

Automated identification of the orientation of mediolateral oblique mammographic images

C.D.M. Henderson

<http://www.craighenderson.co.uk>

cdm.henderson@gmail.com

Abstract. *When comparing mammogram images, whether temporal or bilateral, it is essential that the orientation of the breasts in each image match. This initial image orientation problem is generally omitted from the registration procedure and left to the operator to provide the x-ray images in the order and orientation that the software system requires. In a fully automated system, this is an unacceptable burden on the system operator. This paper describes an algorithmic procedure for identifying the orientation of mediolateral oblique breast images that can be used as a first step in the process of mammogram registration in an automated system.*

Keywords. mammogram orientation; image processing; image analysis

Introduction

The algorithm described in this paper is simple but effective for identifying the orientation of a mammogram image. Since the purpose of the paper is to describe a method of identifying left and right oriented mediolateral oblique (MLO) images, the terms nipple-left and nipple-right will be used with respect to an image, without any implication whether it is the left or the right breast of the patient. This specific laterality is insignificant information to the registration procedure, or a computer-aided detection (CAD) system as a whole, but is important information to report in the results of the analysis.

Image Segmentation

It is desirable to be able to identify the orientation of the image before precise segmentation of the mammogram image to reduce the complexity and increase the efficiency of the segmentation algorithm [1]. An algorithm to segment an MLO image with a known orientation can be confident in making assumptions regarding, for example, the curvature of the breast-air region, and can mirror an image of the wrong orientation in a simple pre-processing step, therefore increasing flexibility of its use.

However, to reliably analyse the image for the orientation, it is necessary to remove non-breast areas such as film clips and as much noise and as many artefacts as possible. To satisfy this need, a very coarse segmentation algorithm is applied to the image to create an estimation of the breast area. The most important attribute to be identified is the overall shape of the breast area. The size and border positions are of no significance for

identifying the orientation of the breast. A simple binary thresholding algorithm is applied to the image to create a black and white image. In a recent paper, Papadopoulos [2] suggested a global threshold value of 20 could be applied to a normalised 256 greyscale mammogram image for segmentation purposes. Most values below this threshold, Papadopoulos argues, are representative of the background area, with a small number within the tissue area close to the breast-air boundary. This has proven to work well for locating breast region, although not an accurate technique for finding the boundaries. The largest non-black area is then extracted from the binary image and assumed to represent the breast area [1].

Orientation Analysis

The procedure of identifying the orientation of a mammogram is described below.

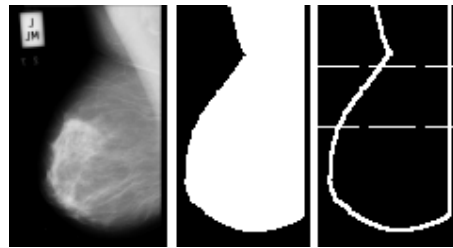


Figure 1 – Breast orientation steps

Figure 1 shows the output of the image processing in steps 1 to 3 applied to image mdb095 from the MIAS MiniMammography Database [3].

1. Threshold the image, setting each pixel to black if the grey scale intensity is less than 20, otherwise the pixel is set to white. Extract the largest white area.
2. Create an outline of the breast representation by applying a Sobel Edge Detector [4] to the binary image.

Using the convolution masks

P ₁	P ₂	P ₃
P ₄	P ₅	P ₆
P ₇	P ₈	P ₉

-1	0	+1
-2	0	+2
-1	0	+1

+1	+2	+1
0	0	0
-1	-2	-1

Convolution Mask for Sobel Edge Detector

the approximate gradient magnitude at each pixel can be calculated quickly with the equation

$$|P_5| = |(P_1 + 2 * P_2 + P_3) - (P_7 + 2 * P_8 + P_9)| + |(P_3 + 2 * P_6 + P_9) - (P_1 + 2 * P_4 + P_7)|$$

3. Divide the image into four horizontal rows of equal width and mask off the second quarter from the top, such that the upper quarter and lower half of the image is ignored in subsequent processing.
4. Use a 3 by 3 mask (below) to inspect the diagonal neighbours of each pixel in the image. Using a variable, *gradient*, with an initial value of zero, increment *gradient* for non-black values in the top-left and bottom-right corner and decrement the Gradient for non-black values in the top-right and bottom-left corner.

+1	0	-1
0	0	0
-1	0	+1

Mask to inspect diagonal neighbours

5. The sign of the gradient determines the orientation of the breast image after step 4 has been applied to the entire image. A negative value identifies a nipple-left image; a positive value identifies a nipple-right image. A value of zero indicates that the algorithm has failed to conclude the orientation of the image.

Initial Results

Initial test results in applying this algorithm to the individual images from the MIAS MiniMammography Database were reasonable, but not sufficiently reliable for a fully automated solution.

