Use of Quadrature Filters for Detection of Stellate Lesions in Mammograms

Hans Bornefalk

Dep. of Physics, KTH, SE-106 91 Stockholm, Sweden
`bornefalk@particle.kth.se`

Abstract. We propose a method for finding stellate lesions in digitized mammograms based on the use of both local phase and local orientation information extracted from quadrature filter outputs. The local phase information allows efficient and fast separation between edges and lines and the local orientation estimates are used to find areas circumscribed by edges and with radiating lines. The method is incorporated in a computer-aided detection system and evaluation by FROC-curve analysis on a data set of 90 mammograms (45 pairs) yields a false positive rate of 0.3 per image at 90% sensitivity.

1 Introduction

Breast cancer is one of the leading causes of cancer deaths for women in the western world. Early detection with mammography is believed to reduce mortality [1], [2] and several countries, including Sweden, have nation-wide screening programs where middle-aged women are routinely invited to breast cancer screening. To reduce intra-reader variability and to increase overall efficacy several research groups have developed computer-aided methods that are intended to work as a “second look”, or spell checker, prompting the radiologist to take a second look at any suspicious areas. While there seems to be a need for more evaluation of the usefulness of computer-aided detection (CAD) systems in screening environments, the evidence is accumulating in favor of it increasing the overall efficacy of the screening process [3], [4]. With the advent of full-field digital mammography, which is expected to replace the nowadays commonplace film-screen mammography, the extra step of having to digitize the film mammogram will disappear and thus make computer-assisted methods even more attractive.

Depending on the type, carcinoma in the breast can manifest itself in several ways. Some are only revealed by the presence of microcalcifications, whereas other have a center mass which is visible as a denser region on the mammogram. This paper is concerned with the detection and classification of spiculated, or stellate, lesions. These are characterized by their branches which spread out in all different directions. Carcinoma of this subtype can either have or not have a well defined center body, and their detection, especially in dense breasts, is often hampered by superposition of normal anatomical structures of the breast. As their presence is a highly suspicious indicator of malignancy, especially combined

with a center mass, the detection of stellate patterns is a very important step in computer-aided detection of malignant lesions. It is therefore not surprising that so many different techniques for identifying these areas have been proposed.

Karssemeijer and te Brake [5] suggested a statistical method based on a map of pixel orientations. The idea is that if many pixels have orientations pointing toward a specific region, this region is more likely to be the center of a stellate distortion. The reported sensitivity was about 90% at a false positive rate of 1 per image. This method has later on been improved to 0.38 false positives per image at the same sensitivity [6]. Another popular method is based on first identifying individual spicules and then, via the Hough transform, accumulate evidence that they point in a certain direction [7], [8]. A third method for the detection of radiating spicules is based on histogram analysis of gradient angles [9]. The basic idea is that if the distribution of gradient angles in a certain local neighborhood has high standard deviation, this indicates that the gradients point in all different directions. This is indicative of a stellate pattern. On the other hand, if the distribution has a well defined peak, this indicates that the gradients in the neighborhood have a well defined orientation and thus that they do not belong to a stellate distortion. A common problem when detecting spiculated lesions is that they range in size from a few millimeters up to several centimeters. The solution adopted by Kegelmeyer [9] was simply to choose the size of the local neighborhood large enough to encompass even the largest lesions. However, a large local neighborhood makes detection of small spiculated lesions more difficult. Karssemeijer and te Brake addressed this issue by estimating the edge orientations at three different scales, and choosing the maximum response.

The objective of the study is to present an alternative algorithm for the detection of malignant tumors in digital mammograms. This method has been presented partially before by the author [10]. The method resembles all three methods mentioned above: pixel orientations are extracted [5], [6], the individual spicules are identified [7] and finally the angular distribution is analyzed in a way similar to Kegelmeyer's method [9]. To be able to evaluate the system, the whole CAD chain is implemented (finding regions-of-interest (ROIs), segmentation, extraction of image characteristics and classification) but the focus is on developing the ROI location technique using quadrature filter outputs.

