11

Evaluation of Mammographic Segmentation and Risk Classification Based on Tabár Tissue Modelling

Wenda He[1], Erika Denton[2] and Reyer Zwiggelaar[1]
[1] *Department of Computer Science, Aberystwyth University, SY23 3DB*
[2] *Department of Radiology, Norfolk & Norwich University Hospital, NR4 7UY*
United Kingdom

1. Introduction

This chapter evaluates recently proposed computer vision methodologies for mammographic segmentation and risk classification based on Tabár tissue modelling (Tabár et al., 2004). The overall objective is to bridge the gap between the theoretical work and their clinical usefulness, with respect to mammographic risk assessment for early breast cancer detection.

In the computer vision literature, there are two lines of scientific investigation into automated mammographic risk assessment. The first focuses on the correlation between mammographic risk and the parenchymal; and the second focuses on the correlation with the overall breast density. In addition, the sensitivity of mammography is significantly reduced by increased breast density which can mask some tumours due to dense fibroglandular tissue (Sickles, 2007). Whilst computer aided mammographic density estimation has been widely investigated as a two-class (i.e. dense and non-dense breast tissue) problem, Tabár breast components composition estimation as a four-class (i.e. nodular, linear, homogeneous densities and radiolucent) problem seems to be overlooked in computer aided mammography. Despite the current standard which is favouring breast density based risk estimation, clinical evidence has shown that the mammographic parenchymal appearances (*e.g.* distribution of various breast tissues) may provide more information than percent density alone (Astley, 2004).

2. Background

Breast density varies from one woman to another. These variations are related to the risk of developing breast cancer; the denser the breast appearance the higher the risk. Mammographic breast density also has a strong association with mammographic screening sensitivity (Couto et al., 2001). Women with high risk density patterns are advised to undergo screening mammography more frequently, and may require additional views (Ciatto et al., 2004).

2.1 Mammographic risk assessment

To assess breast cancer risk, different methods have been developed to measure mammographic density patterns. Two most frequently used classification methods for measuring breast density include Wolfe's mammographic parenchymal patterns (Wolfe, 1976a), and the percentage of the breast densities (Brisson et al., 2003). In Wolfe's scheme, breast cancer risk (i.e. N1, P1, P2 and DY) is assessed based on four mammographic features, including the extent of dense breast tissue in the mammogram, and characteristics of densities (*e.g.* shape and texture). Several other classification methods use parenchymal patterns, including BIRADS classification (i.e. BIRADS I-IV) proposed by the American College of Radiology (American College of Radiology, 2004), and Tabár's scheme (i.e. Tabár I-V) (see Section 2.2). Breast percentage density based scheme, such as Boyd's six-class categories (Boyd et al., 1995) assesses the extent of fibroglandular breast densities in the mammogram, without taking into account the various types of fibroglandular densities, and is expressed by the percentage of the breast showing dense tissue (Boyd et al., 2006).

Experts' subjective appraisal of mammograms can lead to inter and intra observer variability. Automated mammographic risk assessment may be beneficial in a clinical environment, because computerised processes can produce unbiased and consistent results as a means of aiding diagnosis.

2.2 Tabár tissue modelling

In mammographic risk assessment, Tabár *et al.* have proposed a mixture model of four mammographic building blocks, representing nodular, linear, homogeneous and radiolucent tissues (Tabár et al., 2004). These four building blocks compose the normal breast anatomy. Nodular densities mainly correspond to Terminal Ductal Lobular Units (TDLUs). Linear structures correspond to either ducts, fibrous tissue or blood vessels. Homogeneous structureless densities correspond to fibrous tissues which appearance could hide the underlying normal TDLUs and ducts, as well as their alterations due to hyperplastic breast changes. Radiolucent areas are related to adipose fatty tissues, which appear as dark areas in mammographic images (Tabár et al., 2004).

Tabár's tissue modelling is strongly influenced by Wolfe's original work, which divided mammograms, using parenchymal patterns, into four classes (Wolfe, 1976b). For Tabár's risk assessment model, the relative proportions of the tissues belonging to the four building blocks are used to subdivide mammograms into five risk classes. Patterns I to V indicate low to high mammographic risk. The relative composition of the four building blocks, nodular, linear, homogeneous and radiolucent (in that order) are as follows:

- Pattern I has composition [25%, 15%, 35%, 25%]. This pattern is the most common mammographic pattern in premenopausal woman (Tabár et al., 2004). All four building blocks are fairly equally distributed in this pattern. *However, with involution, pattern I will change to either pattern II or pattern III; on the other hand, hormone-replacement therapy often reverses the process of involution. Mammograms of pattern I show normal fibroglandular tissue with partial fatty replacement, and pathological changes can be easily perceived in pattern I, although these breasts can be radiologically "dense"* (Tabár et al., 2004). This pattern is considered the lowest mammographic risk.

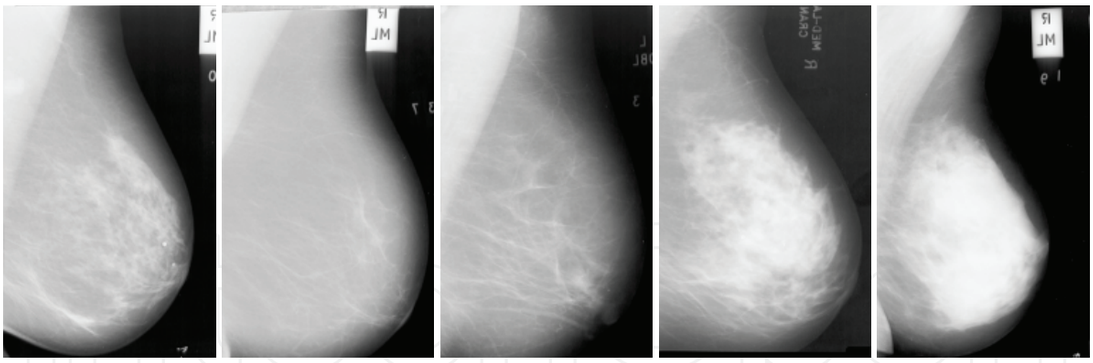


Fig. 1. Mammographic risk patterns with respect to Tabár's risk classification. From left to right showing patterns I-V, corresponding from low to high mammographic risk.

- Pattern II has composition [2%, 14%, 2%, 82%]. This pattern represents the end results of the process of involution, and is characterised by the over representation of radiolucent fatty tissue, which provides excellent background for radiologists to visualise any process of disease (Tabár et al., 2004). Therefore, the detection of abnormalities in mammograms of pattern II is relatively easier.
- Pattern III is similar in composition to pattern II. In this pattern the retroareolar prominent ducts are often associated with periductal fibrosis (Tabár et al., 2004). *Neither of these patterns (i.e. patterns II and III) has nodular densities or diffuse fibrosis; the over representation of radiolucent fatty tissue makes pathological lesions being detected relatively easier on the mammogram* (Tabár et al., 2004).
- Pattern IV has composition [49%, 19%, 15%, 17%]. This pattern is dominated by prominent nodular and linear densities. *Their presence often makes perception of pathological lesions difficult. Also this pattern appears to be resistant to the process of involution* (Tabár et al., 2004). This pattern is considered higher mammographic risk.
- Pattern V has composition [2%, 2%, 89%, 7%]. *This patten is dominated by extensive fibrosis. The overwhelming dominance of this homogeneous, structureless fibrous tissue limits the capabilities of mammography to demonstrate the details of normal anatomy and to reveal small pathological lesions. The radiology reports often state such a mammogram as "extensive fibrosis", which is the only demonstrable feature* (Tabár et al., 2004). This pattern is considered the highest mammographic risk.

Fig. 1 shows example mammographic images with respect to Tabár's five mammographic risk patterns.

3. Literature review

Automatic mammographic risk assessment is expected to play a significant role in the development of breast screening programs and computer-aided detection/diagnosis systems. As mentioned, strong evidence has indicated a correlation between mammographic parenchymal patterns and mammographic risk. This section reviews the existing representative methodologies used for mammographic risk assessment, and related techniques used in computer-aided mammography, with discussions of their merits and shortcomings.

3.1 Textons representation

Recent research has demonstrated the use of textons (Varma & Zisserman, 2004) based mammographic segmentation. Petroudi *et al.* (Petroudi et al., 2003) applied such a technique directly to mammographic images for risk classification. A total of 40 textons were generated via clustering over the filter response space; that is 10 textons for each of the four BIRADS categories. Mammograms were modelled as texton frequency histograms; and by comparing the texton distribution of a mammogram to the learned models using the χ^2 (chi-square) method, the mammographic risk was determined as the BIRADS risk category corresponding to the nearest neighbour model. The evaluation was conducted on a small Oxford database (not publicly available) (Evertsz et al., 2000), the exact agreement with the ground truth was 76% accuracy.

Gong *et al.* (Gong et al., 2006) used a similar statistical approach for mammographic segmentation and risk classification. Instead of using image filtering to obtain the feature vectors, the raw pixel intensities of a $N \times N$ square neighbourhood image patch around a pixel were taken and row reordered, to form a N^2 dimensional feature vector. Four mammographic risk classes were learned based on Wolfe's scheme (Wolfe, 1976b). The same processes used in (Petroudi et al., 2003) were employed for the texton generation, mapping, and risk classification. The Oxford database was again used in the evaluation, and an average classification accuracy of 87% was achieved for 43 randomly selected mammograms.

The statistical texture based classification method (Petroudi et al., 2003) was combined with the multi-vector Markov random fields based approach (Gong et al., 2006) in (Petroudi & Brady, 2006), to perform mammographic density segmentation using textural and structural information based on Wolfe's scheme. At the evaluation stage, 32 mammograms were selected from the Oxford database. The segmentation results show a strong correspondence between the texture variations and the radiologists perception of the different breast density patterns. Despite the use of a small dataset, the breast density segmentation is considered anatomically plausible from an expert radiologist's point of view. It is concluded that texton based statistical modelling is able to generalise the data, and overcome noise issues related to image acquisition and involution of the breast.

