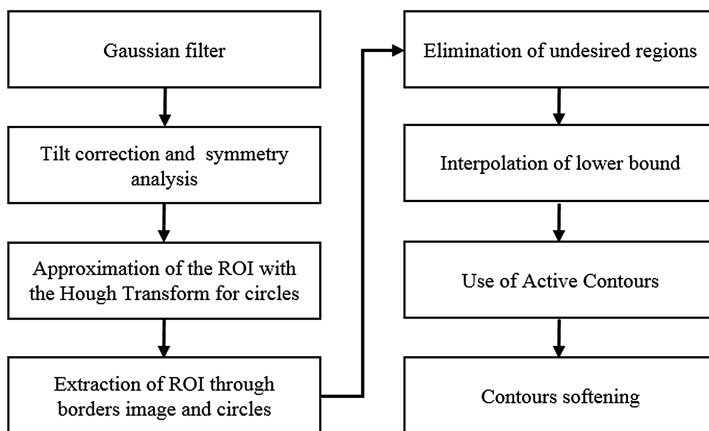


Fig. 1. Block diagram of the proposed algorithm.

Forty thermographic images from the database [7] (only available database found) were used in the development of the ROI extraction algorithm, and other different 100 images from the same database were used to evaluate the performance of the proposed algorithm. The proposed algorithm is then compared to the algorithm of Marques [8] which, through thresholding, obtains some of the lower edges of the breasts, and then

with different interpolation methods completes the lower bounds of the ROI. However, the thresholding method of Marques fails when processing breast thermographies with unclear bounds.

2 Methods

Forty thermographic images from the database from [7] were used in the development of the ROI extraction algorithm. Figure 1 summarizes the entire process in a block diagram.

In order to avoid detection of false borders, the first step of the method is noise reduction through a Gaussian filter ($\sigma = 0.5$) within the breast thermography to generate the image $I_G(x,y)$.

2.1 Tilt Error Correction

Due to difficulty to keep a patient still during the image capture, patients tend to incline toward a side. Tilt error correction is necessary due that it can affect the diagnosis [9]. Symmetry analysis enables to correct tilt errors, as well to obtain important information that can be used to identify the ROI. To perform this correction, the algorithm identifies the symmetry axis in the proper position through an iterative process. A border image, $I_B(x,y)$, is obtained using the Canny operator to the filtered image $I_G(x,y)$. The Canny threshold, C_{TH} , is fixed to a value that allows detecting the lower borders of the breasts on at least 90 % of the input images. Then, a dilation operation (performed 4 times) is used over each edge pixel $p_e(x,y)$ to obtain thicker edges

$$\begin{aligned} & \text{for each } p_e(x,y) \\ & N_8(p_e(x,y)) = 1 \end{aligned} \quad (1)$$

This image is rotated initially -10° , but in future iterations it is rotated with 1° interval increments until 10° is reached (values can be modified to desired range). The rotation operation is defined by

$$x_2 = \cos(\theta)(x_1 - x_0) - \sin(\theta)(y_1 - y_0) + x_0 \quad (2)$$

$$y_2 = \sin(\theta)(x_1 - x_0) + \cos(\theta)(y_1 - y_0) + y_0 \quad (3)$$

where θ is the angle of rotation in a clockwise direction and (x_0, y_0) are the coordinates of the center of rotation [10]. The obtained image is copied and pasted in the center of a blank image with greater size with same proportions (recommended ratio of 1.5) denoted as $I_R(x,y)$. This step allows increasing the working space for the future steps. Afterwards, the size of a window W is defined based on the biggest size of breasts that can be found in any breast thermography. This window will continuously be placed in different positions over $I_R(x,y)$ in order to find the symmetry axis. Initially, it is centered in the y axis at the left border of $I_R(x,y)$. But in future iterations it will be horizontally shifted to the right in intervals of 1 pixel until reaching the right border of $I_R(x,y)$. Then two window images of the same size, W_R and W_L , are obtained by cutting W vertically.

