International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:11, 2008

Frame Texture Classification Method (FTCM) Applied on Mammograms for Detection of Abnormalities

Kjersti Engan, Karl Skretting, Jostein Herredsvela, Thor Ole Gulsrud

Abstract**—** Texture classification is an important image processing task with a broad application range. Many different techniques for texture classification have been explored. Using sparse approximation as a feature extraction method for texture classification is a relatively new approach, and Skretting et al. recently presented the Frame Texture Classification Method (FTCM), showing very good results on classical texture images. As an extension of that work the FTCM is here tested on a real world application as detection of abnormalities in mammograms. Some extensions to the original FTCM that are useful in some applications are implemented; two different smoothing techniques and a vector augmentation technique. Both detection of microcalcifications (as a primary detection technique and as a last stage of a detection scheme), and soft tissue lesions in mammograms are explored. All the results are interesting, and especially the results using FTCM on regions of interest as the last stage in a detection scheme for microcalcifications are promising.

Keywords**—** detection, mammogram, texture classification, dictionary learning, FTCM

I. Introduction

THE word *texture* is in general regarded as surface appearance or tactile qualities. A texture can be regarded as a self-similar object. In image processing the texture of a region describes the pattern of spatial variation of gray tones (or in the different color bands in a color image) in a neighborhood that is small compared to the region. By definition, *texture classification* is to identify the texture class in a region, whereas *texture segmentation* means finding a boundary map between different texture regions of an image [1]. There is an ambiguity here since classification can be used for segmentation. We use the term *texture classification* in the following even though the goal of the classification is segmentation. Most texture classification algorithms start by finding a local feature vector which in turn is used for classification.

Texture classification using learned (overcomplete) dictionaries and sparse representation is a relatively new area in texture classification. The work presented here is a continuation of the work by Skretting et al. in [2] where the algorithm *Frame Texture Classification Method (FTMC)* was introduced. The main motivation of this work is to show that the FTCM method can be useful in a practical application, like a part of a detection system for abnormalities in mammograms. The FTCM is described in Section II, and in short the feature extraction part simply consists of organizing the neighborhood of pixel in a texture image into a feature vector. The classification part is more complicated and should be regarded as a supervised vector classification method. Some extensions to the original FTCM is done by testing different types of smoothing as well as a vector augmentation step described in Section II-C and II-D respectively.

Manuscript received January 4, 2007

Kjersti Engan is with the dept. of Electrical and Computer Eng. at University of Stavanger, N-4036 Stavanger, Norway, (email:kjersti.engan@uis.no)

Karl Skretting is with the dept. of Electrical and Computer Eng. at University of Stavanger, N-4036 Stavanger, Norway, (email:karl.skretting@uis.no)

Jostein Herredsvela is with Statoil - Geophysical Data Quality, Grenseveien 21 NO-4035 Stavanger, Norway (email:JOSHE@statoil.com)

Thor Ole Gulsrud is with Roxar Flow Measurement AS, Stavanger, Norway (email: Thor.Ole.Gulsrud@roxar.com)

A. Mammogram

Breast cancer is the most frequent cancer disease, and a leading cause of cancer death among women. In Norway, populated by 4.5 million people, as many as 2644 women developed breast cancer and 715 died of breast cancer in 2003 [3]. The survival rate is greatly influenced by how early the cancer is discovered and treated. Mammograms are X-ray projections of the breast tissue onto a detector array or a film plate. The exposure of the mammographic film or detector array is a function of the intensity of the electromagnetic waves transmitted through the breast. Most western countries today have *mammographic screening* programs. A screening is defined as the presumptive identification of unrecognized disease or defect by application of tests, examinations, or other procedures. In Norway every women between 50 and 69 years of age are invited to screening every second year. Evaluating screening data is a very labor and time demanding process. Cancer is only found in a small percentage of the cases (approx. 0.6 % in Norway), and due to the vast amount of images relatively rapid interpretation is done. Hence there is a risk that subtle abnormalities can be overlooked. In Norway today, two independent radiologists evaluate all screening data. This is widely recognized as the best way of evaluating screening data. However, due to lack of qualified radiologists, or for economical reasons, screening data are evaluated by a single radiologist in some countries/regions/hospitals.

There are two main types of mammographic findings that indicate a possibly cancer in the breast tissue: Soft tissue lesions (tumors) and clusters of microcalcifications. Soft tissue lesions can be subdivided in different groups, and can be malignant or benign. Soft tissue lesions are often palpable. The two most common groups are spiculated masses (usually malignant) and circumscribed masses (usually benign). The size, shape and density varies much. Clusters of microcalcifications are early signs of possible cancer or precancerous

changes, and are in general *not* palpable. Microcalcifications are small calcifications of different shape and density, approx. 0.1 - 1 mm in diameter, and a cluster of microcalcifications is often defined as minimum three microcalcifications within an area of 1 cm^2. Single microcalcifications are never malignant, but clusters of microcalcifications might be malignant or benign. Tumors and microcalcifications are denser than the surrounding tissue and thus absorb more of the incident X-rays. They can therefore often be seen as bright spots/regions in the mammograms.

A well working Computer-Aided Detection (CAD) system could be used as a second opinion for the radiologists. Whether CAD can *replace one* of two independent radiologists is under discussion and still needs to be proven. However, CAD would, no doubt, be helpful where no second radiologist is at hand. A CAD system can also be helpful in clinical analysis and in training of radiologists. As early as in 1990 Chan, Doi et al. showed that CAD (state of the art in 1990) did significantly improve radiologists' accuracy in detecting clustered microcalcifications under conditions that simulated the rapid interpretation of screening mammograms [4].

A microcalcification is very small and often bright and it s not presumed to have a distinct texture within the calcification. A *cluster of microcalcifications*, on the other hand, has an obvious texture distinctly different from the background tissue. The edges of the microcalcifications becomes an important part of the texture, and exploiting this also single microcalcifications can be detected. Regarding masses, the texture inside the lesion (spiculated or circumscribed) will presumptive be different than the texture of the surrounding tissue. Thus texture classification techniques can be useful for detection and/or classification purposes for both clustered microcalcifications and masses.

A lot of different work has been carried out on detection of microcalcifications and/or masses, some of them using textural features as a part of the detection scheme. Examples are Chan et al. [5], Mudigonda et al. [6]

II. Overcomplete dictionaries and texture classification

Using an overcomplete dictionary matrix $\mathbf{F} \in \mathcal{R}^{N\times K}$ containing K dictionary vectors as the columns, $\{\mathbf{f}_i\}_{i=1}^{K}$, a signal vector $\mathbf{x} \in \mathcal{R}^N$ can be represented as a linear combination of the dictionary vectors. The term *overcomplete* refers to $K > N$. There will be many possible expansion of the signal, and a full expansion is obviously redundant. In many applications, for example in signal compression or when a signal expansion is used for feature extraction, an efficient representation is desired, and when using an overcomplete dictionary this is consistent with a *sparse* representation. The sparseness can be increased further if the application only requires a sparse *approximation* of the signal:

$$\mathbf{x} - \mathbf{r} = \tilde{\mathbf{x}} = \mathbf{F}\mathbf{w}, \tag{1}$$

where $\mathbf{r}$ is a residual vector, and $\mathbf{w}$ is a *sparse* coefficient vector, i.e. with many/most of the entries in the coefficient vector equal to zero. When $K > N$, and especially if $K \gg N$, a highly sparse $\mathbf{w}$ can often give a good solution to $\min \|\mathbf{x} - \tilde{\mathbf{x}}\|^2$ subject to $\mathbf{x} = \mathbf{F}\mathbf{w} + \mathbf{r}$.