Muhimmah *et al.* (Muhimmah et al., 2007) applied a texton related technique for mammographic segmentation based on Tabár's tissue modelling; a texton selection strategy is incorporated using a combination of visual assessment (probability maps) and minimum spanning tree topological information. The initial number of textons (cluster centres) for each tissue type (i.e. nodular, linear, homogeneous and radiolucent) was empirically chosen to be 25. In the texton selection process, the (Euclidean) minimum spanning tree was used to indicate a topologically probable correct connectivity, in high dimensional spaces. Distinct textons (higher discriminative power) tend to be situated towards the outer edges of the tree; common texture, noise and intensity aspects tend to be modelled by textons in the central part of the tree. A visual assessment was employed to determine the appropriate textons, based on the distinctiveness of the textons spatial displacements. Subsequently, a model driven based mammographic segmentation was performed based on the selected textons. A similar approach was applied in (He et al., 2008) with an alternative texton selection process. Both studies have shown realistic segmentation results, but no risk classification experiments were conducted.

3.2 Co-occurrence matrix representation

Arnau *et al.* (Oliver et al., 2007) applied co-occurrence matrices as a means of extracting texture features directly from mammographic images for automatic classification of breast density. A total of 216 features were computed from the co-occurrence matrices and used at the classification stage. The classification accuracy achieved was 78% using a K-Nearest-Neighbour ($K = 3$) classifier over the complete set of the MIAS database (Suckling et al., 1994). Further improvement was made in (Oliver et al., 2008) based on the same methodology, where 10 additional morphological features for the fatty and dense tissue regions were calculated. At the classification stage a combined Bayesian classifier was developed based on the K-Nearest-Neighbour and ID3 Decision Tree; the combined classifier achieved 86% and 77% correct classification for the MIAS and DDSM (Heath et al., 1998) database, respectively. The investigation indicated that classification accuracies can be improved by increasing the combination of the orientations and step intervals of the co-occurrence matrices (as well as additional features). However, the dimensionality of the feature space was increased.

Zwiggelaar *et al.* (Zwiggelaar et al., 2003) introduced set-permutation-occurrence matrices (a set of co-occurrence matrices) as texture feature vectors for texture segmentation of mammographic density. The novel texture modelling approach incorporated spatial information implicitly. The segmentation results were used for automatic mammographic risk assessment. Further improvement was made in (Zwiggelaar & Denton, 2004), where a transportation algorithm (Hitchcock, 1941) was used to select the appropriate set of co-occurrence matrices for the segmentation of mammographic density. It should be noted that using the transportation algorithm can lead to a very computational intensive task (Andoni et al., 2008); therefore it may not be practical for use in a clinical environment.

3.3 Moments representation

As breast cancer develops/progresses, statistical information can be gathered with respect to shape related texture (*e.g.* various types of tumours), which may have significant implication for the staging of breast cancer. Inspired by medical diagnosis of malignant breast cancer, where mammogram masses are classified based on the presence of spicular lesions or diffuse stellate appearance (Tavassoly, 1992); Soares *et al.* (Soares et al., 1998) have shown that moment techniques can be used for shape related texture analysis in digitised mammograms. Specifically, moments were used as features for automated classification of masses in mammography (Soares et al., 1998). It is a logical approach to use moments as shape features for mass classification, because stellate tumours can be characterised as an irregular shape with radiating spicules. A similar technique was used in (Mencattini et al., 2007), and additional texture features (i.e. Haralick features (Haralick et al., 1973)) are incorporated for mass classification. Other types of moments, such as discrete orthogonal Chebyshev moments were employed in (Vyas & Priti, 2007) for malignancy texture classification in digitised mammograms. Eisa *et al.* (Eisa et al., 2009) combined moments with texture features to improve content-based image retrieval in mammography. For an image a set of geometric moments (spatial moments) of different orders are computed, as a global shape descriptor to describe different spatial characteristics of the image intensity distribution. The results indicated that the mixture of features is more discriminative than the texture or moments based features alone. Tucceryan (Tucceryan, 1994) demonstrated the efficiency of obtaining texture features by computing moments over local image regions. A set of

features with respect to specific texture properties of the image, can be used to derive a set of spatial moments from the second and third-order moments. The spatial moments based texture features are able to capture orientation sensitive texture properties, as well as more complicated textural properties. Substantial evidence has shown that moments based techniques are feasible for texture modelling in mammography. It is especially true when modelling Tabár's mammographic building blocks (i.e. nodular, linear, homogeneous and radiolucent) as different breast tissue types have texture patterns related to not only their periodicity but also shape features.

3.4 Signatures representation

In the computer vision literature, signature (2D histogram) techniques have been applied to texture analysis in mammography as a means of detection and recognition. Guliato *et al.* (Guliato et al., 2008) used signature of a turning angle function to encode features (*e.g.* roughness, complexity) that characterise the contours of masses for breast tumour classification, because malignant breast tumours and benign masses can be distinguished by their shape characteristics. Zwiggelaar *et al.* (Zwiggelaar, 2002; Zwiggelaar et al., 1999) investigated abnormality (*e.g.* the central mass of spiculated lesions) detection in mammography using scale-orientation signatures. Such signatures have been described as rich descriptors of the neighbourhood around each image pixel, and were used to label structures in images, to classify the pixels into linear structures, blob-like structures or background texture. Therefore, signature techniques are also capable of modelling shape related texture.

4. Materials

The Mammographic Image Analysis Society (MIAS) database consists of 322 mammographic images in mediolateral views. The mammographic films were digitised to 50 micron $\times$ 50 micron resolution with a Joyce-Loebl scanning microdensitometer, a device linear in the optical density range 0-3.2 and representing each pixel with an 8-bit word (Suckling et al., 1994). The database provides $\{x, y\}$ image-coordinates for the centre of a circle, its approximate radius (in pixels) that encloses the abnormalities, the information about their severities (i.e. benign and malignant), and the abnormality types (*e.g.* calcification, masses, architectural distortion and asymmetry). Ground truth for breast density is classed into three categories (i.e. fatty, glandular, and dense). The ground truth for mammographic risk classification with respect to Tabár and BIRADS was provided by an expert radiologist.

4.1 Mammographic image patches

In order to facilitate the texture modelling and training process, a set of randomly selected mammograms from the MIAS database were subsampled by an expert mammographic screening radiologist to obtain samples representing Tabár's mammographic building blocks, resulting in patches containing (199) nodular, (253) linear, (70) homogeneous and (121) radiolucent tissue examples (Muhimmah et al., 2007). The collection of patches (See Fig. 2 for examples) covers the various mammographic risk classifications. In addition, for 98 subsampled mammographic patches detailed manual annotations were provided by the same radiologist covering (21) nodular, (27) linear structure, (28) homogeneous and (22) radiolucent regions.

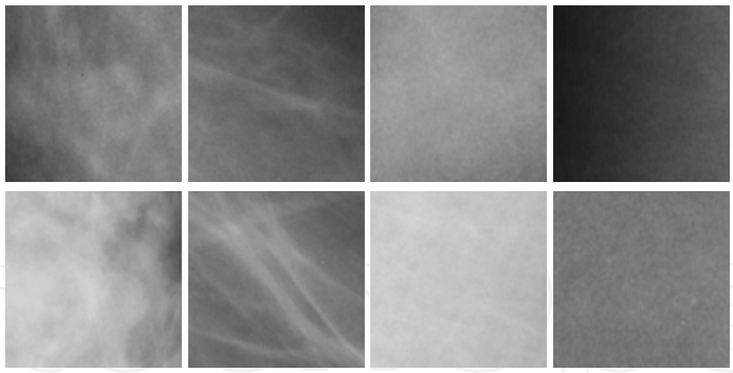


Fig. 2. Subsampled mammographic patches in the top and bottom row belong to the same tissue type. From left to right: nodular, linear, homogeneous, and radiolucent. It should be noted that intra class variation can be seen as large variation in intensity and texture patterns within the same tissue types (Tabár's mammographic building blocks).

5. Methodologies

With respect to mammographic risk assessment using computer vision techniques, there are a few approaches to mammographic segmentation and risk classification using Tabár's scheme, probably because BIRADS is currently more wildly used. However, in computer aided mammography, it is plausible to perform risk assessment based on Tabár's tissue modelling. This section describes two recently developed approaches (He et al., 2009; 2010) for mammographic segmentation and risk estimation based on modelling of various breast components (i.e. nodular, linear, homogeneous and radiolucent tissue) as seen in mammograms. To develop an automated, accurate and repeatable mammographic risk estimation approach based on Tabár's tissue modelling, the major challenge is to segment a given mammographic image according to Tabár's mammographic building blocks, so that the characteristic mixture of these building blocks can be determined, the mammographic risk can be estimated, and the relative proportions of different breast tissue patterns can be quantified. This process is illustrated in Fig. 3.

5.1 Moments based approach

Mammographic segmentation based on spatial moments and prior information (*e.g.* shape related texture features) of mammographic building blocks was investigated. The developed methodology is capable of modelling complex mammographic images and can deal with intra class variation and noise aspects.