Out of the 322 bilateral mammogram images from 161 patients, the orientation of 316 images were correctly identified, 1 image resulted in a zero value gradient, 2 nipple-left images were marked as nipple-right and 3 nipple-right images were marked as nipple-left. Of these failures, contralateral images from a patient were identified as the inverse orientation in only one case. The images mdb287 and mdb288 are a bilateral pair of x-ray images from a single patient and were identified as nipple-left and nipple-right respectively whereas they are in fact nipple-right and nipple-left images. In all other failure cases, the failure was either confined to one image from a pair, or at least one of the failures was self-verifying (i.e. a gradient calculation of zero was produced).

These results represent a success rate of 98.14% in this small low-resolution sample. A failure rate of 1.86%

Failures

There were four categories of failure of the algorithm to identify the orientation of the images.

Cropped Breast Image

The MLO images mdb151 and mdb152 represent a bilateral pair of breast x-rays from one patient (Figure 2). The orientation algorithm calculated a gradient of zero for image mdb151 and incorrectly identifies mdb152 as a nipple-left MLO (with a gradient value of -3).

A gradient of zero always indicates failure, as an MLO image is always either nipple-left or nipple-right. This is the only situation where the algorithm can verify its own results for a single image.

Each breast is larger than the x-ray film, and the complete breast outline is therefore not represented. Significant parts of the breast are missing from the mammogram image, where the images are cropped on three sides, the top, the bottom and the nipple side.

The algorithm for breast orientation relies upon the breast outline to calculate the gradient. In these images, the outline is available only in part, so an adjustment could be made to the algorithm to cope with such images. However, test images of breasts of this scale are less readily available, and this is the only example in the test set used in this research.

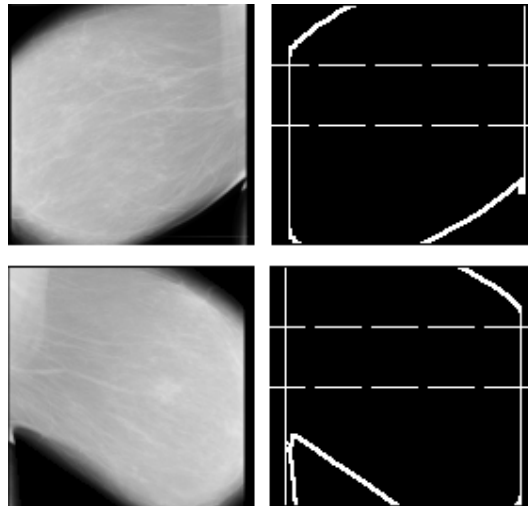


Figure 2 - mdb151 (top) and mdb152 (bottom) - Cropped Breast Images

High Nipple Position

The image mdb179 (Figure 3) represents a nipple-left MLO breast x-ray. The orientation algorithm incorrectly identifies this image as a nipple-right (with a gradient value of 97).

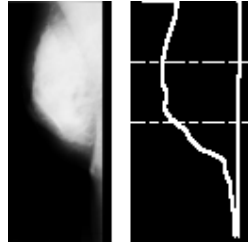


Figure 3 - mdb179 – High Nipple Position

On first inspection, there is no immediate reason why the algorithm should fail on this image as the Sobel Filtered image shows a clear and well-defined breast outline (Figure 3b). However, within the Top 2nd Quarter of the image that is actually used in the algorithm, it is the outline of the nipple that has distorted the gradient and therefore produced an incorrect result.

The result is incorrect because of specific characteristics of the breast outline in the x-ray image. It is easy to see that a correct result could be accomplished using the gradient algorithm on a different part of the image, rather than the Top 2nd Quarter. However, the Top 2nd Quarter has been identified during the research and testing as being a good general case partition for the image processing.

Noise artefacts near the breast contour

The image mdb280 represents a nipple-right MLO breast x-ray (Figure 4). The orientation algorithm incorrectly identifies this image as a nipple-left (with a gradient value of -3).

This is obviously a poor quality x-ray image and the amount of noise on the film will undoubtedly cause problems. The simple segmentation of this image has failed to extract the breast area because of the amount of noise on the image, but that is not the cause of the orientation failure.

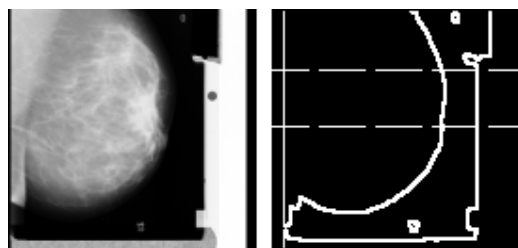


Figure 4 - mdb280 - Noise artefacts near the breast contour

The reason for failure is two fold, and largely coincidental. The Sobel Filtered image in Figure 4b shows noise interference along the breast outline. One of these features lies at the top of our region of interest in the Top 2nd

Quarter, and distorts the results. This distortion is very small, but is just enough to throw the algorithm to give a false result because the sample of breast outline in the Top 2nd Quarter is largely vertical. Even without the noise, the gradient calculation would be very close to zero; so one small piece of noise is enough to generate an incorrect result.

