ANALYSIS OF MAMMOGRAPHIC MICROCALCIFICATIONS USING A COMPUTATIONALLY EFFICIENT FILTER BANK

Thor Ole Gulsrud

Stavanger University College, Department of Electrical and Computer Engineering
Stavanger, Norway
thor.gulsrud@tn.his.no

ABSTRACT

We present a new method for classification of malignant and benign clusters of microcalcifications in digital mammograms. A computationally efficient infinite impulse response (IIR) quadrature mirror filter (QMF) bank is used as a tool for extracting texture and shape features. The filter bank splits the input image into four subbands: low-low band, low-high band, high-low band and high-high band. Texture and shape features based on co-occurrence matrices are computed from the subsampled subbands. The low-low band extracts the information of spatial dependence and the higher frequency bands extract the shape information. The results of an experimental study demonstrate that our approach drastically improves the overall performance compared to a manual system.

1. INTRODUCTION

It is difficult to distinguish between benign and malignant microcalcifications associated with breast cancer. Only 20% - 30% of breast biopsy cases recommended by radiologists prove to be of malignant nature [1]. This high false positive call rate (number of negative biopsies divided by the total number of women screened who did not have cancer) can be attributed to several factors, including poor image quality, radiologist fatigue, and human oversight. Given the high frequency of breast biopsy for benign diagnosis, even a small reduction in the false positive call rate could save a substantial amount of patient anxiety and money.

A computer-aided diagnosis (CAD) system to help decision making for biopsy recommendation can be of significant value. In this context, several investigators have addressed the question of the relation of the shapes and sizes of clustered microcalcifications to the risk of breast cancer. Consequently, many of the proposed CAD systems for analysis of clustered microcalcifications are based on shape features [2]. However, microcalcifications may be very subtle, be of low contrast and have hazy borders. Therefore, shape features of individual microcalcifications may not be reliable. It has however been demonstrated in clinical studies that the grouping of microcalcification regions to define the shape of the cluster is highly dependent on the structure and texture (the tissue texture in regions containing microcalcifications associated with a malignant process may be different from that associated with a benign process) of the image [3]. Our new classification method, presented in this paper, is based on both texture and shape features. First, we use a filter bank to split the input image into different subbands. Thereafter, texture features and a shape feature based on co-occurrence matrices are computed from these subbands. Finally, we use these features in the discrimination between benign and malignant clusters of microcalcifications.

In [4] Chan et al. achieved improved system performance by combining texture and shape features. A disadvantage of their method was that the locations of the individual microcalcifications had to be known. In their work the individual microcalcifications were manually identified. Dhawan et al. [2] computed different texture features from both the whole region of interest (ROI) and from the subbands of the multichannel decomposed ROI. However, features representing cluster formation (e.g. average size of individual microcalcifications and number of microcalcifications in the cluster) were computed from segmented microcalcification regions. In our method, no segmentation of individual microcalcifications is necessary. We extract structural information directly from the subbands of the multichannel decomposed ROI. Consequently, our method is less complex and more elegant than the one proposed by Dhawan et al. As will be demonstrated in an experimental study, this new classification method drastically improves the overall performance compared to a manual system.

2. MATERIALS AND METHODS

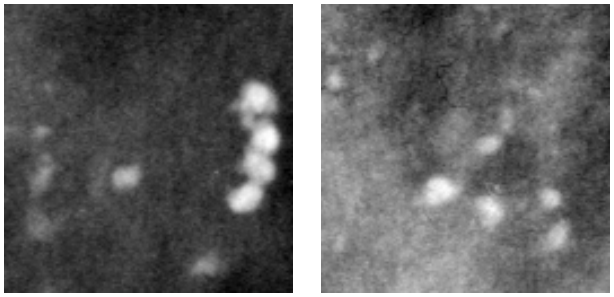
2.1. Data acquisition

The Mammographic Image Analysis Society (MIAS) [5], which is an organization of United Kingdom research groups interested in the understanding of mammograms, has produced a digital mammography database which we have chosen to use in our research. The X-ray films in the database have been carefully selected from the United Kingdom National Breast Screening Programme and digitized with a Joyce-Loebl scanning microdensiometer to a resolution of $50\mu m \times 50\mu m$. Each pixel is represented by 8 bits. An important characteristic of the MIAS database is that each abnormal image comes with a consultant radiologist's truth information, i.e., the locality of the abnormality is given as the coordinate of its center and an approximate radius (in pixels) of a circle enclosing the abnormality.