5.1.1 Local moments calculation

Moments up to order four are computed within a small local window centred at each pixel (i, j) in an image. This number of moments is capable of capturing sufficient texture geometry information on a low dimensional feature. The coordinates of a local window, with size equal to $W \times W$ (with W being odd so that the pixel (i, j) is at the centre of the window) are normalised to be in the [-1, 1] range. The size of the local window was determined by estimating the position of the first side-band in Fourier space, using Fourier analysis on local patches (Gonzales & Woods, 1992). A series of investigations was conducted on parameter settings based on the full range of anatomical objects; the chosen sampling regions consisted

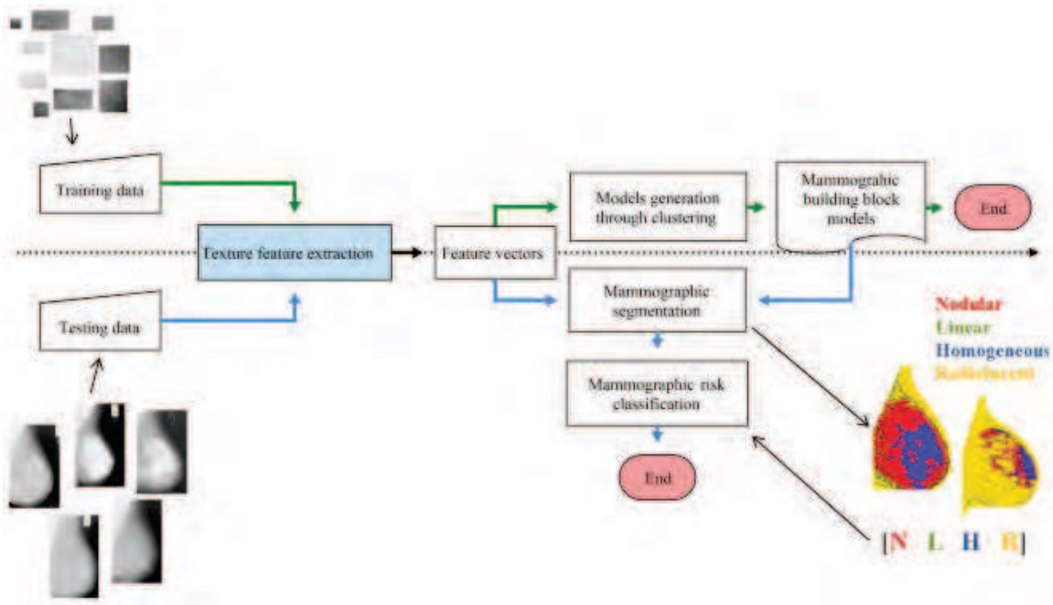


Fig. 3. A flowchart of mammographic segmentation and risk classification methodology. Note that mammographic patches are used at the training (modelling) stage, but full mammographic images are used at the testing (segmentation) stage.

of four square windows with $W = \{7, 13, 33, 63\}$, where the lower limit indicates the feature extraction of small structures (*e.g.* small diameter nodular tissue, thin linear structures), and the upper limit indicates the feature extraction of larger structures. For a pixel (m, n) which falls within the local window $f(m, n)$, the normalised coordinates (x_m, y_n) are defined as $x_m = (m - i)/(W/2)$ and $y_n = (n - j)/(W/2)$. The $(p + q)^{th}$ order moments within the window centred at pixel (i, j) are computed by a discrete sum approximation, which uses the normalised coordinates (x_m, y_n) and is defined as

$$m_{pq} = \sum_{-W/2}^{W/2} \sum_{-W/2}^{W/2} x_m^p \, y_n^q \, f(m, n). \tag{1}$$

The lower-order moments ($p + q \leq 1$) have well defined geometric interpretations. The higher-order moments ($p + q \geq 2$) give more detailed shape characteristics of the polygons (Tucceryan, 1994). The m_{00} moment is used to identify concentrated density (compact high-intensity) at a point, which can be interpreted as a centre of mass detector. The two first order moments, m_{10} and m_{01} are used to identify image pixels at which the brightness changes sharply or has a discontinuity, which can be interpreted as edge or contrast detectors, respectively. The second order moment m_{11} is used to identify image pixels brightness discontinuities at cross like junctions, which can be interpreted as a cross detector. Local moments are inherently integral-based features, which can reduce the effect of uncorrelated noise. The computation of local moments can be interpreted as a convolution of an image with a set of masks (Tucceryan, 1994), this has been illustrated in Fig. 4. The x and y coordinates of the centre of mass can be computed as $\bar{x} = m_{10}/m_{00}$ and $\bar{y} = m_{01}/m_{00}$ with respect to the normalised coordinates. The central moments are computed for each pixel in the local

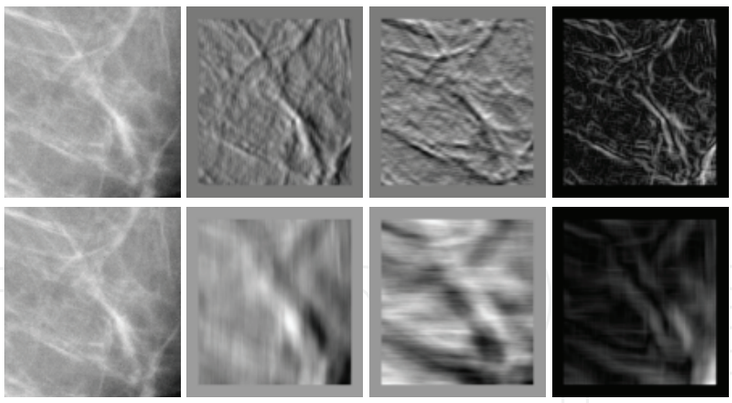


Fig. 4. In both rows, from left to right: original input image, m_{10} moment image, m_{01} moment image, and m_{11} moment image. The local window sizes used were 13 (top) and 63 (bottom).

window, and defined as

$$\mu_{pq} = \sum_{-W/2}^{W/2} \sum_{-W/2}^{W/2} (x_m - \bar{x})^p \, (y_n - \bar{y})^q \, f(m,n). \tag{2}$$

5.1.2 Deriving local image properties

Local image properties are derived from higher order moments (Awcock & Thomas, 1996; Eisa et al., 2009), and determined as

- Variance V_{xy} (spreading) around the centre of mass with respect to the x-axis and y-axis can be computed from normalised 2^{nd} order moments and defined as

$$V_{xy} = \frac{\mu_{11}}{\mu_{00}}, \qquad V_x = \frac{\mu_{20}}{\mu_{00}} \qquad \text{and} \qquad V_y = \frac{\mu_{02}}{\mu_{00}}, \tag{3}$$

  where V_x and V_y are the variances with respect to the x-axis and y-axis, respectively.
- Skewness (symmetry) around the centre of mass can be characterised by the normalised 3^{rd} order moments and defined as

$$S_x = \frac{\mu_{30}}{(\mu_{00} V_x)^{3/2}} \qquad \text{and} \qquad S_y = \frac{\mu_{03}}{(\mu_{00} V_y)^{3/2}}. \tag{4}$$

  Skewness values of $S = 0$, $S < 0$ and $S > 0$ can be interpreted as a Gaussian (normal) distribution, flatter than a normal distribution, and more peaked than a normal distribution, respectively.
- Kurtosis (peakedness) around the centre of mass can be computed from the normalised 4^{th} order moments and defined as

$$K_x = \frac{\mu_{40}}{(\mu_{00} V_x)^2} - 3 \qquad \text{and} \qquad K_y = \frac{\mu_{04}}{(\mu_{00} V_y)^2} - 3. \tag{5}$$

  Kurtosis values of $K = 0$, $K < 0$, $K > 0$ and $K < -1.2$ can be interpreted as a Gaussian (normal) distribution, flatter than normal distribution, more peaked than normal distribution, and bimodal (multi-modal) distribution, respectively.

- The ratio of the longest to the shortest distance vectors from the centroid of the local window to its boundaries is considered as a measure of elongation of the region and is defined as

$$elongation = \frac{(\mu_{20} - \mu_{02})^2 + 4{\mu_{11}}^2}{\mu_{00}}. \tag{6}$$

- For elongated objects, the orientation θ (in degrees) of the major ('long') direction with respect to the x-axis is defined as

$$\theta = \frac{1}{2} \arctan \frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \frac{180}{\pi}. \tag{7}$$

For each pixel, 12 attributes were obtained to form a feature vector (i.e. μ_{00}, m_{10}, m_{01}, $\bar{x}$, $\bar{y}$, m_{11}, elongation, θ, S_x, S_y, K_x, K_y) at four scales; this gives a total of 48 features to represent each pixel.

5.1.3 Feature transformation

A nonlinear transformation, namely a hyperbolic tangent function (Caelli & Oguztoreli, 1987), is used to map moment feature M_k with mean $\overline{M}$ to texture feature F_k defined as

$$F_k(i,j) = \frac{1}{L^2} \sum_{(a,b)\in\omega_{ij}} |\tanh(\sigma(M_k(a,b) - \overline{M}_k))|, \tag{8}$$

where ω_{ij} is an $L \times L$ ($L = 55$) averaging window centred at location (i,j), finer textures require a smaller window to detect smaller features, whereas coarser textures require a larger window; the value $\sigma = 0.01$ controls the shape of the logistic function; and $k = \{1,2,...,12\}$. The parameter values were empirically determined to achieve optimal results, and no significant effect on the results were caused by small variations. Fig. 5 shows example moment images before and after transformation; before the nonlinear transformation, the extracted feature images can be interpreted visually as line detection with different orientations and scales. The enhanced feature images show increased differentiation on tissue specific areas. This is expected to lead to more discriminative feature vectors.

5.1.4 Mammographic model generation

Once the transformed feature vectors were obtained, Tabár mammographic building blocks can be modelled using K-means clustering ($K = 10$). A total of 40 cluster centres were generated to represent nodular, linear, homogeneous and radiolucent tissue (10 for each of the four tissue types). The number of cluster centres (K value) was determined empirically based on the assumption that each mammographic building block contains at least 4 to 6 texture primitives according to their visual appearance covering different orientations and scales. A smaller number of cluster centres can lead to less discriminative models; conversely, too many cluster centres can lead to many similar models and redundancy. Visual assessment of the segmentation of mammographic patches using the manual annotations (*e.g.* outlined tissue specific areas) provided by the expert radiologist, was used to heuristically decide the number of cluster centres. It should be noted that there is no rigorous way to determine exactly how many texture primitives should be used for complex texture patterns as seen in mammographic images. However, it is expected that the set of features describes all mammographic anatomical aspects covering lines and regions at various scales.