Small breasts have near-vertical contours

The MLO images mdb287 and mdb288 (Figure 5) represent a bilateral pair of breast x-rays from one patient. The orientation algorithm incorrectly identifies image mdb287 as a nipple-right MLO with a gradient calculation of 66 and image mdb288 as a nipple-left MLO with a gradient calculation of -88. The gradient in the top 2nd quarter region of the images is in fact the inverse of that expected from the breast outlines, and therefore the orientation algorithm is giving the correct result based on the outline segment given.

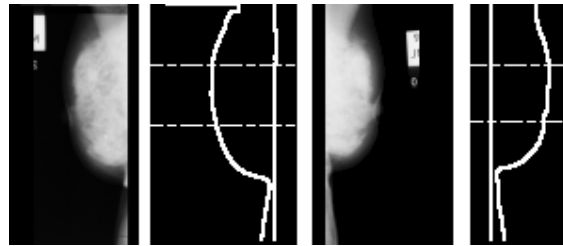


Figure 5 - mdb287 and mdb288 - Small breasts have near-vertical contours

This is a consequence of the shape of the breast outline of this particular patient.

Algorithm Refinements

The algorithm for breast orientation relies upon the breast outline to calculate the gradient. The top 2nd quarter of the image has been identified during research and testing to be a good general case partition for the image processing, producing a correct result in 98.14% of the test images.

The actual causes of failure of the algorithm described above are twofold. Firstly, the position of the vertical region of interest is fixed, and has been selected based on the research and testing. While this position is a good general case, it does not represent the best window of the breast outline contour for all images. Dynamically repositioning the region of interest after a failure will enable the algorithm to continuously reassess the contour until a satisfactory result is established. Secondly, the algorithm is unable to validate its result, unless the gradient is calculated to be zero. If the algorithm was performed on a bilateral pair of images simultaneously, the results from each image can be used to validate the result of both. A revised algorithm is presented below to process image pairs and cross-validate the results.

1. Steps 1 to 3 of the original algorithm are performed on each of the two images to create a partitioned breast outline image of each mammogram.
2. Steps 4 and 5 of the original algorithm are performed on the two images, yielding two gradient values.
3. The two gradient values are compared. If one value is less than zero and the other is greater than zero, then the algorithm has identified one nipple-left and one nipple-right image, so the processing stops.
4. If either gradient has been calculated to zero, then the image orientation is inconclusive in the processing region of interest. For each of the images with a zero value gradient, the region of interest is adjusted upward by half the previous distance from the top of the region of interest to the top of the breast outline. If neither gradient has been calculated to zero, then adjust the region of interest of both images by this amount.
5. Repeat the procedure from Step 2 until the region of interest of either image moves beyond the top of the breast outline.

Error Tolerance

Failure of an automatic algorithm to identify the orientation in a small number of cases is acceptable, and is to be expected when processing images of deformable subjects of such diversity as mammograms. However, the system must be completely reliable in its successes. A failure to deduce the orientation of a pair of images is preferred to reporting an incorrect result.

The self-validation of the revised algorithm presented here fails in one case out of the 161 test cases. In processing the bilateral pair mdb287 and mdb288 (Figure 5), both images were identified as the inverse orientation, i.e. nipple-right and nipple-left respectively whereas they are in fact nipple-left and nipple-right images. Here the algorithm verifies the result as correct, even though it is incorrect, as there is one image of each orientation, resulting in a false positive result.

To identify false positives, a second verification process is used, using prior knowledge of the general appearance of a mediolateral oblique mammogram images. In nipple-left images, the top left area is expected to contain less white pixels than the bottom right area and nipple-right images are expected to contain the greater quantity in the top left area. The segmented image from step 1 of the algorithm is bound to the tightest rectangle covering the breast region and divided into a grid of sixteen, four rows and four columns. The number of white pixels in the top left area of each image is then compared with the number in the bottom right area, which gives a second opinion

orientation. If neither image is identified in the same orientation using this technique as the result from the full algorithm, then the image must be reported as ambiguous.

Conclusion

The revised algorithm with added verification succeeds in correctly identifying the orientation of 160 of the 161 bilateral pairs in the test data set. The remaining bilateral pair is reported as ambiguous. The overall success rate in this small low-resolution sample is 99.38%. A failure to identify the orientation occurred in 0.61% of the test data, and the algorithm reported no false positives.

With a success rate in excess of 99% on the test data, this algorithm appears to be reliable in identifying the orientation of each image in a bilateral pair. Testing on a larger data set will be required to validate these early results for the algorithm's reliability in a fully automated system.

References

- [1] K. Woods, Automated Image Analysis Techniques for Digital Mammography. *Department of Computer Science and Engineering*, University of South Florida, Tampa, Florida, 1994.
- [2] A. Papadopoulos, D. Fotiadis, A. Likas, An automatic microcalcification detection system based on a hybrid neural network classifier, *Artificial Intelligence in Medicine*, 25 (2002) 149-167.
- [3] A. Clark, 2003, The Mammographic Image Analysis Society MiniMammography Database.
<http://www.wiau.man.ac.uk/services/MIAS/MIASmini.html>
- [4] R. Gonzalez and R. Woods, Digital Image Processing, second ed., Prentice Hall, New Jersey, 2002.