Proceedings of the 26th Annual International Conference of the IEEE EMBS
San Francisco, CA, USA • September 1-5, 2004

Automated Image Segmentation for Breast Analysis Using Infrared Images

N. Scales[1], C. Herry[2], M. Frize[2,3]
[1] Department of Electronics, Carleton University, Ottawa, Canada
[2] Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada
[3] School of Information Technology and Engineering, University of Ottawa, Ottawa, Canada

Abstract—**In order to realize a fully automated thermogram analysis package for breast cancer detection, it is necessary to identify the region of interest in the thermal image prior to analysis. A nearly fully automated approach is outlined that is able to successfully locate the breast regions in most of the images analyzed. The approach consists of a sequence of Canny edge detectors to determine the body boundaries and to isolate the most likely candidates for the bottom breast boundary. Three different strategies for identifying the bottom breast boundary are investigated: a variation of the Hough transform to identify the curved edges in the image, an algorithm used to detect the longest connected edges that are not part of the body boundary, and a third approach involving the density of detected edges in the breast region. The last two methods show great promise in successfully segmenting the breasts.**

Keywords— **Thermography, breast cancer, automation, computer-assisted image processing**

I. Introduction

The use of thermography in applications such as detecting the presence of cancer in breasts or in assessing pain in patients is not new [1]-[4]. However, in spite of the enthusiasm of its early proponents, difficulties in differentiating between different levels of gray simply through the human eye made diagnoses based upon the detection of small differences in symmetry between the right and left breast extremely challenging. With advances in computer technology and the power available in current PCs, performing digital processing of these thermal images to detect subtle differences in symmetry for medical diagnosis became an exciting prospect.

Early researchers such as Gautherie et al. [2] believed that thermography would have the potential to detect early changes in blood flow, shown as an asymmetry between the two breasts. Moreover, thermography can be used to screen denser breasts which are typical in young women, whereas mammography is usually not effective until women reach the age of about 40. Thermographic imaging has been shown to offer prognostic and diagnostic information not available in other methods [5]. Unfortunately, it has not become widely used in the medical community due to the highly subjective nature of thermogram interpretation. Researchers are currently attempting to obtain quantifiable measures to help thermogram interpretation become more objective [5]-[7].

In recent studies involving the analysis of thermal images for breast cancer detection by Head and Lipari et al, [5][8][9], the breasts were isolated manually prior to analysis. An automated approach has been attempted by the same research group [10][11], that employs Canny edge detection and the Hough transform (HT) to detect the bottom breast boundaries. However, few details are provided, and the success rate of their method is not discussed. Two different methods for detecting the bottom breast boundary are discussed in detail in this paper, and a third possibility is outlined.

II. Methodology

In this study, 21 different infrared 128x128 8bit grayscale images have been analyzed. The algorithm used to identify the region of interest is in many ways similar to that described by Qi et al [10][11], but has been expanded upon in much greater detail. The method outlined here includes the following steps:

1. Manual removal of the top body edges (for example, the shoulders) and the waistline
2. Edge detection via a series of canny edge detectors
3. Isolation of the left, right, and topmost remaining body edges
4. Detection of the lower breast boundaries (either with a connected-edge detector or the HT)
5. 2^{nd} order interpolation of the curves found in step 4
6. Detection of the region with highest curvature in the left and right body edges
7. Estimation of the top breast boundary using a number of empirical rules and body regions
8. Isolation of the region of interest via the boundaries found in steps 3, 6, and 7

Step 1 is the only portion of the algorithm that has not been automated so far. Automation would be greatly facilitated if all images consistently included the waistline or shoulders. Success of the method typically depends upon steps 2 and 7 (and step 4 in the case of the HT).

The steps of the algorithm are illustrated in Fig. 1 to 3. The original image is shown in Fig. 1. Fig. 2 shows the same image after edge detection and prior to the last step. The isolated right and left breast regions are shown in Fig. 3. In this case, the connected edge algorithm was used to detect the bottom breast boundary.

A. Edge detection

Edge detection was implemented using the Canny edge detector, primarily due to its robustness to noise and its equal treatment of false positives and false negatives [12]. Canny's method implements the first derivative of a

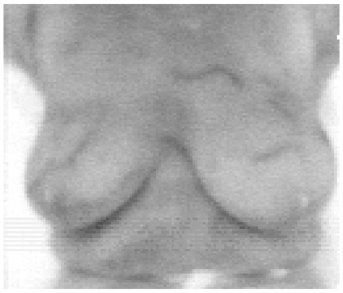

Fig. 1 – Original Image

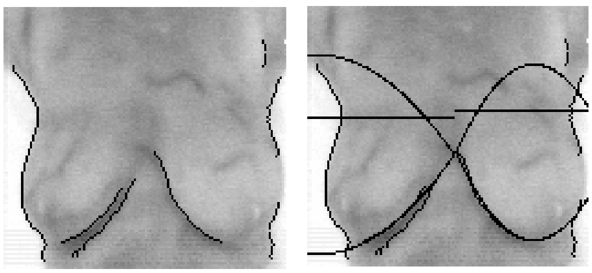

Fig. 2 – Edge Detection Output (left); Interpolation of Curves and Estimation of the Top Boundary (right)