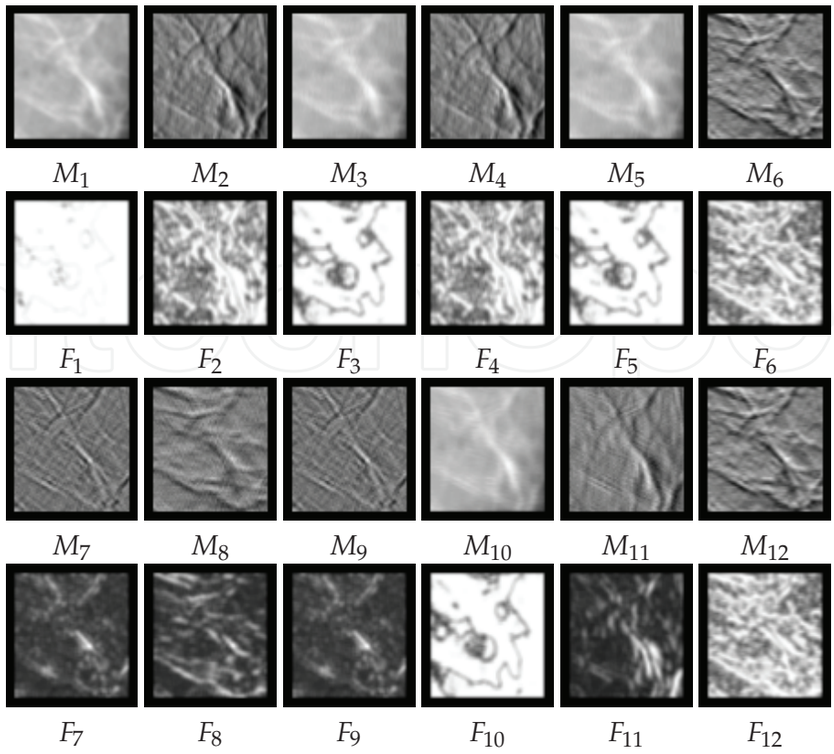


Fig. 5. Example moment images M_k and the corresponding transformed feature images F_k, where $k = 1, 2, ..., 12$. The original linear structure patch is the same as in Fig. 4. The local window $W = 13$.

5.1.5 Model driven segmentation

At the segmentation stage, for each pixel of a mammographic image, the same procedure was applied to obtain transformed feature vectors. For each pixel, the resultant feature vector was compared to the distribution of all the mammographic building block models. A distance weighted *K*-Nearest-Neighbour (KNN) classifier ($K = 9$) was used, to assign a Tabár mammographic building block class to each pixel, which weighted the contribution of each of the *K* nearest neighbours with the Mahalanobis distance (Mahalanobis, 1936) to the query point, giving greater weight to closer neighbours to reduce misclassification. The *K* value was determined empirically, but variation in these parameter indicated robustness.

It is indicated in the computer vision literature that a low intensity background (*e.g.* low or zero grey level values) can lead to meaningless moments (Tucceryan, 1994). When performing texture analysis in mammographic images, this can happen when extracting features of homogeneous regions have high intensity background, and can cause misclassification between high density homogeneous tissue and low density radiolucent tissue. The intensity of homogeneous and radiolucent areas may vary between different mammographic risk patterns; and image normalisation can alter the intensity distribution. To tackle the problem of misclassification due to meaningless moments between high intensity tissues (*e.g.* homogeneous and nodular) and low intensity tissues (*e.g.* radiolucent), a threshold post-processing is incorporated into the classifier to reduce classification errors. This involves reclassifying tissue that is initially classified as radiolucent, but has a very high intensity,

to homogeneous or nodular tissue. The threshold values were determined based on prior knowledge of the intensity distribution and variation of mammographic images across the whole MIAS database, and were defined as the mean values of homogenous and nodular tissues from the collection of annotated mammographic patches. It should be noted that in practice, mammographic images may be obtained through different image acquisition processes. Therefore, for an alternative database, the threshold values may need to be re-evaluated using training images obtained from a specific image acquisition process.

5.2 Texture signatures based approach

Mammographic segmentation using texture signatures based methodology was developed in an attempt to incorporate spatial (*e.g.* textons based techniques) and geometric (*e.g.* moments based techniques) texture features.

For each pixel, a texture signature is generated, consisting of three distinct signatures. Each signature of size $L \times L$ was computed within a local window C, to encode different texture features. To ensure computational correctness, it is necessary to use a circular window instead of a more conventional square window to compute local image features. This is because the proposed methodology uses the distance between a pixel and the centre pixel within the sampling window C as a parameter. The upper limit of the parameter is defined as the value of the max distance between these two points, and the radial distance for any point on a circle to the centre has the same max value. A series of investigations was conducted on parameter settings; covering the full range of anatomical objects. Fourier analysis was used to estimate the first side-band (Gonzales & Woods, 1992) which indicated a circular window of 63 pixels in diameter for feature extraction. Whilst, texture signature size was empirically defined in a dimension of 25×25 pixels, which is expected to be able to encode sufficient mammographic features. A significant amount of randomly sampled points ($\geq$ 60,000) from 78 patches (*e.g.* 16 nodular, 22 linear structure, 23 homogeneous and 17 radiolucent) were used at the training stage, regardless of the associated risk class for the original mammogram.

A texture signature is effectively a stack of three 2D histograms, encoding directional texture features (*e.g.* intensity, orientation and elongation variation).

5.2.1 Texture signature - Part I

The first part of a texture signature consists of a cumulative histogram based on the radial distance between the centre pixel and grey-level intensity within the circular window C. The signature's y-axis represents the grey-level intensity, and the x-axis represents the radial distance to the centre pixel. The occurrence values at any $\{x, y\}$ position within the signature are multiplied with the relevant grey-level value. In the grey-level spatial distribution space, the first part of the texture signature represents the roughness of a given circular texture region. The signature configuration varies rapidly with distance in coarse textures (*e.g.* nodular tissues in Fig. 6 A(a)), and steady with distance in fine textures (*e.g.* homogeneous tissues in Fig. 6 C(a)); the accumulated grey-level close to the top of the signature indicates dominance of low intensity pixel presence (*e.g.* radiolucent fatty tissues in Fig. 6 D(a)), whilst the accumulated grey-level close to the bottom of the signature indicates dominance of high intensity pixel presence (*e.g.* homogeneous tissues in Fig. 6 C(a)).

5.2.2 Texture signature - Part II

The second part of a texture signature is a cumulative histogram based on the angle between the tangent at each segment and grey-level intensity within the circular window C. The angle $\theta = 0^\circ$ is defined when a pixel lies on the positive side of the y-axis of the circular window C, and increases counterclockwise. The signature's y-axis represents the grey-level intensity, and the x-axis represents the tangent at segment within the circular window C. The occurrence values at any $\{x, y\}$ position within the signature are multiplied with the relevant grey-level value. The accumulated grey-level values change with the orientation of the linear structures. Visually speaking, texture signature containing many rough peaks indicate the presence of lines (*e.g.* linear structure in Fig. 6 B(b)). A thick band with spreading tendency appearing in the texture signature indicates the presence of nodular tissues (*e.g.* Fig. 6 A(b)). Note that the linear structure patches may contain noise, and have very low contrast between the foreground and background. These issues make it difficult to extract the related linear structure features. However, the spikes as seen in the corresponding signatures indicate the presence of linear structures and their approximate orientations.

5.2.3 Texture signature - Part III

A texture image may contain repeated occurrence of some grey-level configurations, which vary rapidly in fine textures, and slowly in coarse textures (Haralick et al., 1973). These changes can be encoded in co-occurrence matrices. A co-occurrence matrix is constructed by describing how frequently two pixels, separated by a distance d, with grey-level x_1 and x_2 appear. The occurrence of a grey-level configuration may be described by matrices of relative frequencies. The third part of a texture signature is a modified co-occurrence matrix with $d = 1$ and $\theta = 90^\circ$. In this case, the x and y-axes of the signature represent grey-level information. Note that other configurations are possible, but the experimental results indicated variations in these are not significant. Given a pair of values for the x and y-axis, the absolute value of the intensity difference between two pixels is accumulated, and used as the corresponding histogram value. The statistical variations such as the frequency of some grey-level configurations, and the magnitude of the variance of the configurations, can be used to discriminate breast components with different texture properties (*e.g.* Fig. 6 A(c)-D(c)).

5.2.4 Texture signature - combined (I + II + III)

Finally, all the values in the signature were normalised to zero mean and unit variance. The x and y-axes were normalised to L, to ensure the constructed texture signature has a consistent size of $L \times L$ for the three parts. Fig. 6 shows normalised individual signatures, and the resulting combined texture signatures.