If $K \gg N$, in the limit we can get a good approximation $\tilde{\mathbf{x}} = \mathbf{F}\mathbf{w}$ where only one entry in $\mathbf{w}$ is non-zero. If $\mathbf{w}$ represent a quantized vector, also in the non-zero entry, this is equivalent to do shape-gain Vector Quantization (VQ). Letting K be even larger this can be approximated as $\tilde{\mathbf{x}} = \mathbf{F}\mathbf{s}$ where $\mathbf{s}$ is a select vector, i.e. one entry is equal to one, and the rest is zero. This is equivalent with ordinary VQ, where $\mathbf{F}$ now is the VQ codebook, or *dictionary*. Such a VQ dictionary is often learned using a training set, and the most widely used learning algorithm is called the Generalized Lloyd Algorithm (GLA) [7], which is equivalent to unsupervised learning using *k-means* in pattern recognition. In the following we are seeking to represent vectors, generated from pixels of an image we want to classify, using a sparse representation as in Equation 1.

A. Vector selection/pursuit algorithms

Finding the optimal sparse approximation to a signal vector using an overcomplete dictionary is NP-hard [8]. Instead some vector selection/pursuit algorithm is used to select vectors from a dictionary. There exist different types of vector selection algorithms, such as Matching Pursuit (MP) type of algorithms [9], [10], [11], Basis Pursuit [12], and FOCUSS [13], [14]. In earlier experiments we have seen that the Order Recursive Matching Pursuit (ORMP) is often a good choice when we are concerned with both approximation capability and computational complexity. Thus, a fast implementation of ORMP is used in this work [10].

B. Dictionary Learning

The quality of a sparse representation of a class of signals is highly dependent on the dictionary. If the dictionary is well suited for the class of signals, a good quality of fit, i.e. a low mean square error, can be obtained with a very sparse representation/approximation. Thus finding a good dictionary for a specific class of signals is important.

The algorithm used for learning dictionaries in this work, and presented in this section, is the algorithm called Method of Optimal Directions (MOD) by Engan et al. [15], [16]. This algorithm is the core algorithm of a larger family of algorithms named Iterative Least Squares based Dictionary Learning Algorithms (ILS-DLA) by Engan et al. [17]. MOD is an algorithm for learning unrestricted block based dictionaries, whereas ILS-DLA includes block based and overlapping dictionaries, with or without various restrictions/constraints.

The ILS-DLA algorithms require a training set of signal blocks representative for the class of signal we want to represent: $\{\mathbf{x}_l\}_0^{L-1}$. For a one dimensional signal and the block-oriented case the synthesis equation for an approximation of a signal block, $\mathbf{x}_l$, is $\tilde{\mathbf{x}}_l = \mathbf{F}\mathbf{w}_l$. Ordering the training set of signal blocks as columns in a matrix gives: $\mathbf{X} \approx \mathbf{F}\mathbf{W}$, where $\mathbf{X} = [\mathbf{x}_0\, \mathbf{x}_1 \ldots \mathbf{x}_{L-1}]$ and $\mathbf{W} = [\mathbf{w}_0\, \mathbf{w}_1 \ldots \mathbf{w}_{L-1}]$.

The optimal dictionary will depend on the target sparseness factor and the class of signals we want to represent. We want to find the dictionary, $\mathbf{F}$, of size $N \times K$ where $K > N$, and

the sparse coefficient vectors, $\mathbf{w}_l$, that minimize the sum of the squared errors. The objective can be written as:

$$\text{Argmin}_{\mathbf{F},\mathbf{W}}\|\mathbf{X}-\mathbf{F}\mathbf{W}\|_F^2, \text{ subject to } \forall l,\ \|\mathbf{w}_l\|_0 = \text{Const}, \quad (2)$$

or an alternative objective is:

$$\text{Argmin}_{\mathbf{F},\mathbf{W}}\|\mathbf{w}_l\|_0\ \forall l, \text{ subject to } \|\mathbf{X}-\mathbf{F}\mathbf{W}\|_F^2 \leq \epsilon, \quad (3)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The training vectors $\mathbf{x}_l$ are columns in the matrix $\mathbf{X}$, and the corresponding coefficient vectors $\mathbf{w}_l$ are columns in a coefficient matrix $\mathbf{W}$. Finding the optimal solution to these problems are difficult, if not impossible. The iterative solution strategy presented below results in good, but in general suboptimal, solutions to the two problems.

The algorithm starts with a user supplied initial dictionary $\mathbf{F}^{(0)}$ and then improves it by iteratively repeating two main steps:

1) $\mathbf{W}^{(i)}$ is found by vector selection (ORMP is used in this work) using dictionary $\mathbf{F}^{(i)}$.
2) $\mathbf{F}^{(i+1)}$ is found from $\mathbf{X}$ and $\mathbf{W}^{(i)}$, where the objective function is
   $J(\mathbf{F}) = \|\mathbf{X}-\mathbf{F}\mathbf{W}^{(i)}\|_F^2$.
   This gives:
   $$\mathbf{F}^{(i+1)} = \mathbf{X}(\mathbf{W}^{(i)})^T(\mathbf{W}^{(i)}(\mathbf{W}^{(i)})^T)^{-1} \quad (4)$$
   Before proceeding all the dictionary vectors are normalized. $i = i+1$, go to step 1 as long as a stop criterion is not fulfilled.

i is the iteration number. The first step is suboptimal due to the use of practical pursuit algorithms, while the second step finds the $\mathbf{F}$ that minimizes the objective function.

Which initial objective function we wan to solve is dependent on the required/decided stopping rule used with the pursuit algorithm. If the stopping rule is a predefined sparsity, i.e. a constant number of nonzero entries, it corresponds to the objective of Equation 2. If the stopping rule is to continue to select vectors until a quality criterion on the approximation is fulfilled it corresponds to the objective of Equation 3. Which objective is the most suitable is dependent on the application. In this paper the stopping rule is always a predefined sparsity, thus the objective is the one stated in Equation 2.

C. Texture classification using sparse representation and learned dictionaries

Skretting et al. introduced a method for texture classification using sparse representation and learned dictionaries named *Frame Texture Classification Method (FTCM)* in [2], [18]. The method should be regarded as a supervised vector classification method. The classification of each test vector is done by comparing the representation errors of a number of different sparse representations, each using an overcomplete dictionary. For the purpose of texture classification, the test vectors are generated from pixels in the texture image. The underlying texture model for the development of FTCM is a deterministic model where a texture is modelled as a tiled floor where all tiles are identical. The color, or gray-level, at a given position in the floor is given by an underlying continuous periodic two-dimensional function. Based on this model it is shown that a vector generated from a spatial neighborhood is indeed a sparse combination of at most four vectors from a finite, but often quite large, dictionary. More details can be found in [2].