The obtained image on the right side, W_R , is then reflected horizontally through a morphological operation, and denoted as W_{RR} . In order to find the similarities between the two images obtained from W , an image W_{AND} is generated by

$$W_{AND} = W_L(x,y) W_{RR}(x,y) \tag{4}$$

W_{AND} is a binary image indicating the pixels that W_{RR} and W_L have in common at same positions. The pixels with a value of 1 in W_{AND} are counted and denoted as the symmetry level. The algorithm uses 3 conditions of higher symmetry given by the variables γ_S , θ_A and α_P , which correspond respectively to the symmetry level, the angle of rotation and the position of W in the x axis of $I_R(x,y)$. Iteratively, the algorithm tests the different angles of rotation, and the different horizontal shift positions of W , calculating on every combination the symmetry level. If the symmetry level of the current iteration is higher than γ_S , the variables are overwritten with the current level of symmetry, angle of rotation, and shift position. Once the algorithm has tested all the possible combinations, a W_{AND} image is generated with the values stored in the higher symmetry variables. Ideally, the resulting W_{AND} image will show the borders of a breast and the body. Furthermore, the symmetry axis can be easily and correctly placed as a vertical line over the rotated thermography by an angle of θ_A , and at position λ in the x axis based on α_P and the width of W_{AND} , ω_{AND} :

$$\lambda = \alpha_P + \omega_{AND} \tag{5}$$

The rotated thermography will have no tilt error. The symmetry analysis steps are illustrated in Fig. 2. A more extended description of this method can be found in [11].

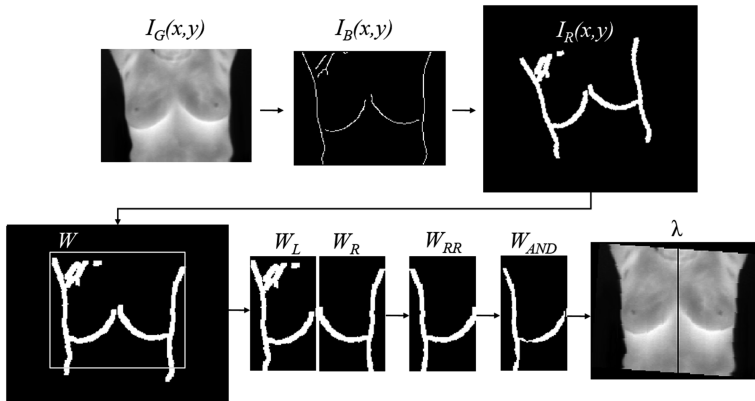


Fig. 2. Symmetry analysis and tilt error correction.

2.2 Breast Location Approximation

In order to approximate the location of the breasts on the tilt corrected thermographic image, the Hough transform is used. Because breasts tend to be circular/oval shaped,

one circle C is located over W_{AND} with the Hough transform. In order to detect the desired circles, it is important to constrain the identification of circles to possible expected breast sizes. Also, the position of circles must be limited by providing a minimum distance of the circle's origin to the upper image limit to avoid detection of false circles on neck/armpit. Due that the desired threshold value for the Canny edge detector, C_{TH} , may vary for different thermographies, the algorithm has an automatic threshold adjustment to ensure a desired amount of detected edges. In order to perform this adjustment, a threshold value for minimum symmetry level is defined experimentally. C_{TH} is initialized at a value permitting detection of few edges, and it is continuously reduced until reaching the desired symmetry level threshold in W_{AND} . This process is exemplified in Fig. 3. In (a) the threshold C_{TH} is adjusted until the y component of the circle C , y_C , is greater or equal to 37 (defined threshold for this size of thermography) making sure that the circle is not in the armpit. In (b) the value of C_{TH} is also continuously decremented until the symmetry level, L_S , is greater than 600 (defined threshold for this size of thermography) allowing a correct C circle detection over the breast on W_{AND} .

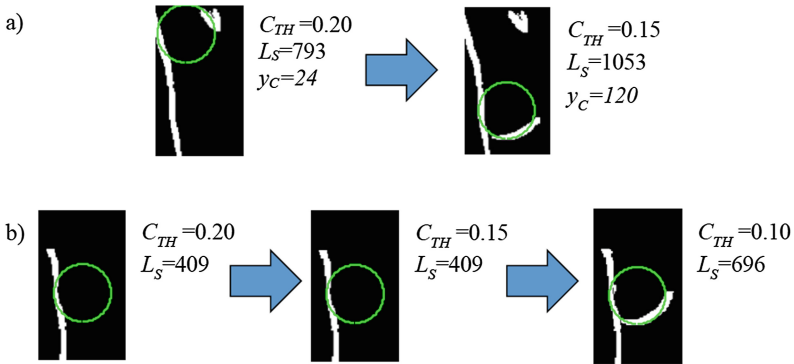


Fig. 3. Automatic threshold adjustment to make sure that (a) the circle C is not in the armpit, and that (b) the symmetry level L_S is above the desired threshold.