5.2.5 Model generation, selection and reduction

Texture signatures generated from the mammographic patches vary due to mammographic images taken under various exposure conditions, linear structures at different orientations, and stochastic parenchymal patterns in nodular tissue. K-means clustering was used to establish initial models based on the assumption that similar models are closely clustered in the feature space, and can be combined through a clustering process. Initially, 30 cluster centres were empirically defined for each mammographic building block. Note

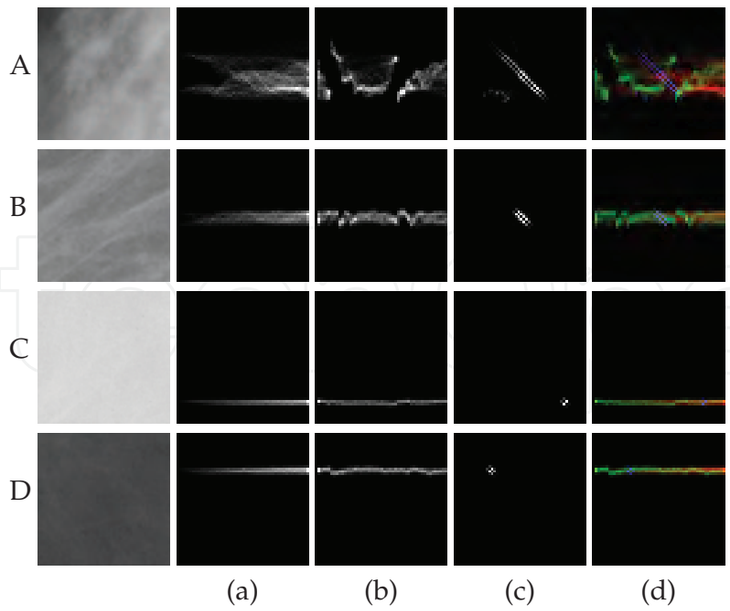


Fig. 6. From top to bottom showing mammographic patches containing nodular, linear structure, homogeneous and radiolucent tissue, respectively. For each row, from left to right showing an mammographic patch, three individual signatures, and the texture signature.

that to predetermine the appropriate number of cluster centres may prove challenging for complicated breast anatomic components. Therefore, the initial number of cluster centres was over-estimated deliberately to facilitate further model selection and reduction. This model generation by clustering resulted in a total of 120 (30×4) cluster centres to represent four types of mammographic building blocks. For each mammographic building block model M_i ($i \in 1, 2, ...120$), a cross voting process was employed to achieve the optimal model subsets, by determining whether it is the nearest neighbour (in a Euclidean sense) of the other models of the same type of mammographic building block. For instance, if model M_i receives votes from a subset of models M_s, it can be interpreted as to be representative of the mammographic building block. Models receiving no or few votes may be outliers, and are removed. The outlier bounds (threshold values) were determined empirically through a series of tests on mammographic patch segmentation. The threshold values for nodular and homogeneous tissue models, were defined as half of the largest standard deviation values of the corresponding models; whereas for linear and radiolucent tissue models, the threshold values were defined as half of the smallest standard deviation values of the corresponding models. As a result, a total of 42 models (i.e. 12 for nodular tissue, 15 for linear tissue, 8 for homogeneous tissue, and 7 for radiolucent tissue) were selected from the initial 120 models.

5.2.6 Model driven segmentation

At the segmentation stage, the same procedure was used to generate a texture signature at each pixel. The resultant texture signatures were compared to the 42 selected mammographic building block models. A distance weighted K-Nearest-Neighbour (KNN) classification ($K = 3$) was used in the Euclidean space, to assign a mammographic building block to each pixel, which weighted the contribution of each of the K nearest neighbours with the distance to the query point, giving greater weight to closer neighbours to reduce misclassification. The

K value was determined empirically, but variation in this parameter indicated robustness, in terms of the number of correctly classified pixels and visual correctness of the segmented tissue specific areas.

6. Evaluation

A comparative evaluation of two mammographic segmentation methods was conducted; evaluation aspects include: 1) mammographic segmentation using the MIAS database, 2) quantitative and qualitatively measurement on the derived tissue compositions from the segmented images, 3) mammographic risk classification based on the tissue compositions, and 4) clinical evaluation of the segmentation to assess the realistic and practical use of the proposed methodologies in a clinical environment.

Customised software was developed for evaluation, to determine the effectiveness of the segmentation methodologies and the potential clinical utilisation for screening mammography or computer aided diagnosis systems. An expert consultant radiologist performed the evaluation by grading a given mammographic segmentation from five available options (i.e. *unacceptable*, *poor*, *acceptable*, *good* and *very good*). In addition, the grades can be combined and referred to as principal grades; the first two grades were considered as *clinically unacceptable* (CUA), the remaining three were considered as *clinically acceptable* (CA)). The statistical results are presented as percentages with respect to the five grades and principal grades, and associated with all Tabár and BIRADS risk classification, as well as the corresponding low and high risk categories. Expert feedback was collected and incorporated in the investigation, including the segmentation accuracy on the tissue specific areas, and the misclassified breast tissues associated with the segmentation methodology issues. All the clinical evaluation aspects were taken into account to conclude the relationship between the risk classification and clinical practice satisfaction.

7. Results and discussion

This section compares the experimental results, and is arranged in line with the evaluation stages described in Section 6. With respect to the mammographic risk classification, both Tabár and BIRADS risk categories are included, in order to facilitate the classification accuracy comparison achieved using existing work in the literature.

7.1 Mammographic segmentation

Mammographic segmentation was performed on the MIAS database using both the moments and texture signatures based methods. Visually speaking, comparing the results achieved using a filter bank based texton approach in (He et al., 2008), the segmentation results produced by both the methods show significant improvements with respect to segmentation on tissue specific areas (see Fig. 7 for examples). In terms of mammographic segmentation accuracy, pectoral muscle and upper abdominal fat are often misclassified as nodular or homogeneous fat within nodular tissues. With respect to segmentation produced using the moments based method, dense areas are often misclassified as radiolucent tissue. Misclassified tissues between nodular and homogeneous classes may be caused by under exposed films. This indicated the lack of discriminative modelling between nodular and homogeneous tissues, as well as between nodular and radiolucent tissues. The segmentation results indicated that both the moments and texture signatures based methods are capable of

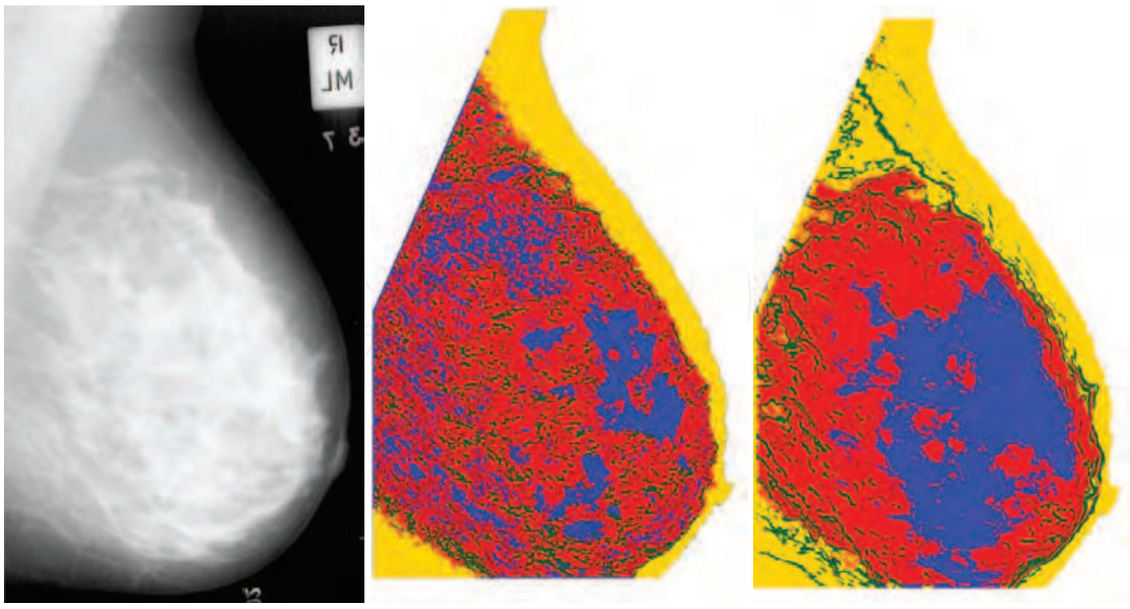


Fig. 7. Example mammographic segmentation. Tabár risk pattern IV; BIRADS risk pattern III. This shows the original mammographic image (left) and the segmentation produced using the moments (middle) and texture signatures (right) based methodologies. The tissue specific areas are: nodular (red), linear (green), homogeneous (blue) and radiolucent (yellow). Note that nodular tissue was over-segmented when using the moments based method.

incorporating shape related texture features during the Tabár tissue modelling. Although the segmentation seems to be less accurate on extremely larger tissue structures. This is due to the use of sampling windows at the feature extraction stage, as the encompassed region for a local window may not be sufficiently large to capture these structures. However, using a larger window can compromise computational efficiency.

With respect to the moments based approach, some of the misclassified pixels within low contrast regions (*e.g.* radiolucent tissue) may be caused by meaningless moments. High intensity homogeneous density regions (*e.g.* homogeneous tissue) can lead to a similar problem. This indicates that the currently used texture features may miss out some discriminative intensity information. It seems necessary to have an approach designed specifically for dealing with intensity variation as seen on digitised mammograms; the zero mean normalisation method used may be too generalised, and can have a direct effect on the model generation. A threshold based post-processing step was used to reduce the false positives using prior knowledge. This process is effective but may require tuning for an alternative database (*e.g.* DDSM (Heath et al., 1998)), because the mammographic image intensity distribution can be affected, due to a different image acquisition process.