2 Region of Interest Location

Several methods for finding seed points for ROIs exist. Most methods are intensity based using the fact that many tumors have a well defined central body. Other methods make use of spiculation features and look for regions where spicules seem to radiate from. We have chosen a combination of these two features, and also added a feature to capture the edge orientation. In order to minimize the risk of missing potential areas the entire image is uniformly sampled and each sampling point is evaluated in terms of contrast and spiculation. This evaluation is performed on three different scales. The contrast measure at node i, j is defined as the contrast between a circular neighborhood with radius r

centered at i, j and the washer-shaped neighborhood with inner and outer radii given by r and $2r$. The spiculation and edge measures are based on orientation estimates extracted from quadrature filter outputs, as described below.

2.1 Introduction to Quadrature Filters

A filter f is a quadrature filter if there is a vector $\hat{\mathbf{n}}$ such that the Fourier transform F of f is zero in one half of the Fourier domain: $F(\omega) = 0$, if $\omega^T \hat{\mathbf{n}} \leq 0$. $\hat{\mathbf{n}}$ is called the directing vector of the filter. Quadrature filters and a method to construct orientation tensors from the quadrature filter outputs are described thoroughly in Granlund and Knutsson [11] and here we will only touch briefly on some essential concepts.

The directing vector of quadrature filter i is denoted $\hat{\mathbf{n}}_i$ with $\varphi_i = \arg(\hat{\mathbf{n}}_i)$. The quadrature filters will be complex in the spatial domain since $F(\omega) \neq F^*(-\omega)$, and thus the convolution output with the image signal s , $\mathbf{q}_i = f_i \star s$, will be complex. We let q_i denote the magnitude of $\mathbf{q}_i$ and similar for the phase angle: $\theta_i = \arg(\mathbf{q}_i)$.

The local orientation in an image is the direction in which the signal exhibits maximal variation. This direction can only be determined modulo π which is the rationale for introducing the double angle representation below. If $\mathbf{v}$ is a vector oriented along the axis of maximal signal variation in a certain point, the double angle representation of the local orientation estimate is given by $\mathbf{z} = ce^{i2\varphi}$ where the magnitude c is the certainty of the estimate and $\varphi = \arg(\mathbf{v})$.

It can be shown [11] that the double angle representation of the local orientation can be constructed as: $\mathbf{z} = \sum_i q_i \hat{\mathbf{m}}_i$ where $\hat{\mathbf{m}}_i = (\cos(2\varphi_i), \sin(2\varphi_i))$. With four quadrature filters and $\varphi_i = (i-1)\pi/4$ this leads to a convenient expression for the 2D-orientation vector:

$$\mathbf{z} = (q_1 - q_3, q_2 - q_4). \quad (1)$$

In the spatial domain, a quadrature filter can be written as the sum of a real line detector and a real edge detector:

$$f(x) = f_{line}(x) - if_{edge}(x). \quad (2)$$

The line and edge detectors are related to each other the following way in the Fourier domain:

$$H_{edge}(u) = \begin{cases} -iH_{line}(u) & \text{if } u < 0 \\ iH_{line}(u) & \text{if } u \geq 0 \end{cases} \quad (3)$$

f_{line} is an even function and f_{edge} is an odd function and this can be used to distinguish between lines and edges. The phase angle introduced above actually reflects the relationship between the evenness and oddness of the signal. A line would give a non-zero response only for the real part of f (since f_{line} is even) and hence the phase angle $\theta = \arg(f \star s)$ would be close to zero. On the other hand, an edge would give a phase angle of $\pm\pi/2$ since only the odd part of Eq. 2 would give a contribution. This is illustrated in the right panel of Fig. 1 for the

Fig. 1. a) Line and edge filters used for construction of a 1D-quadrature filter. b) Lower row depicts filter responses for the signals in the top row

filter pair in the left panel. Note how the phase angle can be used to separate edges from lines.