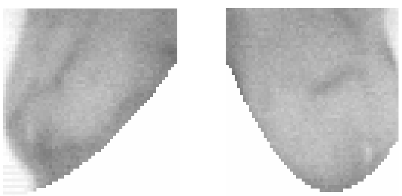

Fig. 3 – Isolation of the left and right breast

Gaussian function to smooth the image and to obtain the magnitude and orientation of the gradient for each pixel [13]. The built-in Canny edge detector in the Matlab Image Processing Toolbox was used, requiring only values for the low and high thresholds as well as the standard deviation of the Gaussian filter. After numerous trials, it was found that most images required different low/high threshold values in order to eliminate all but the strongest edges in the image without removing the bottom breast boundary.

This requirement suggests that the use of an adaptive filter to find the optimum threshold values for each image would be beneficial. However, for simplicity, in this work a series of different thresholds were applied to each image and an empirically determined success indicator was used to select the appropriate edge detector thresholds. It was found that, for the 21 images used in this study, an appropriate filter usually corresponded to one in which the total number of edge pixels in the image after edge detection was between 180 and 290. Thus, the success indicator chosen was the total number of edge pixels in the image after edge detection. A list of approximately a dozen of the most common threshold pair values and standard deviations was generated. The filters were applied so that in general each successive filter would have a more significant impact on the image, i.e. the high threshold for each successive filter was larger than the previous, with the difference between the low and high thresholds increasing as well. Although this algorithm is without any intelligent adaptive capabilities, it was found to perform adequately in most cases.

B. Detection of breast boundaries

The isolation of the outermost body edges was performed simply by finding the edge pixels at the leftmost, rightmost, and topmost of the image. In many cases, if there were no shoulder or neck edges present in the image, the topmost body edge alone could determine the lower breast boundaries. However, as many patients had thermal artifacts in the breast regions, a more robust method was required.

The detection of the lower breast boundaries was originally performed using a variation of the HT by Tsai et al. [12] designed to detect circles. This method takes advantage of the fact that lines normal to the tangent of a circle at any point will cross through the circle centre. The most frequent intersection of normals in the image (in a 2-D accumulator array) therefore determined the most likely candidate for the circle centre. A 1-D accumulator array was then used to find the radius of the circle.

The image was divided into two smaller images (left and right), with the leftmost and rightmost body edges removed. The HT was then applied to each of the smaller images. Typically, a number of candidate circles were detected for each breast boundary, as shown in Fig. 4. The portions of the circles not part of a detected edge were removed. The remaining circle edges were next either passed through an algorithm to detect the largest connected edge, or interpolated directly to find the most likely lower breast boundary candidate.

Unfortunately, the HT proved to be too sensitive to small changes in curvature or concavity, and thus many false edges were detected. The HT algorithm used was quite good at detecting circles, but in cases where the bottom breast boundary was relatively flat (see Fig. 4), the method failed. A more general version of the HT could have more success but at a higher computational cost.

A simpler and more successful strategy was to look for the largest connected edges in the image (after the left and rightmost body edges were removed). A recursive search algorithm was used to locate pixels that were 'connected' to their neighbors by a distance of no more than 2 pixels. The largest connected object in each half of the image was considered the most likely candidate for the bottom breast boundary. This algorithm could successfully detect the bottom breast boundary in a majority of the cases. The image shown in Fig. 4, that failed due to the flat nature of the bottom breast boundary when using the HT, was successfully analyzed using the connected edge technique, and is shown in Fig. 5.

In both techniques, a 2^{nd} order polynomial was used to approximate the bottom breast boundaries. It was found that higher order polynomials tended to change concavity in order to follow small fluctuations in the boundary, which made extrapolation beyond the detected boundary edge impossible.

The next step is to determine the top of the breast. Unfortunately, this region does not usually exhibit any edges in the image, and thus must be estimated from other points in the image, such as the armpit. As suggested by Qi et al. [10][11], the region with the highest curvature in the left and right body edges was used to estimate the position of the armpit. A cubic polynomial function was used to model the local curve around any given point. The point of maximum concavity was assumed to be found at the location of the armpit. Unfortunately, in many instances, poor edge detection resulted in strong points of concavity elsewhere in the body edge, and the point of maximum concavity was often found lower or higher in the image then expected.