7.2 Tissue composition

Table 1 shows the average tissue compositions (the relative proportions of the four mammographic building blocks) achieved, based on the developed segmentation approaches. The differences between the achieved and proposed tissue compositions (see Section 2.2) were measured by the Euclidean distance (ED). The smaller the distance, the more similar

	Nodular	Linear	Homogeneous	Radiolucent	ED
	Tabár pattern I				
Tabár *et al.* (Tabár et al., 2004)	25	15	35	25	
Moment	27±18	15±5	26±17	32±21	11.6
Texture Signature	19±5.8	14±5.3	26±15.6	41±16.3	19.3
	Tabár pattern II/III				
Tabár *et al.* (Tabár et al., 2004)	2	14	2	82	
Moment	14±10	9±4	21±20	56±23	34.7
Texture Signature	10±9.4	12±4.2	10±18.8	68±21.6	18.1
	Tabár pattern IV				
Tabár *et al.* (Tabár et al., 2004)	49	19	15	17	
Moment	38±19	12±5	26±15	24±14	18.4
Texture Signature	44±11.7	11±4.3	13±11.5	32±12.4	17.8
	Tabár pattern V				
Tabár *et al.* (Tabár et al., 2004)	2	2	89	7	
Moment	12±18	4±4	67±27	17±13	26.2
Texture Signature	11±3.3	7 ±4.4	64±11.0	18±10.9	29.2

Table 1. The relative proportions of the four mammographic building blocks and their standard deviations, obtained using the developed mammographic segmentation approaches. The Euclidean distances between the achieved and the tissue compositions proposed by Tabár are shown in the column 'ED'.

the achieved and proposed tissue composition, which also reflects the effectiveness of the developed mammographic segmentation approaches. The ED results indicated that the agreement between the proposed tissue compositions are closer to those achieved using the texture signatures based approach, compared to the moments based approach. The average values of the summations of the variances for the risk patterns for the moments and texture signatures based approach are 90.9 and 84.4, respectively. The smaller average total variance can be linked to the satisfactory classification accuracies achieved using texture signatures based mammographic segmentation (see Sections 7.3 and 7.4).

The resultant tissue composition from each segmented mammographic image was used as a feature vector for the risk classification. With respect to Tabár based risk classification, the feature vectors were compared with Tabár models (risk patterns) and using a Euclidean distance. Whilst BIRADS based risk classification was performed using leave-one (mammographic image)-out methodology. The ground truth was provided by an expert radiologist.

7.3 Risk classification - Tabár's risk categories

In the computer vision literature, very few investigations were conducted with respect to mammographic risk classification based on Tabár's risk categories. These classification results are summarised and compared in Table 3. Details of risk classification confusion matrices are shown in Table 2. A direct comparison with existing mammographic risk assessment methods is not always possible due to differences in databases choice. Jamal *et al.* (Jamal et al., 2006) used 122 mammograms, 30, 21, 20, 27 and 24 for patterns I to IV, respectively. It should be noted that the used dataset is much smaller than the MIAS database and not publicly

		Automatic				
	Tabár	I	II/III	IV	V	Accuracy
Truth	I	33	36	41	9	28%
	II/III	14	66	5	8	71%
	IV	22	6	50	2	63%
	V	1	1	5	21	75%

		Automatic				
	Tabár	I	II/III	IV	V	Accuracy
Truths	I	102	16	1	0	86%
	II/III	7	61	11	14	66%
	IV	9	5	65	1	81%
	V	1	4	0	23	82%

Table 2. Classification confusion matrices based on Tabár risk categories. With respect to the moments based method (left), total accuracy was 53%, $\kappa = 0.36$ (fair agreement). Total accuracy for low and high categories was 71%, $\kappa = 0.40$ (fair agreement). With respect to the texture signatures (right) based method, total accuracy was 78%, $\kappa = 0.7$ (substantial agreement). Total accuracy for low and high categories was 86%, $\kappa = 0.69$ (substantial agreement).

Database		I	II/III	IV	V	Total_1	Low	High	Total_2
	Jamal (Jamal et al., 2006)	73	90	74	88	83			
MIAS	He_M (He et al., 2009)	28	71	63	75	53	70	72	71
	He_TS (He et al., 2010)	86	66	81	82	78	88	82	86

Table 3. Mammographic risk classification based on Tabár's risk categories. He_M (He et al., 2009) used the developed moments based approach. He_TS (He et al., 2010) used the developed texture signatures based approach.

available. Therefore, the distribution over the inter and intra class variations are expected to be different, which can effect the classification results. With respect to the classification achieved using the developed methods, strong correlation was found between the precision of mammographic segmentation and subsequent risk classification. The segmentation accuracy is particular critical when performing risk classification based on Tabár tissue modelling, because tissue compositions for the segmented images were used as feature vectors at the risk classification stage. The results show overall good classification accuracies, except for Tabár pattern I results achieved using the moments based approach. It should be noted that the latter aspect can be linked to previous published results (Muhimmah et al., 2006), which indicated that Tabár Patten I shows weak correlation between BIRADS and Tabár risk models.

7.4 Risk classification - BIRADS risk categories

The majority of the investigations with respect to mammographic risk classification are based on BIRADS risk categories. The classification results are summarised and compared in Table 5. Details of risk classification confusion matrices are shown in Table 4. Note that Bovis *et al.* (Bovis & Singh, 2002) used mammographic images for 377 patients from DDSM. Petroudi *et al.* (Petroudi et al., 2003) used 132 images selected from the Oxford database (not publicly available). A direct comparison with Oliver *et al.* (Oliver et al., 2008) is probably the closest option. Oliver *et al.* (Oliver et al., 2008) shows divergence in the opinion of three expert radiologists, and the inter observer variation is dealt with using the majority vote from the three radiologists. In the experiment, the consensus classification was used and superior classification was achieved. It should be noted that the inter observer variation seems to be minimised by using the consensus classification. In (Oliver et al., 2008) the MIAS database was used, and 82% correct classification was obtained based on 'expert C' (Oliver et al., 2008).

		Automatic				
	BIRADS	I	II	III	IV	Accuracy
Truth	I	65	10	1	11	75%
	II	0	71	21	11	69%
	III	0	18	63	12	68%
	IV	0	5	7	25	68%

		Automatic				
	BIRADS	I	II	III	IV	Accuracy
Truths	I	41	5	10	3	70%
	II	1	68	12	5	79%
	III	0	11	101	29	71%
	IV	0	0	5	29	85%

Table 4. Classification confusion matrices based on BIRADS risk categories. With respect to the moments based method (left), total accuracy was 70%, $\kappa = 0.59$ (moderate agreement). Total accuracy for low and high categories was 79%, $\kappa = 0.58$ (moderate agreement). With respect to the texture signatures (right) based method, total accuracy was 75%, $\kappa = 0.75$ (substantial agreement). Total accuracy for low and high categories was 87%, $\kappa = 0.74$ (substantial agreement).

Database		I	II	III	IV	Total_1	Low	High	Total_2
MIAS	Oliver(A) (Oliver et al., 2008)	80	60	67	46	66	79	77	79
	Oliver(B) (Oliver et al., 2008)	89	75	69	59	75	92	77	85
	Oliver(C) (Oliver et al., 2008)	91	84	89	73	86	89	94	91
	He_M (He et al., 2009)	47	73	87	65	73	73	91	83
	He_TS (He et al., 2010)	70	79	72	85	75	79	94	87
DDSM	Oliver (Oliver et al., 2008)	55	88	77	69	77	89	79	84
	Bovis (Bovis & Singh, 2002)					71			97
Oxford	Petroudi (Petroudi et al., 2003)	91	64	70	78	76	91	94	

Table 5. Mammographic risk classification based on BIRADS risk categories and comparison with existing work using: Oliver(A), the segmentation and feature extraction are described in (Karssemeijer, 1998); Oliver(B), the segmentation approach is described in (Karssemeijer, 1998) and the feature extraction is used in (Oliver et al., 2008); Oliver(C), both the segmentation and feature extraction are described in (Oliver et al., 2008). He_M (He et al., 2009) used the developed moments based approach. He_TS (He et al., 2010) used the developed texture signatures based approach.

It should be noted that the proposed mammographic segmentation is specifically designed with respect to Tabár mammographic parenchymal patterns, and has fundamental differences to density based segmentation described in (Oliver et al., 2008). In this study, BIRADS risk classification was obtained to facilitate the classification performance evaluation when using different risk models (i.e. Tabár and BIRADS), as using mammographic parenchymal patterns rather than density information alone.

7.5 Clinical evaluation

The risk classification results (see Tables 2 and 4) show both Tabár and BIRADS related aspects, and indicate that the results for the high risk mammograms are significantly better than those for the low risk mammograms; also show a clear correlation between the Tabár and BIRADS results, and is in line with previous published work (Muhimmah et al., 2006). Table 6 shows the segmentation assessment results with respect to the moments based approach, in which for Tabár II/III and BIRADS I, at about 41%, are especially low. The other individual

classes are encouraging, with the Tabár V and BIRADS IV classes reaching 100% clinical acceptable results (note that the number of samples for Tabár V or BIRADS IV are small). The segmentation assessment results based on all the images in both Tabár and BIRADS risk categories indicate the robustness of the segmentation methodology, as both achieved 70% clinical acceptable results. Further details of the segmentation assessment results can be found in Table 7, where the classification results were correlated from the risk classification confusion matrices (see Tables 2 (left) and 4 (left)) and Table 6. Table 7 shows the number of cases which were correctly classified (*e.g.* all the cases on the diagonal of a confusion matrix) in the original mammographic risk assessment experiment against the clinical evaluation of the segmentation results. The results for mammographic risk assessment that were misclassified by a single, two and three classes (*e.g.* +/- 1 to II would be I and III) are also included (it should be noted that the misclassification by three classes only happens with a few cases).