The main idea of the algorithm is as follows: Consider an image, $\mathbf{I}_m$, where every pixel should be classified as belonging to a certain texture/class from a predefined number of classes, C. Let $I_m(j,k)$ denote the image pixel at position (j,k). For each pixel in the (possible preprosessed) image, a vector $\mathbf{y}_l$ is made from the specific pixel and a number of neighborhood pixels. The size and shape of the neighborhood is predefined. More information and tests on neighborhood sizes and shapes can bee seen in [19]. The vector $\mathbf{y}_l$ can be regarded as a *feature vector* for pixel $I_m(j,k)$. For each possible texture class, $i = 1,\dots C$, a dictionary, $\mathbf{F}_i$, is learned using a training set. The training vectors are constituted from training images the same way as the feature vector. The dictionaries are learned using the MOD algorithm described in Section II-B. The feature vector, $\mathbf{y}_l$, is then represented sparsely (with a predefined sparseness factor, S, i.e. number of nonzero entries in $\mathbf{w}$) by all the different dictionaries, and the different residuals are calculated:

$$r_l^i = \|\mathbf{y}_l - \mathbf{F}_i\mathbf{w}_l^i\|, \quad i = 1,\dots C. \quad (5)$$

This is done for every single pixel in the (possible preprossessed) image. The resulting residuals are arranged as a new image for each class, a residual image $\mathbf{R}_i$, of the same size as the original texture image.

Texture, by definition, is not pixel-by-pixel local except in the edge between two textures. Thus it is natural to do some sort of smoothing before/during the final classification. Two different types of smoothing are used in this paper: $\mathbf{Sm_1}$ is a smoothing based on lowpass filtering as was done in [2], and $\mathbf{Sm_2}$ is smoothing based on morphological filtering and is a new type of smoothing tested with the FTCM in the present work. Both smoothing techniques give a reduction in the total classification error and better classification within the texture regions, but the cost of the smoothing is often more classification errors along the borders between different texture regions, as expected.

Sm1: The residual images can be smoothed before classification. The C residual images are smoothed using an $a \times a$ pixels separable Gaussian low-pass filter with variance σ^2, as done in FTCM, resulting in $\mathbf{Rs}_i$. The last part is a simple classification. Let $\mathbf{I}_{class}$ be an image of the same size as $\mathbf{I}_m$. Let the value of $I_{class}(j,k)$ correspond to the class no. of pixel $I_m(j,k)$ and in general $A(j,k)$ correspond to the pixel at position (j,k) in an image $\mathbf{A}$:

$$I_{class}(j,k) = \{i : Rs_i(j,k) < Rs_l(j,k);\ \text{all } l \neq i\}. \quad (6)$$

Sm2: Find an unsmoothed classification image, $\mathbf{I}_{class}^{us}$:

$$I_{class}^{us}(j,k) = \{i : R_i(j,k) < R_l(j,k);\ \text{all } l \neq i\}. \quad (7)$$

A classification image $\mathbf{I}_{class}$ is now found by filtering the image $\mathbf{I}_{class}^{us}$ with a morphological filter finding the class that is repeated most times in $\mathbf{I}_{class}^{us}$ within the

neighborhood of the filter, or structure element, $\mathbf{B}$. For $i = 1, 2 \dots C$:

$$p_i(j,k) = |\{i : (i \in \mathbf{I}^{us}_{class}) \cap (1 \in \mathbf{B}_{(j,k)})\}| \quad (8)$$

$$I_{class}(j,k) = \max_i p_i(j,k), \quad (9)$$

where $|A|$ denotes the *cardinality* of a finite set A, i.e. the number of elements in A.

D. Vector augmentation

The vector augmentation technique presented and tested here is an extension of the original FTCM scheme. As previously mentioned, the dictionaries used with FTCM have normalized columns. Consider the situation that one initial dictionary is created from a certain set of vectors and that a second initial dictionary is created from the same set of vectors multiplied by a constant. After normalization the two dictionaries will be exactly the same. Augmenting the vectors with an extra and constant element, z, prior to normalization results in two different dictionaries.

In texture classification sometimes the average gray level is higher for some textures than others. Augmenting the vectors (training and test/use) ensures that the dictionaries respond differently to dark and bright regions. Let $\mathbf{X}_i$ be an $N \times L$ training matrix for class i. The training matrix is augmented with an extra row containing L elements of value z. The new training matrix is denoted $\mathbf{X}_i^{'}$:

$$\mathbf{X}_i^{'} = \begin{bmatrix} \mathbf{z} \\ \cdots \\ \mathbf{X}_i \end{bmatrix}, \quad (10)$$

where $\mathbf{z}$ is a constant vector with equal element values z. The initial dictionary associated with class i is formed by randomly selecting K vectors from the training matrix $\mathbf{X}_i^{'}$, and normalize the vectors, forming $\mathbf{F}_i^{(0)}$. The dictionaries, $\mathbf{F}_i,\ i = 1, 2 \dots C$, are trained using the new training matrices, $\mathbf{X}_i^{'},\ i = 1, 2 \dots C$. When using the dictionaries for classification of textures, the same vector augmentation has to be done on the feature vectors $\mathbf{y}_l$ before classification: $\mathbf{y}_l^{'} = \begin{bmatrix} z \\ \cdots \\ \mathbf{y}_l \end{bmatrix}$. As stated, augmenting the vectors ensures that the dictionaries respond differently to dark and bright regions, and this can be important in some applications where the textures are known to have different average gray level. The present application is texture classification of mammograms, and it is known that the average gray level value is higher for most lesions than for normal tissue. This is an important feature difference between these two textures (lesions and normal tissue), and should be exploited using augmented vectors in training as well as classification.

III. Classification of microcalcifications in mammograms using learned overcomplete dictionaries

This chapter will start by looking at texture classification using FTCM as a first part of a scheme for detection of microcalcifications. It will be seen, by some case studies, that there is an obvious potential in using the FTCM for this purpose, but also that a significant amount of false positives are resulting from connective tissue lines, etc. Thereafter the use of FTCM on regions of interest (ROI) is considered as a last part of a scheme for detection of microcalcifications. The obtained results on ROI experiments are very good.

All experiments in this section are performed on direct digital mammograms supplied by the Breast Diagnostic Center of Stavanger University Hospital (SUS). The images are originally of size 2294×1914 with a resolution of 100 microns (0.1 mm/pixel), 14 bits pr. pixel, and recorded on a GE Senograph DS as a part of the daily clinical work at SUS. The microcalcifications can be very small (0.1 - 1 mm in diameter) thus some microcalcifications only cover a couple of pixels. To be able to detect these small microcalcifications the images *cannot* be downsampled. A binary image containing the breast region, $Ib_{breastregion}$, is easily found for digital images by simple thresholding (the thresholding can be followed by an erode operation to remove the skinline area). This can reduce the image size to some extent, depending on the breast size, by shrinking it to a bounding box around the breast region. However the images will still be large.

The main feature that can be exploited to distinguish microcalcifications from normal/background tissue is the sharp edges of the microcalcifications[1]. The brightness of microcalcifications can, however, vary a great deal. Thus vector augmentation (Section II-D) is not used in the microcalcification experiments.

A. FTCM as a first part of a detection of microcalcifications scheme

Depicted in Figure 1 is a block diagram of the scheme used in this section, with FTCM used for detection of microcalcifications.