The extension of the phase concept to two dimensions is not trivial, but will give us the necessary means to distinguish different events from each other (namely edges, bright lines and dark lines). The reason for the difficulties is that phase cannot be defined independently of direction, and as the directing vectors of the quadrature filters point in different directions (yielding opposite signs for similar events) care must be taken in the summation [12]. If $\Re(\mathbf{q}_i)$ and $\Im(\mathbf{q}_i)$ denote the real and imaginary parts of the filter output from the quadrature filter in direction $\hat{\mathbf{n}}_i$, the weighted filter output $\mathbf{q}$ is given by

$$\Re(\mathbf{q}) = \sum_{i=1}^4 \Re(\mathbf{q}_i), \quad \Im(\mathbf{q}) = \sum_{i=1}^4 I_i \Im(\mathbf{q}_i) \quad (4)$$

where

$$I_i = \begin{cases} 1 & \text{if } |\varphi_i - \varphi| \leq \frac{\pi}{2} \\ -1 & \text{if } |\varphi_i - \varphi| > \frac{\pi}{2} \end{cases} \quad (5)$$

The interpretation of the indicator function I_i is that when the local orientation in the image (φ) and the directing vector of the filter (φ_i) differ by more than $\pi/2$ the filter output $\mathbf{q}_i$ must be conjugated to account for the anti-symmetric imaginary part. The total phase θ is now given as $\theta = \arg(\mathbf{q}) = \arg(\Re(\mathbf{q}) + i\Im(\mathbf{q}))$. As for the one-dimensional case, phase angles close to zero correspond to bright lines, phase angles close to $\pm\pi$ correspond to dark lines and phase angles close to $\pm\pi/2$ correspond to edges.

2.2 Illustration on a Real Mammographic Lesion

By thresholding the filter outputs on certainty and phase, a binary image is produced. This can be used to separate bright lines, and thus candidates for spicules, from the surrounding tissue. Edges can also be extracted if the thresholds are chosen appropriately. This is illustrated in Fig. 2 where phase angles

Fig. 2. a) Malignant lesion with radiating spicules in a washer-shaped local neighborhood. Detected by thresholding on local phase. b) Malignant lesion and prominent edges as detected by local phase

$\theta \in (-\pi/6, \pi/6)$ are extracted in panel a) and $|\theta| \in (5\pi/6, 7\pi/6)$ in panel b). Note how the majority of lines in Fig. 2 seem to radiate away from the center of the lesion, and the edges seem to circumscribe the center. The question is how this can be quantified for subsequent use in classification?

2.3 Measure of Spiculatedness in a Local Neighborhood

Consider a coordinate $\mathbf{x}_0 = (x_0, y_0)$ in the image. Next consider a pixel coordinate $\mathbf{x} = (x, y)$ on a detected bright line. This is illustrated in panel a) in Fig. 3. The direction of maximal signal variation in this pixel is $\mathbf{v}(\mathbf{x})$. Let $\varphi(\mathbf{x}) = \arg(\mathbf{v}(\mathbf{x}))$. From Eq. (1) we get the following expression for the double angle representation of local orientation:

$$\mathbf{z}(\mathbf{x}) = ce^{i2\varphi(\mathbf{x})} = q_1 - q_3 + i(q_2 - q_4). \quad (6)$$

Let $\hat{\mathbf{r}}$ be the normalized vector pointing from $\mathbf{x}_0$ to the pixel $\mathbf{x}$: $\hat{\mathbf{r}}(\mathbf{x}) = \frac{\mathbf{x} - \mathbf{x}_0}{|\mathbf{x} - \mathbf{x}_0|}$. Since the vector $\hat{\mathbf{r}}$ is normalized, it can be expressed as $(\cos \varphi_r, \sin \varphi_r)$. Now define

$$\hat{\mathbf{r}}_{double}(\mathbf{x}) = (\cos(2\varphi_r), \sin(2\varphi_r)). \quad (7)$$