The same clinical satisfaction study was conducted with respect to the results obtained using the texture signatures based approach. Table 8 shows the segmentation assessment results, which for Tabár I, at 42% is especially low; however, the clinical acceptable rate increases as the risk level gets higher which is particularly encouraging, with the Tabár V class reaching over 96% clinical acceptable results. For BIRADS risk classes I-III, the clinical acceptable rate for segmentation is about 55%; and 92% clinical acceptable rate is achieved for BIRADS risk class IV. It should be noted that the number of samples for Tabár V or BIRADS IV are small. These aspects are also clear for the low/high risk results, which show a relative better clinical acceptable result for high risk categories, where the high risk results based on Tabár risk modelling are promising as over 76% are considered clinically acceptable results. Further details of the segmentation assessment results can be found in Table 9, where the classification results from the risk classification confusion matrices (see Tables 2 (right) and 4 (right)) and Table 8 are correlated. Table 9 shows the number of cases which were correctly classified (*e.g.* all the cases on the diagonal of a confusion matrix) in the mammographic risk assessment experiments against the clinical evaluation of the segmentation results.

The results in Tables 7 and 9 show strong correlations between the level of correctness of the mammographic risk classification and clinical evaluation of the segmentation results: i.e. correct mammographic risk classification is associated with the *good* and *very good* segmentation results, whilst misclassification by two and three classes is mainly associated with the *poor* and *unacceptable* segmentation results. This is especially true for the BIRADS based results, where the (almost) correctly classified images each have a clinical acceptable segmentation, as opposed to only a few cases that are in the clinical acceptable segmentation category for the images associated with significant misclassification for mammographic risk assessment.

The decreasing accuracies in all the individual low risk classes indicates the precision of the segmentation are not all up to the clinical standard. From a clinical point of view, a mathematically correct risk classification (possibly due to a robust classifier), does not necessarily reflect the correctness of the associated segmentation. Even though, the clinical evaluation still show strong positive correlations with the automatic risk classification. It is interesting to see the accuracies increased in the individual high risk classes. This seems to indicate an apparent difference between the way in which radiologists perceive mammographic risk and how the segmentation performs, with especially the segmentation of radiolucent tissues maybe less robust. However, in associating the correct classification

Tabár	♯images	U	P	A	G	VG	CA
I	119	8%	11%	14%	29%	37%	81 %
II/III	93	32%	24%	8%	9%	28%	44 %
IV	80	5%	20%	14%	20%	41%	75 %
V	28	0%	0%	4%	43%	54%	100 %
Low	212	19%	17%	11%	20%	33%	65 %
High	108	4%	15%	11%	26%	44%	81 %

BIRADS	♯images	U	P	A	G	VG	CA
I	59	35%	24%	11%	8%	22%	41 %
II	86	20%	20%	9%	19%	32%	60 %
III	141	4%	14%	15%	26%	41%	82 %
IV	34	0%	0%	3%	38%	59%	100 %
Low	145	26%	22%	10%	14%	28%	52 %
High	175	3%	12%	13%	29%	43%	85 %

Table 6. Clinical grading of satisfaction regarding the mammographic segmentation using the moments approach. 'U', 'P', 'A', 'G' and 'VG' denote *unacceptable*, *poor*, *acceptable*, *good* and *very good*, respectively.

Tabár	♯images	U	P	A	G	VG	CA
correct	251	8%	17%	13%	25%	37%	75%
+/- 1	40	28%	20%	3%	5%	45%	53%
+/- 2	28	46%	4%	7%	18%	25%	50%
+/- 3	1	0%	0%	0%	100%	0%	100%
(LH) correct	275	9%	16%	12%	24%	39%	75%
(LH) incorrect	45	42%	13%	4%	13%	27%	44%

BIRADS	♯images	U	P	A	G	VG	CA
correct	239	9%	18%	10%	23%	40%	73%
+/- 1	63	14%	8%	16%	25%	37%	78%
+/- 2	15	80%	13%	7%	0%	0%	7%
+/- 3	3	33%	67%	0%	0%	0%	0%
(LH) correct	279	9%	16%	12%	24%	38%	75%
(LH) incorrect	41	46%	12%	5%	10%	27%	41%

Table 7. Clinical grading of satisfaction regarding the mammographic segmentation using the moments approach, associated with the level of the automatic mammographic risk classification. 'LH' denotes low and high risk.

Tabár	♯imgs	U	P	A	G	VG	CA
I	119	31%	27%	16%	18%	8%	42 %
II/III	93	23%	19%	22%	22%	14%	58 %
IV	80	14%	17%	24%	25%	20%	69 %
V	28	4%	0%	14%	21%	61%	96 %
Low	212	27%	24%	19%	19%	11%	49 %
High	108	11%	13%	21%	24%	31%	76 %

BIRADS	♯imgs	U	P	A	G	VG	CA
I	87	19%	24%	29%	25%	3%	57 %
II	103	28%	16%	20%	17%	19%	56 %
III	93	24%	24%	16%	21%	15%	52 %
IV	37	2%	6%	21%	21%	50%	92 %
Low	190	24%	19%	23%	22%	12%	57 %
High	130	20%	20%	17%	21%	22%	60 %

Table 8. Clinical grading of satisfaction regarding the mammographic segmentation using the texture signatures approach. 'U', 'P', 'A', 'G' and 'VG' denote *unacceptable*, *poor*, *acceptable*, *good* and *very good*, respectively.

Tabár	♯imgs	U	P	A	G	VG	CA
correct	170	11%	21%	22%	24%	22%	68%
+/- 1	68	16%	13%	22%	29%	20%	71%
+/- 2	72	50%	21%	14%	8%	7%	29%
+/- 3	10	40%	40%	20%	0%	0%	20%
(LH) correct	227	11%	20%	22%	26%	21%	69%
(LH) incorrect	93	48%	20%	13%	10%	9%	32%

BIRADS	♯imgs	U	P	A	G	VG	CA
correct	224	16%	16%	23%	25%	20%	68%
+/- 1	68	24%	32%	13%	15%	16%	44%
+/- 2	17	58%	24%	12%	0%	6%	18%
+/- 3	11	82%	9%	9%	0%	0%	9%
(LH) correct	253	15%	19%	22%	25%	19%	66%
(LH) incorrect	67	48%	25%	10%	7%	10%	27%

Table 9. Clinical grading of satisfaction regarding the mammographic segmentation using the texture signatures approach, associated with the level of automatic mammographic risk classification. 'LH' denotes low and high risk.

of mammograms with the assessment of the segmentation results, Tables 7 and 9 indicated a strong correlation between these. This might also indicate that the tissue modelling associated with the segmentation process, does not cover the various tissue classes appropriately, and there might be a strong non-linear component in the expert assessment (*e.g.* when the areas of dense tissues are significant, the areas of linear and radiolucent tissues play a less important role), which is not present in the tissue modelling approach. It is encouraging to see in Table 8 that the accuracies increased in high risk classes (i.e. Tabár V and BIRADS IV). It is also interesting to notice small classification variation between BIRADS risk classes I, II and III. This may indicate that the presented methodology is able to discriminate mammographic parenchymal patterns well based on Tabár risk modelling, but less robust in extracting density information (possible due to meaningless moments) based on BIRADS risk models.

The major concerns from the radiologist who participated in the clinical evaluation include: 1) pectoral muscle and upper abdominal fat are often misclassified as nodular or homogeneous fat within nodular tissues; 2) radiolucent areas are often misclassified as dense tissue; and 3) misclassification of tissue between nodular and homogeneous classes may be caused by (apparently) under exposed films. Clinical feedback indicates some segmentation methodology issues, such as the lack of discriminative modelling between nodular and homogeneous tissues, as well as between nodular and radiolucent tissues. Intra and inter observer variation is noticeable; however, it should be noted that the issue of experts' subjective assignment of risk classification can be dealt with using multiple readers, and the majority risk classifications can be considered as 'final', but full evaluation of these aspects is seen as future work. For automatic mammographic risk assessment based on parenchymal pattern segmentation, it is encouraging that applying advanced machine learning techniques achieves an improved risk classification. At the same time, a follow up clinical evaluation linking the risk classification and segmentation should be emphasised, to ensure a realistic practical usage in a clinical environment.

8. Conclusions

It is possible to use computer vision techniques to estimate the mammographic risk based on the composition of the four mammographic building blocks (i.e. nodular, linear, homogeneous and radiolucent) described by Tabár. The presented results using the moments and texture signatures based methodologies show realistic segmentation on tissue specific areas, improved estimation of the relative proportions of the four building blocks when compared to the results achieved using the textons based approach (He et al., 2008), and leading to promising mammographic risk classification results. The presented clinical evaluation, both quantitative and qualitative, confirms that the technical evaluation of an automatic risk classification based on Tabár's tissue modelling, may not reflect the effectiveness of the methodology in a clinical setting. Even though, the clinical evaluation still shows strong positive correlations with the automatic risk classification. In addition, the segmentation assessment is linked to the correct/incorrect automatic mammographic risk classification, which indicated that anatomically correct segmentation results tend to lead to correct mammographic risk estimation. However, work towards a robust automatic mammographic segmentation using Tabár's tissue modelling is still on going; such an application is expected to be useful as a means of aiding radiologists' diagnosis and treatment planning prior to biopsies.

9. References

American College of Radiology (2004). *Breast Imaging Reporting and Data System BI-RADS*, 4th edn, Reston, VA: American College of Radiology.

Andoni, A., Indyk, P. & Krauthgamer, R. (2008). Earth mover distance over high-dimensional spaces, pp. 343 – 352.

Astley, S. M. (2004). Computer-based detection and prompting of mammographic abnormalities, *The British Journal of Radiology* 77: S194–S200.

Awcock, G. J. & Thomas, R. (1996). *Applied image processing*, McGraw-Hill, Inc., Hightstown, NJ, USA.

Bovis, K. & Singh, S. (2002). Classification of mammographic breast density using a combined classifier paradigm, *In 4th International Workshop on Digital Mammography*, pp. 177–180.

Boyd, N. F., Byng, J. W., Jong, R. A., Fishell, E. K., Little, L. E., Miller, A. B., Lockwood, G. A., Tritchler, D. L. & Yaffe, M. J. (1995). Quantitative classification of mammographic densities and breast cancer risk: results from the canadian national breast screening study, *Journal of the National Cancer Institute* 87: 670–675.