From the MIAS database we use 25 mammograms from which we have extracted 40 non-overlapping subimages - 20 containing a malignant cluster of microcalcifications, and 20 containing a benign cluster of microcalcifications - all with size 128×128 pixels. Figure 1 shows one sample from each class in the data set.

2.2. Texture feature extraction and classification

Our method for automated analysis of clustered microcalcifications is based on a combination of two different approaches to



a b

Fig. 1. Sample (128 × 128 pixels/64 $mm \times 64\ mm$) of a benign- (a) and a malignant (b) cluster of microcalcifications.

texture feature extraction; multichannel filtering and gray level co-occurrence matrices. Ideally, a texture feature should describe both the primitives out of which the texture is composed, and the spatial relationships between them. However, the weakness of the gray level co-occurrence approach is that it does not capture the shape aspects of the gray level primitives, i.e., the two characteristics are not completely and simultaneously described. In other words, we will lose valuable information if the co-occurrence method is applied in a straightforward manner. Thus, the important question is: Can we modify the method in such a way that the two basic characteristics are completely and simultaneously satisfied? In our experimental study we demonstrate that a useful modification is to perform a multichannel decomposition of the texture image prior to the computation of the co-occurrence matrices. We have described the *combined* approach to texture feature extraction in detail previously [6]. Thus, in the following only a very brief description of the texture feature extraction scheme is presented.

2.2.1. The filter bank

In this work we investigate two-channel infinite impulse response (IIR) quadrature mirror filter (QMF) banks. Figure 2 shows the 1-D two-channel analysis IIR QMF bank used in our experiments. The IIR filter bank is implemented with *all-pass* filters, which are

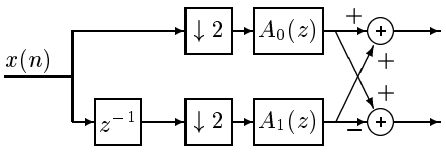


Fig. 2. Two-channel analysis IIR based QMF bank.

attractive due their computational simplicity [7]. The IIR filters are given by

$$H_0(z) = A_0(z^2) + z^{-1}A_1(z^2), \quad (1)$$

$$H_1(z) = A_0(z^2) - z^{-1}A_1(z^2), \quad (2)$$

where $A_i(z)$ is all-pass, that is $A_i(z) \cdot A_i(z^{-1}) \equiv 1,\ i = 0, 1$. To illustrate, employing filters of first order, $A_0(z)$ and $A_1(z)$ can be written as

$$A_0(z) = \frac{a_0 + z^{-1}}{1 + a_0 z^{-1}},$$

$$A_1(z) = \frac{a_1 + z^{-1}}{1 + a_1 z^{-1}}.$$

The extension to 2-D signals is quite straightforward provided the filters employed are separable [8]. In implementing a 2-D system based on separable filters, we can perform 1-D filtering first along the rows of the 2-D signal. This process is then followed by 1-D filtering along the columns of the *result* of the first filtering operation. The 2-D QMF bank splits the image into four subbands: low-low (LL) band, low-high (LH) band, high-low (HL) band, and high-high (HH) band. The LL-band texture is a smoothed version of the input texture, while the LH-band, the HL-band, and the HH-band textures are high-pass filtered textures in the x direction, y direction, and both x and y directions, respectively.

2.2.2. Gray level co-occurrence matrices

The *Gray Level Co-occurrence Matrix* (GLCM) counts how often pairs of gray levels of pixels, that are separated by a certain distance along a certain direction, occur in a digital image. In the present work we have computed the GLCM for four different directions: 0°, 45°, 90°, and 135°. The definitions of the GLCM are given in [9].

The GLCM's are seldom used directly. Instead features based on them are computed. In this paper we want to evaluate the texture features proposed by Haralick et al. [9] in the classification of clustered microcalcifications in digital mammograms.