An empirical measure was derived to successfully detect the top of the breast, using a number of different considerations:

1. The position of the armpit, located by the point of greatest concavity in the body edge
2. The topmost and lowermost points of the interpolated lower breast boundary curve
3. The median of the interpolated lower breast boundary curve
4. A reflected image of the bottom breast boundary
5. A maximum boundary of 1/3 the image size, measured from the top of the image
6. A minimum boundary of 25 pixels above the top of the bottom boundary

Although this method worked for many of the images, in others it tended to underestimate or overestimate the height of the top breast boundary. Improvements in edge detection, and more consistency in selecting the overall region to be imaged in each patient are expected to help in this problem. Some of the curves generated to help determine the top boundary are shown in Fig. 2 and 5. The boundary between the left and right breasts was simply taken to be the midpoint between the leftmost and rightmost boundaries. Once all the boundaries had been detected, the left and right breasts could be isolated for further analysis.

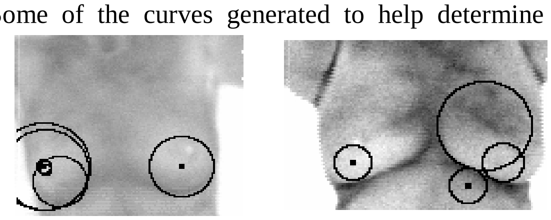

Fig. 4 – Circle detection using multiple circles to find the lower breast boundaries (left); Circle detection failure (right)

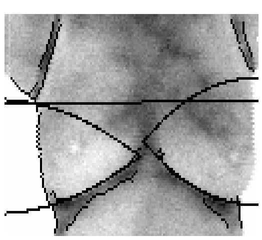

Fig. 5 – Successful use of connected edges to find the lower breast boundary. The edge detector output, the reflected bottom breast boundary, and the final top breast boundaries are shown.

III. Results and Discussion

The methods discussed above were used to isolate the breast regions in 21 different grayscale images. The breast regions were considered to be successfully isolated when visual inspection showed minimal clipping of the breasts. The HT method was able to successfully identify the regions of interest in only 4 of the 21 images. The failures were attributed to the detection of the bottom boundary in 12 of these cases, while the remainder of the failures was attributed to the edge detection step.

In the case of the connected-edge method, the regions of interest were successfully isolated in 13 of the 21 images. However, two more images could be analyzed successfully if the edge detection parameters were inputted manually. An additional image that had a gap in the detected edge for the bottom breast boundary could be successfully analyzed if the neighborhood size used to determine which pixels were connected was increased to 6. Thus, 16 of the 21 images could be isolated with only a slight variation of the method outlined above. The remaining 5 images, however, did not possess strong edge lines for the breast regions at all (see Fig. 6), or possessed strong thermal edges in or near the breast regions that were not related to physical boundaries of the body (see Fig. 7). Thus, although an adaptive, intelligent filter could definitely improve the success rate, alternative methods will have to be used in cases where edge boundaries are not detected at all, regardless of the filter used to detect the edges.

Some success in addressing these issues has been achieved using another method currently being developed by our group. The image contrast was first enhanced by using a disk of 5 pixel radius to perform top-hat and bottom-hat morphological filtering operations, thereby emphasizing the contours and the edges in the images. The threshold for the Canny edge detector was determined adaptively by fixing a lower threshold on the density of the edges detected, i.e. the ratio of the pixels belonging to edges to the total number of pixels in the region. All outermost body edges were neglected in the edge detection step, and detected edges were further trimmed of all edge contours corresponding to vertical or horizontal lines since they were likely artifacts from the edge detection or from the image itself and did not relate to intuitive breast features.

It was hypothesized that the center of mass of all the edges was likely located within the breast area; thus, edges closer to this center of mass were considered more likely to belong to the breast area. It was found that using an inclusion factor (which determined the cutoff distance from the center of mass) of 0.8 to 1.8 times the standard deviation associated with the position of edge pixels helped remove edges that were likely to be outside the breast region. This factor also proved to be a good estimate of the size of the breast (larger breasts in general required larger coefficients). Finally, the smallest convex region enclosing the remaining edges was sought and the ellipse that best fitted its contour was chosen as the most likely breast area. The ellipse fitting algorithm was taken from Halir and Flusser's paper [14]. The same procedure was repeated for both sides.

This method was found to be less sensitive to the number of successfully detected edges, as can be seen in Fig. 6, which shows the best ellipses delimiting the breasts in an image for which no lower boundary could be identified using the connected edges method. The correct selection of the inclusion factor was crucial for the proper delimitation of the breasts.