Boyd, N. F., Martin, L. J., Sun, L., Guo, H., Chiarelli, A., Hislop, G., Yaffe, M. & Minkin, S. (2006). Body size, mammographic density, and breast cancer risk, *Cancer Epidemiology Biomarkers & Prevention* 15(11): 2086–2092.

Brisson, J., Diorio, C. & Mâsse, B. (2003). Wolfe's parenchymal pattern and percentage of the breast with mammographic densities: Redundant or complementary classifications?, *Cancer Epidemiology Biomarkers & Prevention* 12(8): 728–732.

Caelli, T. & Oguztoreli, M. N. (1987). Some tasks and signal dependent rules for spatial vsion, *Spatial Vision* 2: 295 – 315.

Ciatto, S., Visioli, C., Paci, E. & Zappa, M. (2004). Breast density as a determinant of interval cancer at mammographic screening, *Journal of Medical Screening* 90: 393–396.

Couto, E., Harrison, D., Duffy, S., Myles, J., Sala, E., Warren, R., Day, N., Luben, R. & Chen, H. (2001). Estimation of disease progression parameters from case-control data: application to mammographic patterns and breast cancer natural history, *Journal of Epidemiology and Biostatistics* 6: 235–242.

Eisa, M., Refaat, M. & El-Gamal, A. F. (2009). Preliminary diagnostics of mammograms using moments and texture features, *International Journal on Graphics, Vision and Image Processing* 9: 21–27.

Evertsz, C. J. G., Bodicker, A., Hendricks, J. H. C. L., Karssemeijer, N., Weber, U., Bohnenkamp, S., Dechow, D., Berger, L., Woudend, S. van., Brady, M., Holland, R. & Peitgen, H. -O. (2000). High throughput soft-copy reading system for digital mammography in nationwide european screening mammography programs: Requirements and first solutions, *Proceedings of the Radiographic Society of North America*.

Gong, Y. C., Brady, M. & Petroudi, S. (2006). Texture based mammogram classification and segmentation, *8th International Workshop on Digital Mammography*, pp. 616–625.

Gonzales, R. & Woods, R. (1992). *Digital Image Processing*, Addison-Wesley Publishing Company.

Guliato, D., de Carvalho, J. D., Rangayyan, R. M. & Sergio, S. A. (2008). Feature extraction from a signature based on the turning angle function for the classification of breast tumors, *Journal of Digital Imaging* 21: 129–144.

Haralick, R. M., Shanmugam, K. & Dinstein, I. (1973). Textural features for image classification., *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 3, pp. 610–621.

He, W., Denton, E. R. E. & Zwiggelaar, R. (2009). Mammographic segmentation based on mammographic parenchymal patterns and spatial moments, *9th International Conference on Information Technology and Applications in Biomedicine*, pp. 1–4.

He, W., Denton, E. R. E. & Zwiggelaar, R. (2010). Mammographic image segmentation and risk classification using a novel texture signature based methodology, *Lecture Notes in Computer Science*, Vol. 6136, Springer, pp. 526–533.

He, W., Muhimmah, I., Denton, E. R. E. & Zwiggelaar, R. (2008). Mammographic segmentation based on texture modelling of Tabár mammographic building blocks, *Lecture Notes in Computer Science*, Vol. 5116, pp. 17–24.

Heath, M., Bowyer, K., Kopans, D., Kegelmeyer, W. P., Moore, R., Chang, K. & MunishKumaran, S. (1998). Current status of the digital database for screening mammography, *International Workshop on Digital Mammography*, pp. 457–460.

Hitchcock, F. L. (1941). The distribution of a product from several sources to numerous localities, *Journal of Mathematical Physics* 20: 224 – 230.

Jamal, N., Ng, K. H., Looi, L. M., McLean, D., Zulfiqar, A., Tan, S. P., Liew, W. F., Shantini, A. & Ranganathan, S. (2006). Quantitative assessment of breast density from digitized mammograms into Tabár's patterns, *Physics in Medicine and Biology* 51(22): 5843–5857.

Karssemeijer, N. (1998). Automated classification of parenchymal patterns in mammograms, *Physics in Medicine and Biology* 43(2): 365–378.

Mahalanobis, P. C. (1936). On the generalised distance in statistics, *In Proceedings National Institute of Science* 2: 49 – 55.

Mencattini, A., Salmeri, M., Lojacono, R., Rabottino, G. & Romano, S. (2007). Mammographic image analysis for tumoral mass automatic classification, *European Conference on Medical Physics*, Castelvecchio Pascoli, Italy.

Muhimmah, I., He, W., Denton, E. R. E. & Zwiggelaar, R. (2007). Segmentation based on textons and mammographic building blocks, *Medical Image Understanding and Analysis*, pp. 228–232.

Muhimmah, I., Oliver, A., Denton, E., Pont, J., Perez, E. & Zwiggelaar, R. (2006). Comparison between Wolfe, Boyd, BI-RADS and Tabár based mammographic risk assessment, *8th International Workshop on Digital Mammography*, pp. 407–415.

Oliver, A., Freixenet, J., Marti, R., Pont, J., Perez, E., Denton, E. R. E. & Zwiggelaar, R. (2008). A novel breast tissue density classification framework, *IEEE Transactions on Information Technology in BioMedicine* 12: 55–65.

Oliver, A., Lladó, X., Martí, R., Freixenet, J. & Zwiggelaar, R. (2007). Classifying mammograms using texture information, *Medical Image Understanding and Analysis*, pp. 223–227.

Petroudi, S. & Brady, M. (2006). Breast density segmentation using texture, *8th International Workshop on Digital Mammography*, pp. 609–615.

Petroudi, S., Kadir, T. & Brady, M. (2003). Automatic classification of mammographic parenchymal patterns: A statistical approach, *Engineering in Medicine and Biology Society*, Vol. 1, pp. 798–801.

Sickles, E. A. (2007). Wolfe mammographic parenchymal patterns and breast cancer risk, *American Journal of Roentgenology* 188(2): 301–303.

Soares, L. M., Conci, A. & Vianna, A. D. (1998). Automated classification of masses on mammography, *International Symposium on Computer Graphics, Image Processing and Vision*, pp. 1–4..

Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S., Taylor, P., Betal, D. & Savage, J. (1994). The mammographic images analysis society digital mammogram database, *in* Dance, Gale, Astley & Gairns (eds), *Excerpta medica. International Congress Series*, Vol. 1069, Elsevier, pp. 375–378.

Tabár, L., Tot, T. & Dean, P. B. (2004). *Breast Cancer: The Art And Science Of Early Detection With Mamography: Perception, Interpretation, Histopatholigic Correlation*, 1 edn, Georg Thieme Verlag.

Tavassoly, F. A. (1992). *Pathology of the Breast*.

Tucceryan, M. (1994). Moment-based texture segmentation, *Pattern Recognition Letters* 15(7): 659–668.

Varma, M. & Zisserman, A. (2004). Unifying statistical texture classification frameworks, *Image and Vision Computing* 22: 1175 – 1183.

Vyas, V. S. & Priti, P. R. (2007). Malignancy texture classification in digital mammograms based on Chebyshev moments and logpolar transformation, *ICGST International Journal on Bioinformatics and Medical Engineering* 7: 29 – 35.

Wolfe, J. N. (1976a). Breast patterns as an index of risk for developing breast cancer, *American Journal of Roentgenology* 126(6): 1130–1137.

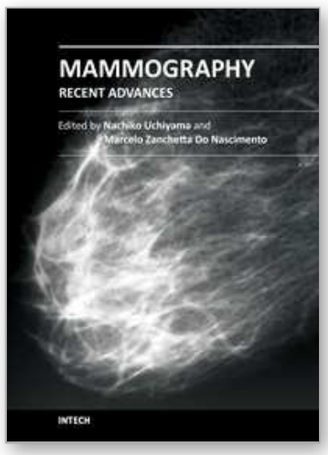
Wolfe, J. N. (1976b). Risk for breast cancer development determined by mammographic parenchymal pattern, *Cancer* 37: 2486–2492.

Zwiggelaar, R. (2002). Image processing in scale-orientation space, *British Machine Vision Conference*, pp. 203–212.

Zwiggelaar, R., Blot, L., Raba, D. & Denton, E. R. E. (2003). Set-permutation-occurrence matrix based texture segmentation, *Iberian Conference on Pattern Recognition and Image Analysis*, pp. 1099–1107.

Zwiggelaar, R. & Denton, E. R. E. (2004). Optimal segmentation of mammographic images, *International Workshop on Digital Mammography*, pp. 751–757.

Zwiggelaar, R., Taylor, C. J. & Rubin, C. M. E. (1999). Detection of the central mass of spiculated lesions - signature normalisation and model data aspects, *Information Processing in Medical Imaging*, Springer-Verlag, London, UK, pp. 406–411.



Mammography - Recent Advances
Edited by Dr. Nachiko Uchiyama

ISBN 978-953-51-0285-4
Hard cover, 418 pages
Publisher InTech
Published online 16, March, 2012
Published in print edition March, 2012

In this volume, the topics are constructed from a variety of contents: the bases of mammography systems, optimization of screening mammography with reference to evidence-based research, new technologies of image acquisition and its surrounding systems, and case reports with reference to up-to-date multimodality images of breast cancer. Mammography has been lagged in the transition to digital imaging systems because of the necessity of high resolution for diagnosis. However, in the past ten years, technical improvement has resolved the difficulties and boosted new diagnostic systems. We hope that the reader will learn the essentials of mammography and will be forward-looking for the new technologies. We want to express our sincere gratitude and appreciation?to all the co-authors who have contributed their work to this volume.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen