Spicule descriptors designed in complex wavelet domain to recognize breast cancer symptoms

Magdalena Jasionowska*, Artur Przelaskowski #, and Rafał Jóźwiak #

*The Faculty of Electronics and Information Technology, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw
# The Faculty of Mathematics and Information Science, Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw

Corresponding Author:
Rafał Jóźwiak
e-mail: R.Jozwiak@mini.pw.edu.pl
tel: +48 22 234 73 07
fax: +48 22 825 37 69

The total number of words in the manuscript, including entire text from title page up to figure legends: 7085
The number of words of the abstract: 200
The number of figures: 4
The number of tables: 1
Abstract

The subject of presented paper is effective recognition of radiating spicules on digital mammograms. The presence of the spicules is a dominant symptom of neoplastic breast lesions called architectural distortions. The originality of the proposed method is the extraction of effective descriptors concentrated with the local directional activity of mammographic texture. Applied methodology was based on the analysis and constructively modeling the conditioning of spicules distribution in complex wavelet domain. It is because of potentially acquisition-invariant energy compaction across directionally spread, especially for piecewise linear structures represented across scales, directions and locations of normalized domain. Additionally, the use of non-directional properties of mammographic findings completes essential conditioning of abnormal spicule appearance. Adaptive analysis of breast tissue distribution in complex wavelet domain allows separating specific local centers of increased multidirectional texture activity among slightly varied but the dominant tendency of tissue directional distribution. Optimized and empirically verified numerical descriptors of local tissue spiculation were fundamental aspect of the proposed effective method to recognize radiating cancer symptoms. Experimental study with a test set of 280 regions of interests, containing normal and abnormal breast tissue of clinically confirmed ADs, have revealed the recognition sensitivity and specificity close to nearly 77% and over 73.5%, respectively.

Keywords

Computer-aided breast cancer diagnosis, mammographic spicule recognition, content-based descriptors, complex wavelets, image analysis.

Glossary of terms

AD – architectural distortion
SM – spiculated mass
CAD – computer-aided diagnosis
FPI – false positive per image
CW – complex wavelets
DT-CWT – dual-tree complex wavelet transform
ROI – region of interests
1. Introduction

Early diagnosis of breast cancer increases the survival rate and gives additional treatment options to improve cure effects [11]. Thus as early as possible detection of breast cancer is a very important challenge for screening interventions and programs. However, image interpretation procedures have still limited the precision or accuracy despite the significant development of continually refined imaging systems such as: mammography, tomosynthesis, MRI, USG. The error rate in mammography, i.e. commonly used method of screening for breast cancer [1] can be even up to 30% for false positives and 20% for false negatives [24]. Among many reasons there are relative and unstable pathology patterns, variable technological conditioning of imaging, and limited expression of diagnostically important symptoms of image features. Other screening methods, e.g. screening MRI is recommended only to women with high risk [1] due to many false-positive findings, a considerably longer duration, and the high cost of MRI examination (in relation to mammography) [25]. Screening USG is also indicated only in high-risk patients who cannot tolerate MRI [1]. Because of these limitations, computer-aided diagnosis systems (CAD) have proved to be useful in the support of image interpretation carried out by radiologists. These systems minimize the mistakes of radiologists in decision-making for a large number of screening cases interpreted in a short time [9]. However, efficiency of CAD systems is still insufficient, in particular in automated detection of subtle spiculated pathological findings manifested in a form of SMs and ADs. For example, the commercial CAD systems (R2 Image Checker and CADx Second Look) detect ADs with sensitivity less than 40% and approximately 1.0 FPI [31]. It is also worth noting that the R2 Image Checker operates with 86% sensitivity at 0.24 FPI for generic collection of SMs, but for the subtle SMs the sensitivity is only approximately 50% [31]. Thus, research directed to effective spicule descriptors design is still ongoing.