The density of the breast can vary a great deal from patient to patient, but also inside the breast, as a function of the thickness of the breast. The reduced tissue thickness in the uncompressed periphery of the breast causes a gray level decay due to the intensity increase of X-rays. When searching for microcalcifications these density variations are not relevant (as opposed to when searching for masses). They can, however, complicate the process of detecting microcalcifications, therefore a preprocessing scheme to suppress the density variations is implemented. *The preprocessing is used in both training and classification/testing.*

The main part of the preprocessing is a morphological operation called top-hat (sometimes opening top-hat or white top-hat):

$$I_{tophat} = I \hat{\circ} S = I - (I \circ S), \quad (11)$$

where I is the image and S a structure element. $I \circ S$ is the opening of image I by structure element S, and is defined as an erosion followed by a dilation [20]. The structure element is generally small compared to the image. The top-hat operation should keep all details in the image but remove the main

[1]Note that a simple edge filter will give far to many false detections

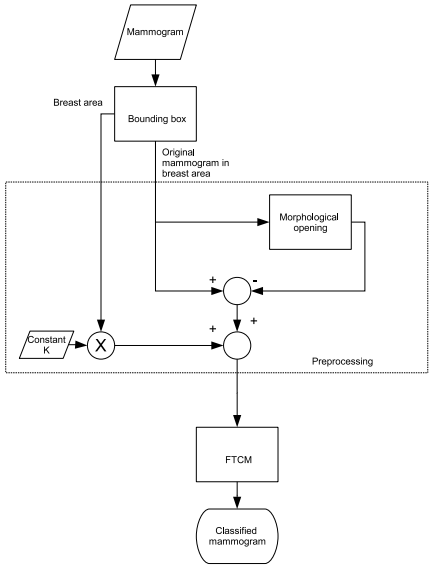


Fig. 1. Scheme for texture classification to detect microcalcifications. The dotted box surrounds the preprocessing eliminating variable breast densities.

density variations in the breast. Therefore the main density variations in the breast should be captured by the opening operation, but *not* the smaller details. Consequently a relatively large structure element is chosen: A circular element with radius $D = 50$. After the top-hat transform the values in the breast region are shifted back to a typical breast-region value, but now by adding a constant, K, to the entire breast region:

$$I_{pre} = I_{tophat} + K \cap Ib_{breastregion} \tag{12}$$

1) Training: The problem is formulated as a two-class problem: *microcalcification* and *other*, thus two dictionaries are needed; F_1 that represent the class *microcalcification* well, and F_2 that represent the class *other* well. These two dictionaries are learned using the dictionary learning method in Section II-B. F_1 is learned using a set of training vectors generated from areas in a number of preprocessed training mammograms containing microcalcifications, and F_2 is learned using a set of training vectors generated from areas in in a number of preprocessed training mammograms *not* containing microcalcifications. The training vectors are made by a predefined neighborhood centered around every pixel in the preprocessed training areas The ordering of the pixels are of no relevance as long as the same ordering (and same neighborhood) is used in training as well as in testing/classification. Different neighborhood sizes and shapes was briefly tested before a simple 5×5 pixel neighborhood was chosen for the experiments.

2) Testing: Following the preprocessing step in testing/classification, every single pixel in the preprocessed image, $I_{pre}(x_i, y_j)$, is made into a feature vector by a predefined neighborhood centered around the pixel at position (x_i, y_j). The neighborhood must be the same as was used in training the dictionaries.

The feature vectors, one for every single pixel in the texture image, are approximated using $\mathbf{F}_i$ resulting in the approximation error image $\mathbf{R}_i$ for $i = 1, 2$. $\mathbf{R}_i$ is smoothed using smoothing method 1, **Sm1**, explained in Section II-C, resulting in $\mathbf{Rs}_i$. The parameters used for the 2D gaussian filter are $a = 5$ and $\sigma = 8$. Finally classification is done by choosing the class with the lowest value at the corresponding position in $\mathbf{Rs}_i$.

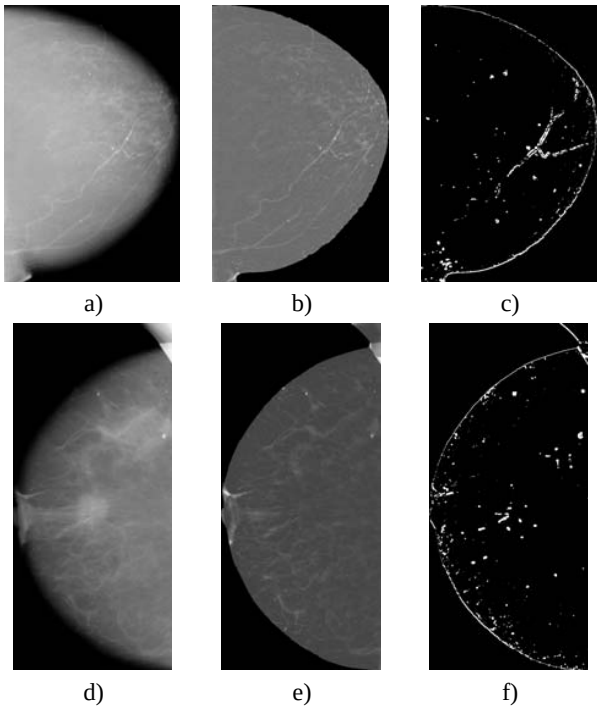
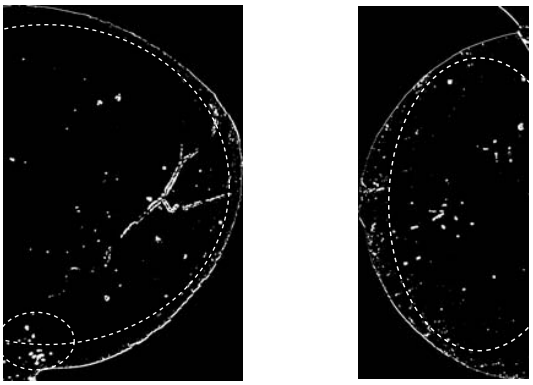


Fig. 2. a) and d) Example image 1 and 2 respectively, b) and e) corresponding images after preprocessing to remove density variations in the breast tissue, c) and f) corresponding results after running FTCM. Contrast is adjusted for visibility in image a,b,d,e.

3) Results: Two typical examples images *not* included in the training set are depicted in Figure 2. a) and d) shows the two original images, example image 1 and 2 respectively. In b) and e) the corresponding preprocessed images, with removed density variation in the breast, are depicted. These four images have all been adjusted by increasing the contrast for visualization. The results after classification using FTCM on the images in b) and e) are seen in c) and f). All calcifications in the breasts are correctly classified as calcifications. In addition there are, unfortunately, some areas that are wrongly classified as calcifications. These are mainly on, and close to, the edges of the breasts. In Figure 3 this is illustrated by some added ellipses on the figures. Roughly speaking the areas classified as calcifications *inside* the ellipses are correct, whereas the areas (all small) classified as calcifications *outside* the ellipses are wrong. There are some wrongly classified pixels inside the ”core” breast area (inside the ellipses) as well, but these are in general very small objects. The calcifications inside the blood vessel seen in Figure 2 a) and b) (appearance of a railway track) are also classified as calcifications by the FTCM. This can hardly be considered incorrect by the classification



a) b)

Fig. 3. Results for example images 1 and 2, as seen in Figure 2. The areas inside the ellipses are, roughly speaking, classified correctly. The areas outside the ellipses have a number of small areas wrongly classified as microcalcifications.

scheme. However this type of calcifications are never malign and thus they should not be marked as a calcification in a CAD system.

Visual inspection of these, and similar examples indicates that FTCM has a very good potential in classifying microcalcifications in breast tissue, but also that we get misclassifications in the boundary areas of the breast. In the interior of the breasts, the classification results are very good. More advanced preprocessing, or using more than two classes in FTCM, might reduce the misclassifications in the boundary areas to some extent. However, they indicate that FTCM is not necessarily well suited as a primary detection scheme. Another issue is that doing FTCM on the entire image is computationally expensive. In the next section we have used FTCM as a *last* part of a scheme for detection/classification of microcalcifications.

B. FTCM used on Regions of interest (ROI) as a last part of a scheme for detection of microcalcifications

Two important texture-features of the cluster of microcalcifications are the brightness of the calcifications as well as the the relatively sharp edges, i.e. high frequency information. The FTCM exploits both these features when classifying microcalcifications as opposed to background tissue. Unfortunately similar features characterize tissue connectivity lines and calcifications inside vessels as well. This gives rise to a number of ”false detections” as seen in the previous section, especially in the edge area or due to calcifications inside blood vessels. In addition, FTCM is computationally expensive when employed for finding a sparse approximation to the feature vector of every single pixel in the entire image.

Most techniques has the problem of false detections when they are tuned not to loose any true clustered calcifications, thus many systems are divided in two: a first detection step and a second step for classification of true or false detections. The results in the previous section motivates for trying FTCM as such a second classification step of suspicious areas found by ”a first step” feature extraction technique. Thus region of interests (ROI) can be found using some sort of (presumptive less computational expensive)feature extraction techniques, for example optimal filtering [21]. In the following, ROI’s are manually chosen to test FTCM on areas with and without single and clustered microcalcifications. The experiments show that using FTCM on an ROI can be very helpful in deciding whether the ROI contain a true cluster of microcalcifications or a false detection.

After preprocessing four different images as described in the previous section, three example segments of size 200×200 pixels, i.e. 20×20 mm, from each image are manually extracted to simulate ROI’s. The segments are selected so that some contain cluster of microcalcifications, some single calcifications, and some no calcifications to simulate true and false detections. The FTCM is tested on the chosen segments with smoothing method 1 and no vector augmentation.

In the Figures 4 and 5 the segments and the corresponding FTCM results are depicted. The ROI’s are plotted with maximum contrast for visualization purpose. The first column of Figure 4 shows three ROI’s from example image 1 (the same image as the top row of Figure 2). All these ROI’s contain clusters of microcalcifications. The second column shows three ROI’s from example image 2 (the same image as the bottom row of Figure 2). The top and middle ROI includes clusters of microcalcifications and the bottom ROI contains one large calcification. The two columns of Figure 5 show the ROI from two different images, image 3 and 4 respectively. The topmost ROI for image 3 and 4 both contain a small cluster of microcalcifications, the middle ROI’s contain one single calcification (small) each, and the bottommost ROI’s contain no calcifications for either of these images. The three first columns (image 1,2 and 3) all show very good classification results. The areas classified as calcifications are somewhat dilated because of the smoothing filter, but this introduces no significant problem. The results for image 4, however, are included to show the problems that might occur in some images. This image has relatively strong/sharp connective tissue lines in the background tissue. Since all the ROI’s in this figure are plotted with maximum contrast, the difference between the images might not seem to be significant. However the connective tissue lines in image 4 are actually more pronounced. As seen from the figure, some of these lines are misclassified as calcifications.

The overall classification potential seems to be very good. No calcifications are missed and relatively few pixels are wrongly classified. From a texture classification point of view, even the results for image 4 are not bad. The number of wrongly classified pixels is approx. 1.3 % and 1.1 % for the top and the middle ROI’s, and only 0.1 % for the bottom ROI. Note that when calculating these numbers the dilation of the correctly classified areas, due to the smoothing filter, is not taken into account. This does not affect the bottommost ROI, but the two others have a somewhat higher ratio of wrongly classified pixels. Since the microcalcifications cover a much smaller area than the background tissue the number of misclassified pixels is not not necessarily a good error

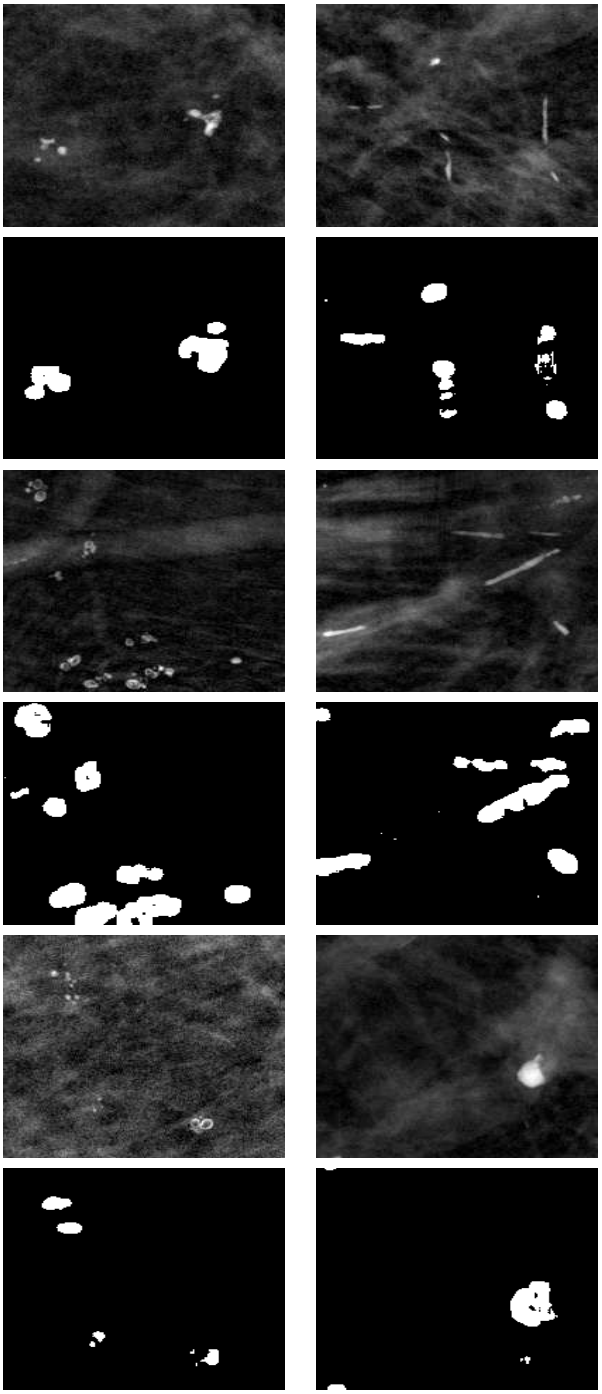


Fig. 4. ROI's (20×20 mm) from *preprocessed* example images with corresponding FTCM results shown beneath each depicted ROI. *First column*: three ROI's from example image 1, all containing clusters of microcalcifications. *Second column*: three ROI's from example image 2. Top and middle ROI's: clusters of microcalcifications, bottom ROI: single calcification (large).