If $\mathbf{x}$ is located on a line radiating away from the center coordinate (in the same direction as $\hat{\mathbf{r}}(\mathbf{x})$), the angles between $\hat{\mathbf{r}}_{double}$ and $\mathbf{z}(\mathbf{x})$ will be π . On the other hand, if $\mathbf{x}$ is located on a line perpendicular to $\hat{\mathbf{r}}$, the angle between $\mathbf{z}$ and $\hat{\mathbf{r}}_{double}$ will be zero. To see that, consider panel a) in Fig. 3 where ψ denotes the angle between $\hat{\mathbf{r}}(\mathbf{x})$ and $\mathbf{v}(\mathbf{x})$. From the figure we see that $\arg(\mathbf{v}) = \varphi_r + \psi \pm \pi/2$. This means that

$$\arg(\mathbf{z}) = 2\varphi_r + 2\psi \pm \pi = 2\varphi_r + 2\psi + \pi \pmod{2\pi}. \quad (8)$$

Since $\arg(\hat{\mathbf{r}}_{double}) = 2\varphi_r$ the absolute value of the difference between the angles (modulo 2π) is

$$\phi = |\arg(\mathbf{z}) - \arg(\hat{\mathbf{r}}_{double}) \pmod{2\pi}| = |2\psi + \pi \pmod{2\pi}|. \quad (9)$$

Now, with ψ close to zero, as it would be if the line is part of a stellate pattern, the angle difference will be close to π , as proposed above. On the other hand, if the

Fig. 3. a) Illustration of relevant angles. b) Angle difference distribution ϕ for the thresholded bright lines in a local neighborhood of a spiculated lesion. Note that the distribution is clearly skewed to the right. c) Angle distribution for edge pixels. The leftward skew indicates that the edges are orthogonal to the radial direction

line is perpendicular to $\hat{r}$ the angle difference ϕ will be close to zero. Thus, if the distribution of the angle differences corresponding to the pixels identified in the line image in a local neighborhood is skewed towards π , this is an indication that many lines are radiating from the center. Conversely, if angle differences of the edge image are skewed towards the left, this is an indication that the prominent edges are perpendicular to lines radiating from the center. An illustration, using the actual data from Fig. 2, is shown in panels b) and c) of Fig. 3.

2.4 Training an ROI Extractor

Five features (or local image characteristics) are used in the ROI extractor. The first is the contrast as mentioned above. The second and the third are the fraction of points in the line image in the washer-shaped neighborhood that have angle deviations ϕ in the intervals $(\pi/2, \pi]$ and $(3\pi/4, \pi]$, respectively. The fourth and fifth features are similar measures for the edge image.

A freely available support vector machine (SVM) implementation [13] is trained to distinguish between areas that could potentially be malignant and those that could not. A washer-shaped neighborhood with inner radius matching the size of the radiologist supplied ground truth is used in extracting the features. Once the five features in the malignant area have been extracted, the same washer-shaped mask is used to extract the same five features in 100 randomly chosen neighborhoods in the image. These are the normal samples that will be used when training the SVM. To ensure independency between training and evaluation data, the leave-one-case-out technique is employed when the SVM is used to extract ROIs.

When features are extracted from a mammogram in the ROI extraction step, the size of the possible lesion is unknown and three different radii of the washer-shaped neighborhood are evaluated. The radius where the corresponding features give the highest SVM response is taken as the size of the ROI - much the same way as Karssemeijer and te Brake do [5], [6].