Generally, the breast tissue in the mammographic image is visible as a directionally oriented image texture which normal pattern converges not always uniformly toward the nipple [29]. Moreover, imaged breast tissue contains several other piecewise linear structures of directionally varied texture caused by ligaments, ducts, blood vessels and other. It hinders often subtle signs of cancerous tissue and makes accurate recognition of spiculated pathologies extremely difficult. Dominant but hardly perceptible feature of malignancy are SMs and ADs. Moreover, abnormal spicules of ADs are more subtle and slightly differ from several types of other spicules. Briefly,

a) a manifestation of probably malignant lesion, most often recognized as AD (excluding postsurgical scar), is characterized by relatively short (<1cm) radiating lines, emanated from invisible dense center; the size is approximately 5-50mm [8,11]; moreover, the spicule width and length are not reproducible by radiologists [33];

b) cancer-like lesion (e.g. fat necrosis) contains radiating lines, but shorter and wider than malignant spicules; there is round radiolucent area in the center [11];

c) SM is distinguishable by radiating lines from well-visible central mass; these linear structures retain the same width along their entire length; probably SMs are identified with slightly higher number of spicules than AD; moreover, the size of spicule and the length of the major axis of the central region of SM are reproducible with high probability [33];

d) benign lesion (e.g. radial scar) is observed with long thin spicules (>1cm) arranged in beams, related to bright central mass or disproportionately small center [8,11].

However, cardinality of the compact radiating spicules of malignant characteristics of appearance, their size and angular distribution, linearity, form of intersections and correlation with surrounding tissue are highly unstable and relative case-dependent. Thus, precise and consequent analysis of oriented-texture patterns is really essential challenge of CAD system for mammograms. Mostly, the research studies are based on spicule directionality modeled with piecewise linear structures: the analysis of local oriented edges [21], statistical analysis of a map of pixel orientations [20], skeleton analysis [23].
Otherwise, top-hat partial reconstruction was used to enhance the spicules [19], the Dixon and Taylor line enhancement algorithm with a line strength map (as their result) indicated the potential presence of oriented lines [39] or the integrated intensity along oriented lines (using generalized DTLO) [4]. Moreover, the extraction of linear structures was realized by estimation of a mean curvature sign and the concentration index [26], Gabor filtering and phase portrait [10,29] or a curvilinear structure ridge detection [15,38]. Another approaches concentrated on non-directional properties of spiculated lesions include: i) the intensity distribution of pixel context, matched to symptom template by thresholding procedures [27], ii) a fractal texture analysis [5,13,36]. Researchers emphasize the importance of using various image domain representation to linear structure recognition – the Radon [32] and Hough domain [23] or multiscale domain including discrete tensor wavelets [30], streeble complex filtering [35], Gabor wavelets [28] or dual-tree complex wavelets [6,7]. Most of the above approaches are closely identified with unacceptable number of FPI at the high enough sensitivity of spicule detection.

The fundamental reason for limited detection efficiency of the spicules is difficult extraction of representative descriptors identifying all significant properties of analyzed abnormalities, especially reliable local variation of texture directionality and complementing conditions of intensity distribution of approximated center of piecewise structure convergence.

Main objective of this study was to find more efficient descriptors of spicular findings. Intended effect is improvement of breast cancer recognition according to CAD concept. However, other applications such as content-based retrieval of mammograms, image fidelity enhancement or content normalization are possible.

The novelty of this paper is designed and optimized computational descriptors of abnormal spicules in mammograms. The most important advantage is specific modeling of directional activity disorders done in local centres of possible spiculations. Our approach was based on last years achievements of harmonic analysis and wavelet theory in sparse modeling of signal properties. Initial study was directed to optimize transformed image representation with criteria of multiscaleability, shift and rotate invariance, low redundancy, and high-enough directional resolution [17,18]. As the result of previous study, computationally tractable CW was found to be useful identifying all significant properties of analyzed subtle structures. The phase information of CW complementing directional distribution of signal energy is really useful to capture invariantly subtle data dependencies. To the best of our knowledge, none specific descriptors of spicules were designed in complex wavelet domain.

Additionally, the original proposition is extended diagnostic model of spiculated findings in mammogram, including background of spicule appearance, spicule profile and convergence that have been formalized in presented set of descriptors. Based on previous studies and experimental experience, representative database of abnormal and normal cases, optimization process of descriptor design and implementation were carried out to maximize specificity and sensitivity of spicule description for breast cancer detection. The discussion of achieved experimental results have been concluded with important issues of further development.