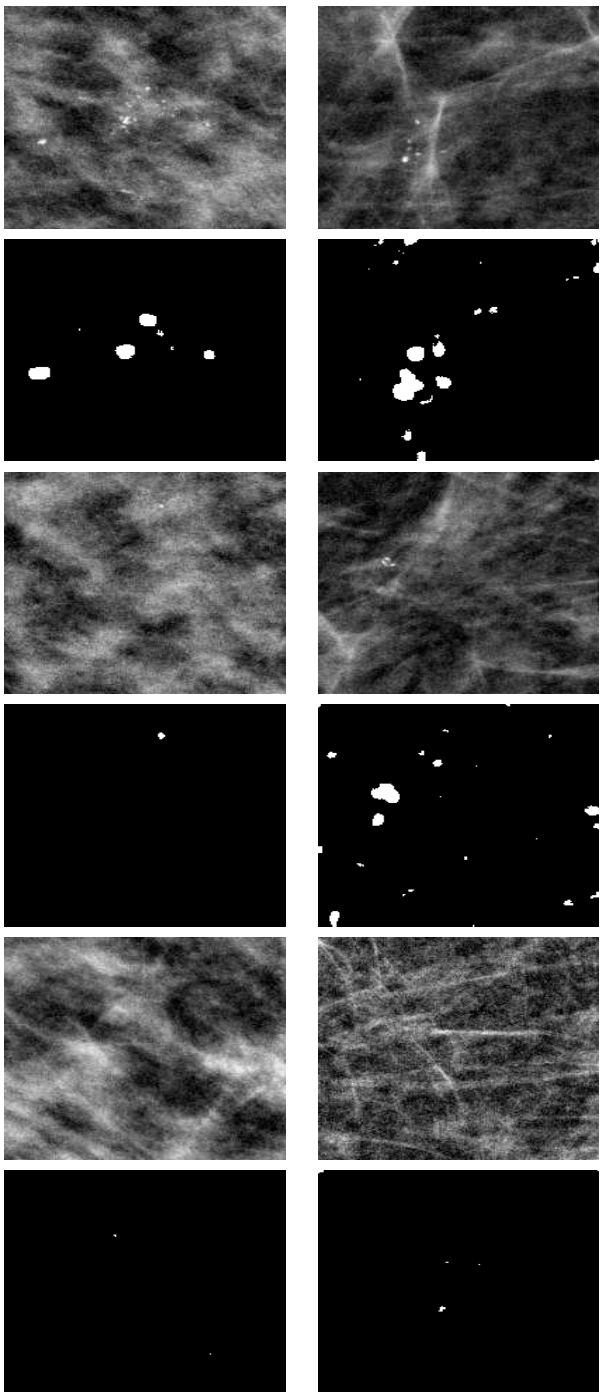


Fig. 5. ROI's (20×20 mm) from *preprocessed* example images with corresponding FTCM results shown beneath each depicted ROI. *First and second column*: three ROI's from example image 3 and 4 respectively. Top row: cluster of microcalcifications, middle row: single calcification (small), bottom row: no calcifications.

measure.

The experiments in Figure 4 and 5 are shown to see, in detail, the potential of using FTCM as the last part of a detection of clustered microcalcifications scheme. However, a reasonable error measure for detection of abnormalities in mammograms is the percentage of true detections and the number of false detections pr. image. To obtain such results our complete detection system, MammoScan μCaD must be tested. In the following we report the results of such experiments but remember that the FTCM is here just a part of a larger system. The scope of this paper is to investigate FTCM on a real world application, thus the other part of the detection system is just briefly mentioned.

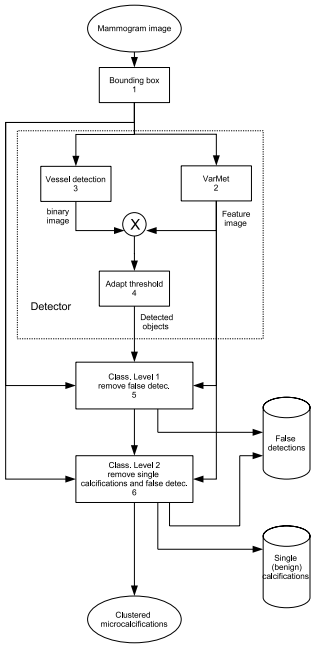


Fig. 6. Overview of the MammoScan μCaD system. The Classifier level 2 (last part) is based on FTCM.

In the complete CAD system for clustered microcalcifications, MammoScan μCaD, the FTCM is used as the last part of the system [22], where the first feature extraction is done by the VarMet method of Engan and Gulsrud [23]. An overview of the MammoScan μCaD system can be seen in Figure 6. The detector inside the dotted box detects suspicious areas. Most of the obviously false detections are removed by a series of simple tests in classifier level 1, including a test that removes connective tissue lines - the problem of image 4 in Figure 5. The main purpose of classifier level 2 is to separate single (and thus benign) calcifications from clusters of microcalcifications (benign or malign), but some remaining false detections are removed here as well. The FTCM works very well as the last step in the MammoScan μCaD scheme. Table I shows some results from a test performed at the Breast Diagnostic Center at SUS, where the test cases all contained microcalcifications. The test was performed on images from 51 selected patients, with in average approx. 3 images pr. patient. The number of images pr. patient varies from 1 to 8, giving a total of 155 images. All of the patients have μCa in some form. Some of the patients have only spread out/single benign μCa (17), others have clusters of μCa (34). Of the latter type some have been tested (cytological or histological) to be malignant (12), the rest is benign (22). Some patients have more than one cluster (malignant or benign), and the clusters can be visible on several of the images since there are often images from different views of the breast. Since these data are from clinical investigation there are often images zoomed in on the region of interest, however these type of images are *not* a part of this test. Our 155 images are all images of a complete breast, but from different views (cranio caudal, medio lateral oblique). The results from the MammoScan μCaD scheme are evaluated by a radiologist at SUS thus the detections are classified as clusters of microcalcifications (true detections), single microcalcifications (and these are not wanted in a complete CAD system) and neither (false detections). The number of "unwanted detections" in the table includes single calcifications as well as false detections.

These results are comparable to other reported test results [24], [25], [26], [27]. More information on MammoScan μCaD, and other test results, can be seen in [22].

True detections (image based) 84 %		True detections (case based) 94 %
Unwanted detections pr. image 1.1		False detections pr. image 0.3

TABLE I
RESULTS AFTER MAMMOSCAN muCAD TEST PERFORMED AT THE BREAST DIAGNOSTIC CENTER AT SUS.

IV. CLASSIFICATION OF SOFT TISSUE LESIONS/MASSES IN MAMMOGRAMS USING LEARNED OVERCOMPLETE DICTIONARIES

All experiments in this chapter are performed on digitized mammograms from the MIAS database provided by the Mammographic Image Analysis Society (MIAS) in the UK [28]. The images from this database have a resolution of 50 microns (0.05 mm/pixel), 8 bits pr. pixel. Four training mammograms were used for learning the dictionaries used in the classification, and 14 different mammograms where used for testing. Each mammogram contains at least one circumscribed lesion. The lesions are relatively large and this makes it possible to downsample the images without loosing any lesions. The mammograms were downsampled to $1/16 \times 1/16$ of their original size before classification, reducing the computational load significantly.

The proposed method segments/classifies the input mammogram into suspicious and non-suspicious regions (i.e. normal breast tissue) using two learned dictionaries of dimension $(N+1) \times K$, and was first tried in [29]. In [30] we used the dictionary classification on already segmented regions, segmented by the watershed transform. The lesion regions in the training images are found by a combination of watershed algorithm and visual inspection. A large number, L,

of overlapping image blocks of dimension $n \times n$ from the correct lesion regions are reshaped into training vectors of dimension $N = n \times n$ and collectively used as an $N \times L$ initial training matrix $\mathbf{X}_1$ for the lesion class dictionary. Since the breast tissue as seen in mammograms often have a dominant direction that varies from image to image, all image blocks are rotated $90°$, $180°$, and $270°$, prior to reshaping them into training vectors. The unrotated blocks are used as well. Ending up with dictionaries with directional qualities is thus avoided. In addition to blocks from the interior of each region, image blocks extending a few pixels outside the region boundaries of the true lesions are used. The training vectors for the normal tissue class dictionary, $\mathbf{X}_2$, are reshaped image blocks from normal tissue regions in the same four mammograms. Since an important feature of the lesions is that they are normally brighter than the surrounding tissue, vector augmentation, as described in Section II-D, is used in all experiments in this section in both training and testing. The added element, z, has the same value in training and testing, and the final training matrices become $\mathbf{X}_i^{'}$ as in Equation 10.