A typical intermediate result of the ROI extractor is shown in Fig. 4. In panel a) the output from the support vector machine is shown. Note that this

Fig. 4. a) Original mammogram with a stellate lesion in the upper part of the image (marked A). b) SVM output from ROI extraction step. The pixel with the highest intensity (marked B) is taken as the first seed point for the level set boundary refinement algorithm, the result of which is shown in panel c)

is not the final classifying decision of the CAD system, but rather the first step localizing the ROIs that should be further processed. The coordinates with the highest response, as shown in panel b) are then extracted and passed on to the segmentation step.

3 Segmentation and Feature Extraction

From the image in panel b) of Fig. 4 the coordinates of the intensity maxima are extracted and a boundary refinement algorithm is initiated around this neighborhood. The local neighborhood and the result of the segmentation is shown in panel c). In this paper a level set method is used. Once an ROI has been segmented from the background, its immediate background is determined as all pixels within a distance d from the ROI where d is chosen such that the area of the background roughly corresponds to the area of the ROI. Then the extended ROI (the ROI and its immediate background) is removed from the ROI extractor grid output. This process is repeated until we have 10 regions of interest to pass on to the next steps in the algorithm: feature extraction and classification. Using the segmentation results, three of the five features are recalculated using the segmented ROI and its immediate surrounding instead of the washer-shaped neighborhoods used in the ROI extraction step. These are the contrast and the angle orientations in the interval $(\pi/2, \pi]$ and similar for the edge orientation. Four additional features are added to aid in the classification. The standard deviation of the interior of the ROI normalized with the square root of the intensity is a texture measure capturing the homogeneity of the area. An equivalent feature is extracted

for the immediate background as is the raw standard deviation of the of the surrounding pixel values. These three very simple texture features are complemented with the order at which the ROI was extracted in the ROI location step (1-10).

The same SVM implementation as mentioned above is trained with the features from these refined areas. The result is presented as a free-response receiver operating characteristic (FROC) curve.

4 Data Set, Scoring Protocol, and Evaluation Method

The data set consists of 90 mammograms from 45 cases of proven malignant cancers (one image from each of the CC and MLO view). All mammograms contain exactly one malignant tumor classified as a spiculated lesion by expert radiologists. The malignant areas have been manually annotated by radiologists and all images are taken from the Digital Database for Screening Mammography¹ (DDSM) at the University of South Florida [14]. The size and subtleties of the tumors are depicted in Figure 5. The tail towards large diameters is to some extent explained by some radiologists not marking the core outline of the tumor, but rather all tissue affected by metastasis. When available, the smaller core outline of the tumor has been used as ground truth. The skewed distribution of subtleties is somewhat concerning at first since a subtlety rating of 1 indicates that a tumor is very difficult to spot and a rating of 5 that it is obvious. However, this seems to be an effect of the rating protocol used by the radiologists at the institution where they were acquired and other researchers concluded that no clear bias towards easier cases is present in the DDSM (when compared to other data sets) [15].

A CAD-marking is considered a true positive if at least 1/3 of the automatically segmented area lies within the area annotated by radiologists. If there is an overlap, but less than a third, the marking is considered a false positive as are all markings not overlapping the annotated area.

The FROC curve quantifies the inherent trade-off between high sensitivity (true positive fraction) and false positive markings per image and its use in evaluation of mammography CAD schemes is well established. The FROC curve is a parameter curve traced out by successively lowering the threshold level of the classifying machine above which to consider a region malignant. For high thresholds, there will be few false positive markings but at the cost of missed cancers. Lowering the threshold level will increase the sensitivity but at the cost of more false positive markings.

The training and evaluation is performed with the leave-one-out method: for each of the 45 runs training is performed on 44 cases and evaluation on the 45th. The evaluation is performed pair-wise, *i.e.* a cancer is considered found if it is correctly marked in one of the two projections.