With the origin coordinates (x_C, y_C) of the circle C , and its radius r_C , it is easy to generate an image with circles located over the breasts positions by drawing 2 circles of radius r_C at the coordinates (x_{C1}, y_C) and (x_{C2}, y_C) where

$$x_{C1}, x_{C2} = \lambda \pm (\omega_{AND} - x_C) \quad (6)$$

The following condition is used to draw the circles in an image denominated $I_P(x, y)$

$$\text{if } \sqrt{(x_{C1} - x_p)^2 + (y_C - y_p)^2} \approx r_C \vee \sqrt{(x_{C2} - x_p)^2 + (y_C - y_p)^2} \approx r_C \quad (7)$$

$$I_P(x_p, y_p) = 1$$

evaluating every pixel $I_P(x_p, y_p) \in I_P(x, y)$.

The image $I_P(x,y)$ is then added to the border image of the rotated thermography obtained with the automatically adjusted threshold C_{TH} . The resulting image $I_{BM}(x,y)$ shows the edges of the body and two circles over the breasts. Since false edges can appear below the breasts, all the pixels below the circles (plus a small tolerance) are cleared. If there is a camouflage problem that does not allow the correct detection of the lower bounds of the breasts, the found circles provide a closed area and a good approximation of the breasts, considering that the circle C was placed correctly on the W_{AND} image. Figure 4 illustrates these steps.

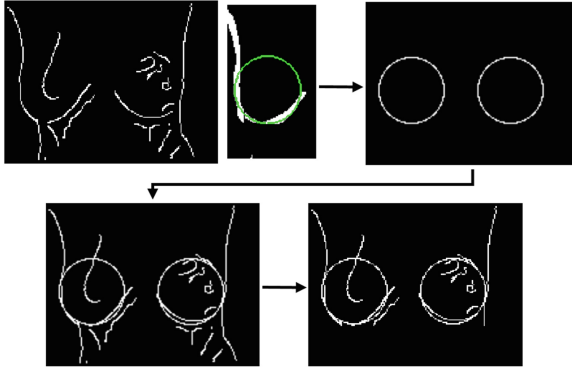


Fig. 4. Approximation of the location of the ROI through the borders image and circles' location over the breasts.

To obtain an approximation of the ROI area based on $I_{BM}(x,y)$, the next group of steps are needed. First, W_R and W_L are obtained from $I_{BM}(x,y)$ by placing W on the conditions of higher symmetry, and dilating the borders as previously explained. The dilation allows connection of separated borders. Then, a process of search and elimination of small areas takes place, removing small borders caused by noise that do not belong to the ROI. In order to provide closed ROI, through vertical and horizontal scans, the two closing pixels (endings) on the edges of every breast are detected. Having the coordinates of these points, it is possible to close the breast borders by drawing straight lines. Afterwards, the closed regions in W_R and W_L with greater areas are filled. If it is desired to obtain two separate ROI for each breast (recommended), the pixels in the last column of W_L are set to zero in order to create a separation line between the breasts. Since the generated areas are thickened, an erosion procedure is performed, allowing having borders closer to where initially the borders were detected. Additionally, with this process, edges that are not part of the ROI are deleted. Next, these images are placed on a blank image of the same size as $I_R(x,y)$, in the same positions where W_R and W_L were initially extracted. The resulting image $I_E(x,y)$ is a mask containing the ROI. This group of steps is illustrated on Fig. 5.

It is very possible that the ROI generated in $I_E(x,y)$ have an error of shape where the outer body edges may be connected to the lower edges of the breasts due to the strength

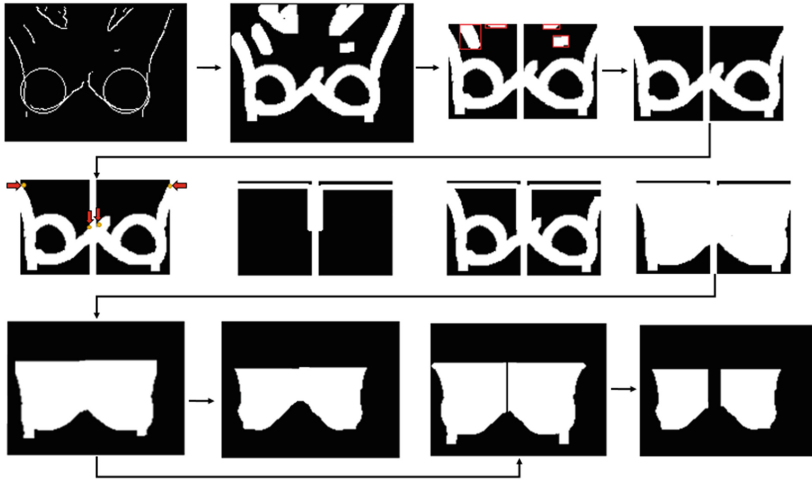


Fig. 5. Extraction of ROI through approximation of ROI.