2. Method of spicule recognition

Abnormal spicules are commonly identified with characteristic background of spicule appearance - containing brightness gradient field and exhibiting a consistent radial structure with approximated center of tumor. The fundamental criterion of optimized AD detector should be ability to distinguish malignant spicules of specific directional activity between textured background appearance with underlying dominant orientation and such structures as ligaments, ducts etc. For this purpose, abnormal spicules of high directional activity in respectively small (local) image area have been modeled. Consequently, adjusted numerical descriptors dedicated to local centers of multidirectional activity were assumed to be particularly important in malignant spicules recognition.
2.1. Directional activity of spicules

Abnormal deformation of breast tissue is characterized as clearly and specifically increased variability of local tissue density. Malignancy in breast tissue (manifesting itself as radiating spicules) causes distortion of the normal oriented breast tissue, i.e. mainly towards the nipple. It results in increased directional activity interpreted as local variation in the angles of approximated line segments representing spicules. Directional activity was defined as the number of different directions of local spicule group, i.e. recognized piecewise linear structures convergent to the approximate center. Such activity is assumed to be decisive sign of abnormality. Berks et. al. estimated such activity with gradient based orientation histograms and patch-oriented models [6]. In [3] analyzed spicules were construed as disruption of the normal tissue in form of the distortion of architectural texture specificity derived from gradient differences between malignant and benign tissue. Spicules show a characteristic microstructure of thin, narrow, brighter, slightly curved lines with important length to width relation and higher brightness level than background, not always distinctly visible. Spicular structure is often considered as a ridge-like, Gauss declining gray level profiles modeled as piecewise linear singularity (with a discontinuity along the line) [19]. Gradient vectors concentrate on the ridge of the line.

Our idea focused on the construction of effective descriptors concentrated on the local directional activity estimates. Hence, specific orientation distribution of spicules can be defined. Consequently, background of spicule appearance was modeled in scalable multidirectional domain with preferable high energy compaction along piecewise linear gradients. On the basis of imaged normal breast tissue and training set of digital mammograms containing spiculated lesions, difficulty in stable estimation of the level of piecewise smoothness, linearity and regularity of the directional structures was noticed. Selectively discriminating approximates of spicules were studied to represent efficiently subtle line singularities of diverse directionality concentrated in the local image area. Accurate recognition of such symptoms possibly distinctive from the background of directionally more homogeneous architecture of surrounding tissue was the operating criterion of optimized representation.

However, the essential difficulty is the extreme variability or instability of symptom signatures in the acquired signals. Consequently, the problems of spiculated structures recognition become quite often ill-posed. Therefore, more emphasis should be placed on solution of adaptive regularization framework. It works in most circumstances. Such solution can be achieved with respective image and specific representation of spicule according to the following requirements:

a) for differentiating feature extraction:
   — precise description of local directionality including diversity, selectivity and energy distribution of nonstationary random line singularities,
   — changes or deviations from the general trend of directional (angular or phase) tissue characteristic,
   — multi-scalability due to different lesion size and spicule extension,

b) for effective spicule recognition:
   — sufficient directional resolution in higher dimensions,
   — shift-invariance,
   — information sparsity and selectivity, i.e. significant energy compaction across multi-directional line singularities.

The main problem of our research was the formalized approximation of diagnostic lesion model with spicules. Precise representation of local directional activity in mammograms was studied and estimated for examined structures. The criterion of such optimized approximation was as high as possible recognition efficiency of malignant ADs in the actual images. Moreover, the attention has been focused also on nondirectional properties of analyzed findings. There are a dynamics of local orientation distinction against dominating tissue background, an estimate of convergence point of spiculated structures. It seems to be helpful in obtaining higher sensitivity of spicule recognition method.
2.2. Directional representation in complex wavelet domain

It is well known that multiscale analysis with wavelets is useful in a wide range of applications. This is possible due to an extremely efficient representation for regular (near smooth) signals containing informative singularities. Such compatibility with non-stationary random processes makes wavelets a powerful tool for analyzing of biomedical signals. Wavelet transforms facilitate highly complex time-frequency characteristics of signal features [2,37]. Interscale dependency, approximation rate and subband correlations are useful to characterize the majority of representative features for symptom detection [12].

Discrete wavelet transform (DWT) using Mallat’s fast algorithm with adjusted orthonormal or biorthogonal bases provides well localized spatially compact, non-redundant representation of the signal. However, wavelet coefficients tend to oscillate around singularities with shifted impulse response. Because of down sampling, aliasing and shift sensibility limited recognition efficiency was noticed. Tensor extension to 2-D wavelets retains the disadvantages and produces low directionality with limited angular resolution of multidirectional singularities in decomposed data domain.

Preferable forms of wavelet-based analysis should primarily produce high directionality with unlimited angular resolution of multidirectional singularities in decomposed data domain, but also deal with 2-D nonseparable kernels and redundant domains of transformed images. CW give an important compromise between significant data dependency in expanded domain and independence of pathology pattern due to acquisition conditions [22]. Despite the increased computational complexity, such analysis reduces aliasing and decreases invariance limitations increasing the precision of determining directionality of line singularity characteristics. There are several alternative forms of CW transform, including preferable in the analysis form of DT-CWT [34]. Providing perfect reconstruction and computational efficiency with linear time complexity, DT-CWT possesses important advantages for specific pattern recognition as follows:

- a) nearly shift invariant magnitude with linear phase encoding of signal shifts, good rotational invariance; small shifts in the input signal preserves the distribution of energy between wavelet coefficients at different subbands and scales; in consequences data processing is almost free of aliasing (i.e. aliasing artifacts, manifested as irregular representation of edges and ridges);
- b) good directional selectivity of linear singularities achieved with nonseparable directional wavelets; angular resolution is determined by six subbands (bandpass sub images of complex coefficients) produced at each level; directional subbands are approximately oriented at angles of ±15°, ±45°, and ±75°;
- c) perfect reconstruction of DT-CWT implementation based on relatively short, linear-phase, separable SOI FBs in Mallat’s fast algorithm;
- d) considerably low redundancy, independent of the number of scales and equal to 4:1 for images.

Consequently, magnitude and phase distribution analysis of complex wavelet coefficients across directional subbands and decomposition levels is useful to select local directional activity (diversity, deviation from the dominant trend or distortion of the most signal characteristics) in detected areas of interests. Reliable approximation of precisely oriented singularities is possible because of complementary phase information. An optimized selection of high-energy centers of image directional activity, followed by the analysis of spicule orientations with high angular resolution, was a consequence of assumed signal models and criteria of effectiveness.

2.3. Recognition procedure
General research outline of optimized method to recognize spiculated findings in mammograms consists of four key stages with several intentional functionalities (Fig. 1) [16, 18]:

a) preprocessing to enhance imaged content, reduce diagnostically irrelevant noise and decrease influence of less informative low-frequency components of imaged tissue (a sequence of adapted low-pass and high-pass filtering, normalization, local contrast enhancement);

b) approximation of regions of interests with increased directional energy concentration; for this purpose a set of Gabor filters is designed accordingly to the adjusted single spicule model; the ROIs with potential findings are chosen on the basis of estimated probability map (described exhaustively in [16]);

c) recognition of abnormal breast tissue imaged in selected ROIs; extraction of spicule-distinction features was first of all based on local directional activity of line-singularities in various transformation domains (image domain, 2D polar Fourier domain [16] and multiscale domain [18]), matched to a thoughtfully adopted model of spiculated lesion;

d) feature vector formulation according to assumed spicular model and subsequently support vector machine classification to detect subtle directional spiculated lesions [16].

![Fig. 1. Scheme of proposed detection method for subtle spiculated findings in mammograms. The main issue of this paper, i.e. feature extraction in complex wavelet domain is marked in bold frame.](image)

Reported study has been focused solely on abnormal spicule recognition from selected regions of interests in complex wavelet domain (Fig. 1) using both magnitude and phase information. Designed algorithm was based on two fundamental assumptions: i) dominant role of reliable directional activity estimate in distinction of star-shaped structure of spicules; ii) attractive properties of complex wavelets to accurately represent the distinguishing activity of line-singularity distribution. Implemented algorithm of spicule recognition in complex wavelet domain (Fig. 2) fulfills the following requirements:

a) initial characteristics of texture orientation activity by Gabor directional filtering to estimate a magnitude map and an orientation angle map;

b) complementing characteristics of textured background of spicule appearance: instead of statistic textural features calculated in enhanced and normalized image domain, acquisition-invariant and case-invariant domain of reduced redundancy and energy compacted across oriented structures were examined; a level of oriented structure distinction in relation to surrounding unoriented texture was determined from the Gabor magnitude map;

c) precise description of each spicule of interests considered as linear (piecewise linear or curvilinear) structure modeled with line-singularity; especially, a distribution of spatial and directional locations in the complex wavelet domain is preferable for detailed analysis of the orientation angle map;
d) essential fine estimation of directional activity sensitive to a concentrated locally radiating spicules of clearly diversified orientation: in details, the representatives of local and global direction deviations are calculated to estimate dominant orientation distributions of specific singularities; it is a form to express and distinguish the significant concentration and composition of abnormal spicules; the superposition of Gabor orientation discriminant and clarification of the complex wavelet phase was optimized;

e) measure of oriented structure convergence to the hypothetical center: local analysis of singularity orientation distribution in domain of high directional resolution; the complex wavelet domain is preferable.