¹ <http://marathon.csee.usf.edu/Mammography/Database.html>

Fig. 5. Diameter and subtleties of tumors in the data set. Trade-off between false positive markings and sensitivity. The evaluation is done pairwise. At 90% sensitivity there are 0.32 false positive markings per image

5 Results

The resulting FROC curve is shown in Figure 5 and a brief comment on the specificity might be in place. On average, each image will have 0.3 false positive markings at a chosen sensitivity of 90%, thus making the method useless for diagnosis. The purpose of CAD (at least to date) has however not been diagnosis but detection, as mentioned in the introduction. The goal has rather been to draw the radiologists' attention to subtle areas that might otherwise have been overlooked, and then pass on the main responsibility for the classification to the human experts. In this context, 0.3 false positive markings per image is acceptable and is comparable to state-of-the-art algorithms. It must be noted however, that reported results are very dependent on the scoring protocol and data sets used [16], [17] and that the data set size used here (45 pairs of images) is too small to draw any strong conclusions about the relative performance of different algorithms. However, it has been demonstrated that spiculation and edge measures extracted using local phase and local orientation information from quadrature filter outputs could work well in a CAD algorithm designed to find spiculated lesions.

References

1. Tabar, L.: Control of breast cancer through screening mammography. *Radiology* 174 (1990) 655–656
2. Tabar, L., Yen, M-F., Vitak, B., Chen, H-H.T., Smith, R.A., Duffy, S.W.: Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. *Lancet* 361 (2003) 1405–1410

3. Freer, T.W., Ulissey, M.J.: Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology* 220 (2001) 781–786
4. Destounis, S.V. *et al.*: Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience, *Radiology* 232 (2004) 578–584
5. Karssemeijer, N., te Brake, G.M.: Detection of Stellate Distortions in Mammograms. *IEEE Trans. Med. Im.* 15(5) (1996) 611–619
6. te Brake, G.M., Karssemeijer, N.: Segmentation of suspicious densities in digital mammograms. *Med. Phys.* 28(2) (2001) 259–266
7. Kobatake, H., Yoshinaga, Y.: Detection of Spicules on Mammogram Based on Skeleton Analysis. *IEEE Trans. Med. Im.* 15(3) (1996) 235–245
8. Ng, S.L., Bischof, W.F.: Automated detection and classification of breast tumors. *Comput. Biomed. Res.* 25 (1992) 218–237
9. Kegelmeyer Jr, W.P.: Computer Detection of Stellate Lesions in Mammograms. In: *Proc. SPIE 1660* (1992) 446–454
10. Bornefalk, H.: Use of phase and certainty information from quadrature filters in detection of stellate patterns in digitized mammograms. In: *Proc. SPIE 5370* (2004) 97–107
11. Granlund, G.H., Knutsson, H.: *Signal Processing for Computer Vision*. Kluwer Academic Publishers, Dordrecht (1995)
12. Haglund, L.: *Adaptive Multidimensional Filtering*. PhD thesis, Linköping University, Sweden. Dissertation No. 284 (1992)
13. Cawley, G.C.: *MATLAB Support Vector Machine Toolbox (v0.50 β)*. University of East Anglia, School of Information Systems, Norwich, 2000. <http://theoal.sys.uea.ac.uk/~gcc/svm/toolbox>
14. Heath, M., Bowyer, K.W., Kopans, D.: Current status of the Digital Database for Screening Mammography. In: *Digital Mammography*. Kluwer Academic Publishers, Dordrecht (1998)
15. Petrick, N., Sahiner, B., Chan, H., Helvie, M.A., Paquerault, S., Hadjiiski, L.M.: Breast cancer detection: Evaluation of a mass-detection algorithm for computer-aided diagnosis—Experience in 263 patients. *Radiology* 224 (2002) 217–224
16. Nishikawa, R.M., Giger, M.L., Doi, K., Metz, C.E., Yin, F., Vyborny, C.J., Schmidt, R.A.: Effect of case selection on the performance of computer-aided detection schemes. *Med. Phys.* 21(2) (1994) 265–269
17. Giger, M.L.: Current issues in CAD for mammography. In: *Digital Mammography '96*. Elsevier Science, Philadelphia (1996) 53–59