of the outer body edges. In order to reduce this error, as shown in Fig. 6, a function to eliminate the undesired regions given by the following code:

for all white pixels $m(i, j)$ in $I_E(x, y)$

if $m(i, j - 5) = 0$ AND $m(i, j + 5) = 0$ THEN $m(i, j - 4) = 0$; $m(i, j - 3) = 0$; ...
 $m(i, j + 4) = 0$;

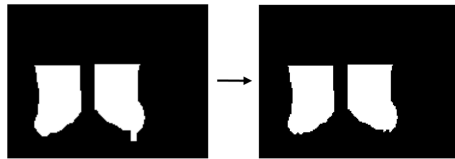


Fig. 6. Elimination of undesired regions in $I_E(x, y)$.

In order to generate a single ROI, as the ground truths provided in [7], an interpolation is needed. For this task, the coordinates of the lower bounds of the breasts in the ROI are obtained through a vertical scan. Additionally, the edge pixels in $I_{BM}(x, y)$ between the two extracted regions are considered and filtered, given that the interpolation will consider only coordinates from $I_{BM}(x, y)$ forming a monotonically incremental function before the symmetry axis, and a monotonically decremental function after the symmetry axis. The interpolation algorithm is cubic spline ($k = 3$), as it provides an excellent fit to the tabulated points and is not unduly complex. Subsequently, $I_E(x, y)$ is filled with pixels with a value of 1 in the middle region from the upper bound to the interpolated function generating a single ROI.

2.3 Breast Area Determination by Active Contours

At this stage, the performance of the algorithm is satisfactory. However, the accuracy can be improved by using active contours or snakes. This technique, shown in Fig. 7, is used to match a deformable model to an image by minimizing the energy which corresponds to the sum of the internal and external energies. For our application, the current ROI is the a priori information used as the initial deformable model. The active contour method used was reformulated to consider local image statistics rather than global, allowing segmentation of the ROI with heterogeneous characteristics that would be hard to extract using standard methods [12]. Without the given initial deformable model with the shape of the breasts, active contours by itself would provide accurate ROI. Rather it would amorphously adjust to the detected edges as it happened in the algorithm presented in [13] which also uses active contours.

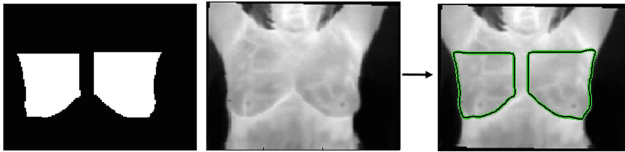


Fig. 7. Active contours adjust the borders of the ROI.

Since the edges of the ROI obtained from the active contour technique can be somewhat irregular, a contour softening is performed as the final process of the algorithm.

3 Discussion and Experimental Results

The algorithm was tested with 100 images (not used for the development of the algorithm) from the data base reported in [7] in order to evaluate the proposed method. It was observed that the algorithm had an excellent performance for the symmetry analysis, correcting tilt errors and providing accurate symmetry axes.

The approximation of the breasts was also very precise since in 100 % of the processed images, the circles were placed correctly over the breasts.

Given the ground truths from [7], the proposed method was compared to the algorithm of Marques [8] since he uses the same database and was found as the algorithm with best performance from all the methods found in the literature. The evaluation showed that the proposed algorithm had an accuracy and sensibility of 0.987 and 0.984, respectively, while Marques' algorithm showed an accuracy of 0.982 and sensibility of 0.974. Additionally, the proposed method had lower standard deviations in accuracy (0.007) and sensibility (0.017) compared to Marques (0.027 for accuracy and 0.066 for sensibility), showing that the suggested method is more reliable. Results also indicated that the proposed algorithm had specificity, positive and negative predictivity values of 0.988, 0.979 and 0.991, respectively.

Due to the characteristics of the proposed method, the algorithm detects breasts of all sizes and of all asymmetry conditions, even when there is a camouflage problem. Figure 8 shows the final extracted ROI for several thermographies. In its first image (upper left), the edges of the patient's left breast are not very visible (as in many other thermographies), and the algorithm presented in [8] failed to incorporate this breast into the ROI. However, the proposed method was able to estimate where this breast was more likely to be located, and included it into the ROI.

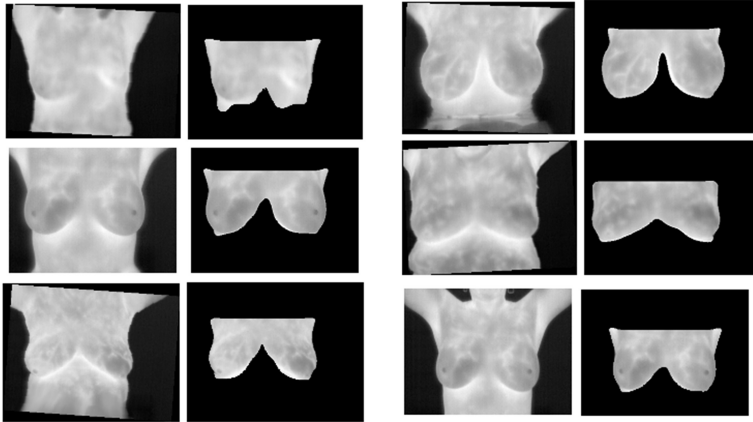


Fig. 8. Tilt corrections and extracted ROI for different breasts thermographies.

4 Conclusions

A novel algorithm for breast segmentation in digital thermographic images, with potential to be used in a CAD system, was presented.

The developed technique is able to correct the tilt error, providing a symmetry axis and detecting the breasts' positions even when their borders are not very visible. Results showed that the algorithm had an excellent performance, providing ROI with higher precision, accuracy, specificity and predictivity metrics compared to another algorithm found in the literature performing the same task.

Acknowledgements. The authors thanks to TNM, for the support of this research. Dr. Aura Conci and the personal of the Visual Lab at the Federal Fluminense University are also acknowledged for providing the used data base. Without their support, this research work would not have been possible.

References

1. World Health Organization: Cancer: fact sheet N°297 (2014). <http://www.who.int/mediacentre/factsheets/fs297/en/>

2. Drosu, O., Hantila, F., Maricar, M.: Non-invasive method for screening and early detection of breast tumors using thermal field analysis. *J. Electr. Electron.* **2**(2), 21–25 (2009)
3. Yokoe, T., Yuichi, I., Yoshiki, T., Hidetada, A., Noritaka, S., Tohru, K., Susumu, O., Yasuo, M.: Breast cancer only detected by thermography: a case report. *Ann. Cancer Res. Ther. Jpn. Soc. Strat. Cancer Res. Ther.* **4**(2), 89–90 (1995)
4. Hobbins, W., Amalu, W.: Beating breast cancer. *Original Internist* **16**(1), 46 (2009)
5. Ng, E.Y.K., Kee, E.C.: Advanced integrated technique in breast cancer thermography. *J. Med. Eng. Technol. Singap.* **32**(2), 103–114 (2008)
6. Borchardt, T.B., Concia, A., Limab, R.C.F., Resminia, R., Sanchez, A.: Breast thermography from an image processing viewpoint: a survey. *Sig. Process.* **93**(10), 2785–2803 (2013)
7. Silva, L.F., Saade, D.C.M., Sequeiros-Olivera, G.O., Silva, A.C., Paiva, A.C., Bravo, R.S., Conci, A.: A new database for breast research with infrared image. *J. Med. Imaging Health Inf.* **4**(1), 92–100 (2014)
8. Marques, R.S.: Segmentação automática das mamas em imagens térmicas. M.S. thesis, Instituto de Computação, Universidade Federal Fluminense, Niterói, RJ, Brasil (2012)
9. Borchardt, T.B.: Análise de imagens termográficas para a classificação de alterações na mama. Ph.D. thesis, Universidade Federal Fluminense, Icaraí, Brasil (2013)
10. Ballard, D., Brown, C.: *Computer Vision*. Prentice-Hall, Englewood Cliffs (1982)
11. Chacon-Murguía, M.I., Villalobos-Montiel, A.J., Calderon-Contreras, J.D.: Thermal image processing for breast symmetry detection oriented to automatic breast cancer analysis. In: Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A., Olvera-Lopez, J.A., Salas-Rodríguez, J., Suen, C.Y. (eds.) *MCPR 2014. LNCS*, vol. 8495, pp. 271–280. Springer, Heidelberg (2014)
12. Lankton, S., Tannenbaum, A.: Localizing region-based active contours. *IEEE Trans. Image Process.* **17**(11), 2029–2039 (2008)
13. Jaimes, R., Osorio, J.E.: Segmentación de Imágenes Termográficas de Glándulas Mamarias Utilizando Técnicas de Detección de Discontinuidades. Bachelor thesis. Universidad Industrial de Santander, Santander, Spain (2008)