V. Conclusion

An automated method for identifying the breast regions in thermal images was devised and shown to be capable of correctly identifying the breast regions in approximately 75% of the images analyzed. After an initial stage that involved removing waistlines and shoulders from the image, a series of Canny edge detectors was applied to remove unwanted thermal edges from the image. Two different techniques were used to identify the lower breast boundary; the technique based upon the detection of the largest connected edges was far more successful than that based on the HT. The upper breast boundary was determined by a number of different measures based upon the left and right body edges as well as the lower breast boundary. All failures could be attributed to the edge detection step.

Future research should focus on the use of more intelligent edge detection as well as the identification of regions that do not show edges in the image, such as the upper breast boundary, and in approximately 25 % of cases, the lower breast boundary as well. Future work will also investigate more robust and precise methods to find the appropriate inclusion factor to determine the breast regions, such as integrating knowledge from other techniques (snake models, connected edges, region growing, and breast boundaries), which could provide a very efficient and robust tool for the segmentation of breast areas.

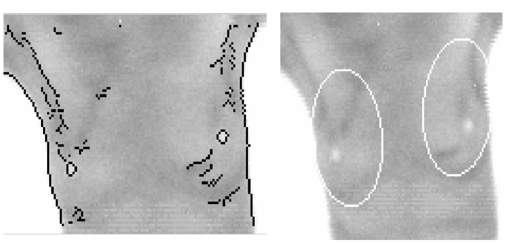

Fig. 6 – Example of an image with very few lines with which to locate the lower breast boundary (left); Best fitted ellipses (right).

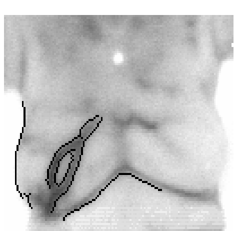

Fig. 7 – Large thermal lines in the image prevent successful detection of the bottom breast boundary

Acknowledgment

The authors thank The Terry Fox Foundation in Fredericton for funding the 1984 study, Dr. Roberge, at Moncton Hospital, for providing the expertise, patient base, and clinical results for the breast cancer research. The authors also thank the National Science Research Council Canada (NSERC) for funding the present study.

References

[1] J. Head, C. Lipari and R. Elliott, “Comparison of mammography and breast infrared imaging: sensitivity, specificity, false negatives, false positives, positive predictive value and negative predictive value,” *Proc. Joint BMES/EMBS Conf.*, Atlanta, 1999, p. 1116.

[2] M. Gautherie and C. Gros “Breast thermography and cancer risk prediction,” *Cancer*, vol. 45, pp. 51–61, 1980.

[3] M. Aubry-Frize, G. Quartey, H. Evans and D. Lapalme, “The thermographic detection of pain,” *Proc. 3rd Can Clin Eng Conf.*, pp. 82–83, Saskatoon, 1981.

[4] Y. Poussart, M. Frize and R. Roberge, “The re-evaluation of thermography in breast cancer detection by new image enhancement techniques”, Terry Fox Foundation, Fredericton, NB, Canada, Research Report, 1988.

[5] C. Lipari, J. Head, F. Wang and R. Elliot, “Image analysis of digitized infrared images of the breasts from a first generation infrared imaging system” *Proc 19th Int Conf IEEE/EMBS*, Chicago IL, 1997, pp. 681–684.

[6] M. Anbar, L. Milescu, A. Naumov, C. Brown, T. Button, C. Carty et al., “Detection of cancerous breasts by dynamic area telethermometry,” *IEEE Eng Med Biol Mag*, pp. 80–91, 2001.

[7] M. Frize, C. Herry and R. Roberge, “Image processing of thermal images to detect breast cancer,” *Proc 24th Int Conf IEEE/EMBS*, Houston, TX, 2002.

[8] J. Head, F. Wang, C. Lipari and R. Elliot, “The important role of infrared imaging in breast cancer,” *IEEE Eng Med Biol Mag*, pp. 52–57, May/June 2000.

[9] C. Lipari and J. Head, “Advanced infrared image processing for breast cancer risk assessment,” *Proc 19th Int Conf IEEE/EMBS*, Chicago IL, 1997, pp. 673–676.

[10] J. Qi, W. Snyder, J. Head and R. Elliot, “Detecting breast cancer from infrared images by asymmetry analysis,” *Proc 22nd Annual EMBS Int Conf.*, Chicago IL, 2000, pp. 1227–1228.

[11] J. Qi and J. Head, “Asymmetry analysis using automatic segmentation and classification for breast cancer detection in thermograms,” *IEEE EMBS Conf. Turkey*, 2001.

[12] D. Tsai, “A machine vision approach to detecting and inspecting circular parts,” *Int J Adv Manuf. Technol.*, 1999: 15: 217–221.

[13] A. Bovik, “Handbook of Image and Video Processing”, Academic Press, 2000.

[14] R. Halir, J. Flusser, “Numerically Stable Direct Least Squares Fitting of Ellipses”, *WSCG’98*, pp. 125-132, 1998.