In both training and testing the same sparsity constraint, S, is used. That means that the (test/training) vector is represented using S vectors from the dictionary. In testing the dictionaries, both smoothing methods described in Section II-C are tried, and the results are compared. In Figure 7 three test images are depicted in the first column, together with classification results in the other columns. The second and third column show results after smoothing method **Sm1**, i.e. the residual images are smoothed using a Gaussian filter before a simple pixel by pixel classifier. Results using two different values of the standard deviation of the Gaussian filter are shown. Small σ will lead to more areas that are misclassified as lesions, i.e. false positives (FP). Too large σ might result in lesions being rejected. The last row in Figure 7 shows an image with one normal sized and one small lesion. σ=1.5 keeps both true detections, but there are a number of FP (5-6). Enlarging σ to 3 removes most of the misclassified areas (FP), but the small lesion is unfortunately also removed. The forth and fifth column shows results after smoothing method **Sm2** was applied with two different sizes of the structure element. The structure element used is circular with radius $R = 4$ and 6 pixels respectively. With a small radius the results show too many FP. With a larger radius some of these FP are removed but small lesions can be discarded as well. This is seen on the last row of the figure, where the small lesion is lost using the larger radius.

13 of the 14 test mammograms contain one lesion each while one mammogram contains two lesions. 14 of the 15 lesions are detected using **Sm2**, $n = 9$, $z = 250$, $S = 3$, $K = 164$, and $R = 4$. A relatively good segmentation quality was achieved. The lesion not correctly detected had low contrast and was located in very dense tissue, see Figure 8. There is however a number of FP present in all the images. A FP reduction technique should be implemented before comparing to reported results like Te Brake and Karssemeijer [31] (75 % sensitivity at 1 FP/image) and Kobatake [32] (90.5 % sensitivity at 1.3 FP/image). The scope of this paper, however, is the texture classification technique and its potential.

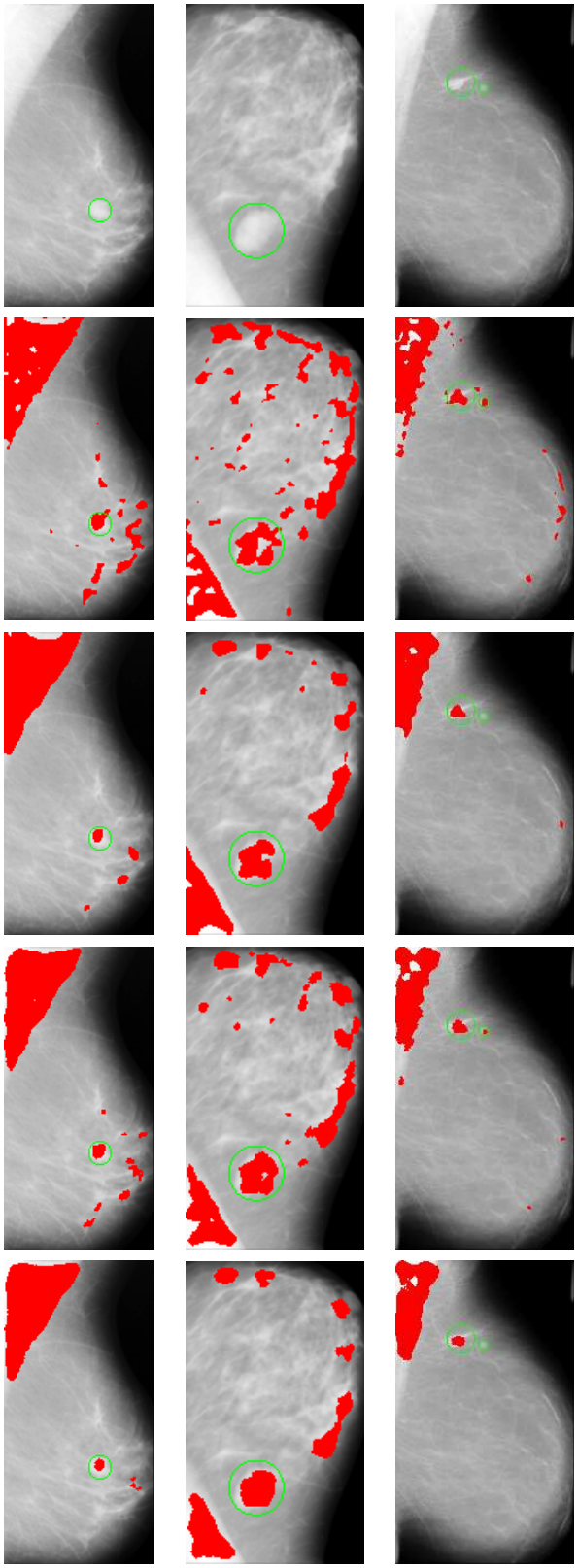
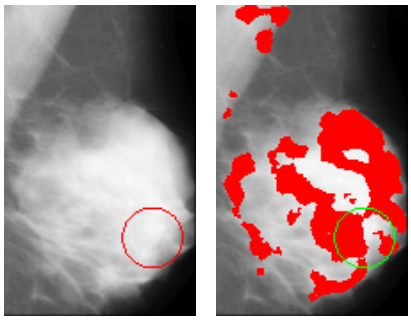


Fig. 7. Test images for lesion detection, and corresponding results with different smoothing. *First row*: downsampled images with truth marking. *Second and third row*: Smoothing by **Sm1**, $\sigma = 1.5$ and 3 respectively. *Fourth and fifth row*: Smoothing by **Sm2**, radius of structure element/filter $R = 4$ and 6 respectively.



(a) (b)

Fig. 8. The mammogram *mdb010rm* for which the lesion is not detected. (a): Mammogram and truth circle. (b): Mammogram and detected regions. **Sm2**, $K = 164$, $n = 9$, $R = 4$, $S = 3$, and $z = 250$.

V. Discussion and Conclusion

The recently presented Frame Texture Classification Method (FTCM) of Skretting et al. [2] has been tested on the real world application *detection of abnormalities in mammograms*. FTCM is tested for both detection of clustered microcalcifications and lesions in mammograms. Two different smoothing techniques and a vector augmentation technique useful for the current application are implemented as extensions to the original FTCM.

The results for detection of microcalcifications are promising, but there is a problem of too many false detections, especially in the uncompressed breast area at the edges. Using FTCM as a last part of a complete detection of microcalcification scheme, as done in MammoScan μCaD [22], performs very well. Using FTCM on lesion detection shows interesting results, but not as promising as for the microcalcifications. We experience a number of False Positives (FP) present in all the images. An FP reduction technique should be developed, perhaps one based on another set of learned dictionaries. For mammograms of dense breasts there are at present a potential for improvement, as with most other methods. The method has only been tested on a limited number of mammograms, and needs testing on a larger database to provide more reliable sensitivity/specificity data.

References

[1] M. Tuceryan and A. K. Jain, "Texture analysis," in *Handbook of Pattern Recognition and Computer Vision*, C. H. Chen, L. F. Pau, and P. S. P. Wang, Eds. Singapore: World Scientific Publishing Co, 1998, ch. 2.1, pp. 207–248.

[2] K. Skretting and J. H. Husøy, "Texture classification using sparse frame-based representations," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. Article ID 52 561, 11 pages, 2006, doi:10.1155/ASP/2006/52561.

[3] www.kreftregisteret.no.

[4] H. P. Chan, K. Doi, C. Vyborny, R. Scmidt, C. Metz, K. L. Lam, T. Ogura, Y. Wu, and H. Macmahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms. the potential of computer-aided diagnosis," *Investigative Radiology*, vol. 25, no. 10, pp. 1102–1110, 1990.

[5] H. P. Chany, B. Sahiner, and N. P. et al., "Computerized classification of malignant and benign microcalcifications on mammograms: texture analysis using an artificial neural network," *Phys. Med. Biol.*, vol. 42, no. 3, pp. 549–567, Jan. 1997.

[6] N. R. Mudigonda, R. M. Rangayyan, and J. E. L. Desautels, "Gradient and texture analysis for the classification of mammographic masses," *IEEE Trans. Medical Imaging*, vol. 19, no. 10, pp. 1032–1043, Oct. 2000.

[7] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, USA: Kluwer Academic Publishers, 1992.

[8] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM journal on computing*, vol. 24, pp. 227–234, Apr. 1995.

[9] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397–3415, Dec. 1993.

[10] M. Gharavi-Alkhansari and T. S. Huang, "A fast orthogonal matching pursuit algorithm," in *Int. Conf. on Acoust. Speech and Signal Proc.*, Seattle, U.S.A, May 1998, pp. 1389–1392.

[11] K. Skretting and J. H. Husøy, "Partial search vector selection for sparse signal representation," in *NORSIG-03*, Bergen, Norway, Oct. 2003, available at http://www.ux.his.no/˜karlsk/.

[12] S. S. Chen, "Basis pursuit," Ph.D. dissertation, Stanford University, Nov. 1995.

[13] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Processing*, vol. 47, pp. 187–200, Jan. 1999.

[14] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado, "Subset selection in noise based on diversity measure minimization," *IEEE Trans. Signal Processing*, vol. 51, no.3, pp. 760–770, Mar. 2003.

[15] K. Engan, S. O. Aase, and J. H. Husøy, "Method of optimal directions for frame design," in *Proc. ICASSP '99*, Phoenix, USA, Mar. 1999, pp. 2443–2446.

[16] ——, "Multi-frame compression: Theory and design," *Signal Processing*, vol. 80, pp. 2121–2140, Oct. 2000.

[17] K. Engan, K. Skretting, and J. Husøy, "A family of iterative LS-based dictionary learning algorithms, ILS-DLA, for sparse signal representation," *Digital Signal Processing, Elsevier*, vol. 17, no. 1, pp. 32–49, 2007, doi:10.1016/j.dsp.2006.02.002.

[18] K. Skretting, "Sparse signal representation using overlapping frames," Ph.D. dissertation, Stavanger University College/NTNU, Stavanger, Norway, 2002, available at http://www.ux.his.no/˜karlsk/.

[19] K. Skretting and J. Husøy, "Frame based texture classification by considering various spatial neighborhoods," in *Proc. NORSIG '05*, Stavanger, Norway, Sept. 2005.

[20] E. R. Dougherty and R. A. Lotufo, *Hands-on Morphological Image Processing*. Washington, USA: SPIE PRESS, 2003.

[21] T. O. Gulsrud and J. H. Husøy, "Optimal filter based detection of microcalcifications," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 11, pp. 1272–1280, 2001.

[22] K. Engan, T. O. Gulsrud, K. F. Fretheim, B. Iversen, and L. Eriksen, "A complete computer aided detection (CAD) system for microcalcifications in mammograms - MammoScan μCaD," *Submitted for journal publication*.

[23] K. Engan and T. O. Gulsrud, "VarMet - a method for detection of image singularities with application to mammography," *WSEAS Transactions on Signal Processing*, vol. 2, no. 9, pp. 1222–1229, Sept. 2006.

[24] G. Horvath, J. Valyon, G. Strausz, M. Pataki, L. Sragner, L. Lasztovicza, and N. Szekely, "Intelligent advisory system for screening mammography," in *Proceedings of IEEE Instrumentation and Measurement Technology Conference, IMTC 04*, vol. 3, May 2004, pp. 2071–2076.

[25] I. El-Naqa, Y. Yang, M. W. N.P., Galatsanos, and R. Nishikawa, "A support vector machine approach for detection of microcalcification," *IEEE Trans. Medical Imaging*, vol. 21, no. 12, pp. 1552–1563, 2002.

[26] W. J. H. Veldkamp and N. Karssemeijer, "An improved method for detection of microcalcification clusters in digital mammograms," in *Proc. of SPIE International Symposium Medical Imaging, Image Processing 1999*, vol. 3661, San Diego, Ca, USA, May 1999, pp. 512–522.

[27] www.r2tech.com.

[28] *www.wiau.man.ac.uk/services/MIAS/MIAScom.html*.

[29] J. Herredsvela, K. Engan, T. O. Gulsrud, and K. Skretting, "Detection of lesions in mammograms using learned dictionaries and sparse representations," in *Proc. of EMBEC05*, Prague, Czech Republic, Nov. 2005.

[30] ——, "Detection of masses in mammograms by watershed segmentation and sparse representasions using learned dictionaries," in *Proc. of NORSIG 2005*, Stavanger, Norway, Sept. 2005.

[31] G. M. te Brake and N. Karssemeijer, "Single and multiscale detection of masses in digital mammograms," *IEEE Trans. Medical Imaging*, vol. 18, no. 7, pp. 628–639, July 1999.

[32] H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, "Computerized detection of malignant tumors in digital mammograms," *IEEE Trans. Medical Imaging*, vol. 18, no. 5, pp. 369–378, May 1999.

Kjersti Engan (M'01) was born in Bergen, Norway, in 1971. She received the Ing. (B.E.) degree in electrical engineering from Bergen University College in 1994 and the Siv.Ing. (M.S.) and Dr.Ing. (Ph.D.) degrees in 1996 and 2000 respectively, both in electrical engineering from the University of Stavanger, Stavanger, Norway.

She is an associate professor with the department of Electrical and Computer Engineering at University of Stavanger, Norway. Her research interests include signal and image representation and compression, sparse signal representation, image analysis, and medical signal and image segmentation and classification.

Dr. Engan is a member of NORSIG (Norwegian Signal Processing Society) and currently on the board of NOBIM (Norwegian Image and Pattern Recognition Society).

Karl Skretting was born in Nærbø, Norway in 1962. He started to study signal processing at University of Stavanger (UiS), Norway in 1996. He received his M.Sc. and Ph.D. degree from the Electrical and Computer Engineering Department at UiS in 1998 and 2002 respectively.

Currently, he is employed as an associate professor with the department of Electrical and Computer Engineering at University of Stavanger, Norway, and he is a part of the Cybernetics Group. The topics of interest are signal representation and classification, data compression, and system identification.

Thor Ole Gulsrud was born in Stavanger, Norway, in 1965. He received the Ing. (B.E.) degree from the University of Stavanger, Stavanger, Norway, in 1986 and the Siv.Ing. (M.S.) degree from the University of Strathclyde, Glasgow, Scotland, in 1991, both in electrical engineering. In 2001 Dr. Gulsrud received his Ph.D. degree in digital image processing from Aalborg University, Aalborg, Denmark.

In the period 1999-2006 he was an associate professor with the department of Electrical and Computer Engineering at the University of Stavanger, Norway. He is currently a technical product manager at Roxar Flow Measurement AS, Stavanger, Norway. His research interests include image analysis, medical image segmentation and classification, and ultrasonic techniques for fluids characterization.

Dr. Gulsrud is a member of NOBIM (Norwegian Image and Pattern Recognition Society).