2.2.3. The combined approach

A block diagram of the combined approach is shown in Figure 3. The LL-band is used to extract the information of spatial dependence, i.e., the Haralick features are computed from this subband. In order to extract structural information we use a texture feature called *zero-crossing* (ZC) [10]. In computing the ZC feature the high-pass filtered subband textures are first quantized into binary textures, i.e., textures having only the two gray levels black (0) and white (1). This is due to the fact that the histograms of the high-pass filtered subbands textures are mostly distributed at these two gray levels [10]. In the following we assume that the texture images have 256 gray levels. The quantization is then performed using binary decision with a threshold value of 128, i.e., if a pixel in the input image has gray level less than 128 its gray level is set to 0, otherwise it is set to 255. Having only two gray levels, the size of a GLCM of a high-pass filtered subband texture is 2×2. The ZC feature based on any of the high-pass filtered bands is then computed as the number of times we get a transition $0 \rightarrow 255$ or $255 \rightarrow 0$, given a certain direction θ and a certain distance d.

In order to form one particular texture feature from the high-pass filtered bands, a ZC feature is first calculated from each of the four ($\theta = 0^\circ,\ 45^\circ,\ 90^\circ,\ 135^\circ$) co-occurrence matrices based on the quantized subbands. The results from the twelve calculations (four calculations in each of the three subbands) are then averaged together, thus providing some degree of rotation invariance. As shown in Figure 3, the features from the LL-band are combined with the ZC feature in order to create a multi-dimensional feature vector.

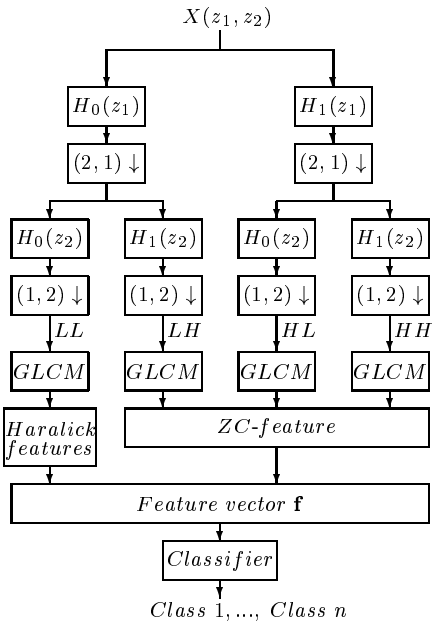


Fig. 3. The combined approach.

The investigation of different classification algorithms is not a topic in this work. Thus we have chosen a classifier which is already implemented, well tested and has shown good performance in many practical applications. Such a classifier is the Learning Vector Quantization (LVQ) algorithm of Kohonen [11]. We employ the LVQ classifier in a *supervised* manner, i.e., it is trained using a labeled data set. Note that due to the small size of the data set we employ the LVQ classifier by means of the *leave-one-out* method. This procedure takes N training samples, trains the classifier on $N-1$ samples, then uses the remaining one sample to test. Classification is continued in this manner until all N samples have been used as the test sample. Final performance is reported as an average of results for all N trails.

2.3. Performance evaluation

For the performance evaluation of our method we use the following definitions:

- *True Positive (TP) rate*: The ratio of the number of malignant cases correctly classified to the total number of malignant cases in the test set.
- *False Positive (FP) rate*: The ratio of the number of benign cases incorrectly classified to the total number of benign cases in the test set.

It should be mentioned that each image used in the experimental study contains one case (benign or malignant).

In addition we define the *Overall Performance* (OP) rate of a classification method as follows:

$$OP \text{ rate} = TP \text{ rate} \cdot \frac{MC}{NIm} + (100 - FP \text{ rate}) \cdot \frac{BC}{NIm} \quad (3)$$

where MC is the number of malignant cases in the test set, BC is the number of benign cases in the test set, and NIm is the total number of images in the test set.

Feature	Without ZC feature			ZC feature included		
	TP	FP	OP	TP	FP	OP
ASM	71.7%	58.4%	56.7%	73.0%	35.3%	68.9%
COR	52.0%	59.7%	46.2%	71.5%	31.1%	70.2%
VAR	75.7%	74.6%	50.6%	75.2%	35.9%	69.7%
IDM	62.9%	67.5%	47.7%	74.7%	33.3%	70.7%
SENT	55.5%	61.9%	46.8%	72.8%	29.6%	71.6%
ENT	70.6%	59.8%	55.4%	75.2%	34.8%	70.2%
DENT	58.0%	62.2%	47.9%	76.1%	29.0%	73.6%
IMC1	54.7%	60.8%	47.0%	74.9%	33.3%	70.8%
IMC2	51.9%	52.7%	49.6%	71.6%	31.7%	70.0%

Table 1. Classification results for IIR filter $i\_2\_1\_09$.

3. RESULTS AND DISCUSSION

We investigate the performance of the following nine Haralick features [9]: Angular second moment (ASM), Correlation (COR), Variance (VAR), Inverse difference moment (IDM), Entropy (ENT), Sum entropy (SENT), Difference entropy (DENT), Information measures of correlation 1 (IMC1), and Information measures of correlation 2 (IMC2).

In order to find the filters best suited for our application we performed experiments on different combinations of filters and the selected texture features. Based on computed OP rates, we selected the IIR filter $i\_2\_1\_09$ to be used in the further analysis. This filter has formerly been used in both texture classification [12, 6], and texture segmentation [13].

One problem with the combined approach is the arbitrary choice of the distance parameter d that determines the relationship between two pixels for the second order calculations. Initially, we tested the performance of the combined method using distance parameters of $d = 1$, 2, 3, and 4. The results demonstrated that the combined method using GLCM's computed with the distance parameter $d = 1$ is better, - in terms of classification accuracy, than the method using GLCM's computed on distance parameters of d =2, 3, and 4. Hence, the textures in our data set may be defined as *fine*, resulting in poorer classification performance when using texture features based on GLCM's computed on distance parameters greater than one. Due to this observation, only the discrimination power of selected texture features extracted from the $d = 1$ GLCM's are examined in the further experiments.

In Table 1 we show the classification results for the nine single Haralick features in combination with the IIR filter $i\_2\_1\_09$. From the far right columns we observe that the combined method based on this filter yield approximately similar classification results ($\approx 70\%$) for all nine features. More interesting: If we do not include the information from the higher frequency bands (i.e., the ZC feature) in the feature vector, the OP rates drop significantly. As shown in Table 2, similar results are obtained if we compute the Haralick features from the co-occurrence matrices of the whole test image.

Having evaluated the classification performance of each of the nine texture features, the next step is to evaluate the classification performance of different combinations of features. However, considering the fact that the nine features can be combined in 511 different ways, it is obvious that some kind of feature selection procedure needs to be carried out. Our selection of "best" features are based on the OP rates in Table 1. Only the five best features in this table are kept for further analysis. Thus, from Table 1 we select the features COR, IDM, SENT, DENT and IMC1. The work by Gabrielsen [15] shows that no improvements in the classification

Feature	TP	FP	OP
ASM	60.1%	40.3%	59.9%
COR	62.7%	59.2%	51.8%
VAR	70.3%	70.4%	50.0%
IDM	59.8%	66.0%	46.9%
SENT	57.8%	33.5%	62.2%
ENT	55.1%	32.6%	61.3%
DENT	66.6%	57.2%	54.7%
IMC1	56.4%	62.9%	46.8%
IMC2	56.3%	52.8%	51.8%

Table 2. Classification performance of the Haralick features with the filtering step left out (i.e., "pure" co-occurrence method).

Feature-set	TP	FP	OP
COR, IDM, SENT, ZC	82.0%	28.6%	76.7%
IDM, DENT, ZC	81.3%	28.8%	76.3%
COR, IDM, ZC	83.0%	30.8%	76.1%
IDM, SENT, ZC	79.7%	28.3%	75.7%
COR, SENT, DENT, ZC	82.0%	30.7%	75.7%
SENT, DENT, ZC	79.1%	27.9%	75.6%

Table 3. IIR filter $i\_2\_1\_09$: The feature-sets yielding the five highest OP rates.

performance are gained by using feature vectors of higher dimensionality. For a texture classification problem similar to the one addressed in the present paper, Gabrielsen based the selection of best features on an OP rate $\geq 68\%$, implying that six features were selected for further analysis. The classification performance of all 63 combinations of the selected features were evaluated. The results showed that no improvements were gained by including more than three Haralick features in the feature vector.

The feature sets yielding the five highest OP rates are presented in Table 3. We observe that by increasing the dimensionality of the feature vector the classification performances are somewhat improved. However, no improvements are gained by including more than three Haralick features in the feature vector. Note that there is a trade-off between the TP and FP rates. For example, the relatively high FP rate of the feature set COR, IDM, and ZC in Table 3 is compensated by a high TP rate. Also, note that some features that are not very useful by themselves can become effective features when they are combined with other features. As an example, from Table 1 we observe that the feature COR, which is a measure of gray level linear dependencies in the texture image, yields relatively low OP rates by itself. However, this feature is included in 50% of the feature sets in Table 3.

4. CONCLUSION

Contribution to a practical computer based mammographic interpretation system is made with the introduction of a novel texture feature extraction method for diagnosis of detected clustered microcalcifications. Our method is based on a combination of two different approaches to texture feature extraction: The multichannel filtering approach and the co-occurrence approach. A great advantage of our method is that structural information can be extracted without knowledge of the localization of the individual microcalcifications contained in a cluster.

In the experiments we have compared the classification capability of the combined method with that of statistical descriptors based on co-occurrence matrices of the whole image. For the combined method we have used a IIR QMF bank, implemented with first order all-pass filters, which are attractive due to their computational simplicity. The experiments showed that the combined method have better classification performance than that of statistical descriptors computed from the co-occurrence matrices of the whole image.

Considering the fact that only 20% - 30% of breast biopsy cases recommended by radiologists prove to be of malignant nature, the results of the experiments showed that our proposed classification method can provide radiologists with a second opinion.

5. REFERENCES

[1] E. A. Sickles, "Breast calcifications: Mammographic evaluation," *Radiology*, no. 160, pp. 289–293, 1986.

[2] A. P. Dhawan, Y. Chitre, C. Kaiser-Bonasso, and M. Moskowitz, "Analysis of Mammographic Microcalcifications Using Gray-Level Image Structure Features," *IEEE Trans. Med. Imag.*, vol. 15, no. 3, pp. 246–259, 1996.

[3] M. Lanyi, *Diagnosis and differential diagnosis of breast calcifications*. Springer Verlag, 1986.

[4] H.-P. Chan, B. Sahiner, N. Petrick, M. A. Helvie, K. L. Lam, D. D. Adler, and M. M. Goodsitt, "Computerized classification of malignant and benign microcalcifications on mammograms: texture analysis using an artificial neural network," *Phys. Med. Biol.*, vol. 42, pp. 549–567, 1997.

[5] J. Suckling, J. Parker, D. R. Dance, S. Astley, I. Hutt, C. R. M. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S.-L. Kok, P. Taylor, D. Betal, and J. Savage, "The Mammographic Image Analysis Society digital mammogram database," in *Proceedings of the 2nd International Workshop on Digital Mammography*, (York, England), pp. 375–378, Jul 1994.

[6] T. O. Gulsrud and J. H. Husøy, "Image texture classification using Quadrature Mirror Filter bank in combination with Co-occurrence matrices," in *Proc. SPIE International symposium on optical instrumentation and applied science: Applications of Digital Image Processing XVII*, (San Diego), pp. 497–506, July 1994.

[7] J. H. Husøy, "Low complexity subband coding of still images and video," *Optical Engineering*, vol. 30, pp. 904–911, July 1991.

[8] M. Vetterli, "Multi-dimensional sub-band coding: Some theory and algorithms," *Signal Processing*, vol. 6, pp. 97–112, Apr. 1984.

[9] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cyb.*, vol. 3, pp. 610–621, Nov. 1973.

[10] A. Kundu and J. L. Chen, "Texture classification using qmf bank-based subband decomposition," *CVGIP: Graphical models and image processing*, vol. 54, no. 5, pp. 369–384, 1992.

[11] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, pp. 1464–1480, Sept. 1990.

[12] J. H. Husøy, T. Randen, and T. O. Gulsrud, "Image texture classification with digital filter banks and transforms," in *Proc. SPIE Int. Symp. on Optical Inst. and Applied Science: Appl. of Digital Image Proc. XVI*, (San Diego), pp. 260–271, July 1993.

[13] T. Randen and J. H. Husøy, "Multichannel filtering for image texture segmentation," *Optical Engineering*, vol. 33, pp. 2617–2625, Aug. 1994.

[14] C. M. Kocur, S. K. Rogers, L. R. Myers, T. Burns, M. Kabrisky, J. W. Hoffmeister, K. W. Bauer, and J. M. Steppe, "Using neural networks to select wavelet features for breast cancer diagnosis," *IEEE Engineering in Medicine and Biology*, vol. 15, no. 3, pp. 95–102, 1996.

[15] S.-O. Gabrielsen, "Automated classification of microcalcifications using a multichannel filtering approach," Master's thesis, Stavanger College, Norway, 1995.