Fig. 2. Flowchart of the feature vector extraction in complex wavelet domain useful in the recognition of spiculated structures in mammographic image. Gabor filtering was used in to carry out mammogram analysis only on image area of high amplitude, i.e. area containing potential spicules.

2.3.1. Characteristics of texture orientation activity

Initial analysis of mammogram ROIs was directed to texture orientation extraction through determining the maximum magnitudes of directional filtering and corresponding dominant orientations at each pixel of interests. Following Rangayyan et al. [29] we have found that a bank of Gabor filters of different orientations (adjusted to approximated model of line segments of interests) is really useful to analyze locally oriented structures in mammogram. The Gabor image decomposition into the magnitude map and orientation angle map provides robust performance in the presence of noise and variations in scale. Gaussian envelope multiplied against the sinusoidal carrier of the Gabor filter kernel is oriented at the angle \( \theta_k \) according to the formula:

\[
G_k(x, y) = e^{\frac{-1}{2} \left( \frac{(x')^2}{\alpha_x^2} + \frac{(y')^2}{\alpha_y^2} \right)} \cos(2\pi f x')
\]  

with rotation angle range fixing filter support as \( x' = x \cos \theta_k + y \sin \theta_k \) and \( y' = y \cos \theta_k + x \sin \theta_k \), \( \theta_k \in \left[ -\frac{\pi}{2}, \frac{\pi}{2} \right] \). The Gabor magnitude map of image \( I(m, n) \) contains the highest filtering magnitude response \( G^{(d)}_l(m, n) = \max_k \left| G^{(d)}_{l,k}(m, n) \right| \) where \( G^{(d)}_{l,k}(m, n) = g_k \times l(m, n) \). Respectively, the Gabor orientation angle map is determined at each pixel as \( G^{(\theta)}_l(m, n) = \theta_{\arg \max_k} \left| G^{(d)}_{l,k}(m, n) \right| \).

Empirically optimized parameters are: i) a number of necessary Gabor filters to assure enough resolution of studied texture orientation: \( k = 1, 2, ..., 180 \), ii) the values of \( S_x = 4 \) and \( S_y = 8 \) defining the width and extending the Gaussian envelope; the elongation in the \( y \) direction is to double in comparison to the extent in the \( x \) direction; it means the quite moderate orientation precision due to the relatively little elongated filter shape, reflecting
only averaged model of strongly diversified spicule length to width ratio for abnormal ADs. The fixed ratio $\frac{S_y}{S_x}$ is generally enough to detect directional linearity and too vague to identify the nature of spicules. However, the crucial benefit is increased detection sensitivity of curved, curvilinear quasi spicules at the initial stage of selected tissue analysis. On the other hand, we verified the determination rule of the period of cosine term $f = \frac{0.7}{\tau}$, where $\tau = \frac{2.35}{S_x}$ is full-width at half-maximum of the Gaussian term along the $x$ axis. Consequently, $f = 0.075$ was used to reduce cosine oscillations along elongated Gaussian slope and model respective Gabor filter response to be ridge-like, according to coarse spicule characteristics. It is worth noting that high magnitude $G^{(s)}_i(m,n)$ indicates distinctly oriented texture at the pixel $(m,n)$ while corresponding low magnitude means lack of highly oriented structure. Thus, the relation of minimum to maximum of the Gabor magnitude map characterizes a dynamics of local orientation distinction against dominating tissue background. Simple normalized scalar feature indicating point dynamic variety was defined by formula:

$$PDV = \frac{\min_{m,n} G^{(s)}_i(m,n)}{\max_{m,n} G^{(s)}_i(m,n)}$$  \hspace{1cm} (2)$$

In this way it is possible to complement the characteristics of textured background of spicule appearance and the possible presence of visual texture density in the center of mammographic findings.

2.3.2. Precise directional description of active texture

The Gabor orientation maps of mammogram regions were used for further, more precise extraction and efficient recognition of abnormal directional distortions compared to dominant architecture of mammary gland. The distribution of angular singularities, determined in locally estimated $G^{(s)}_i$, reflects diversity of orientation trends by non-random possible abnormal pattern. Multiscale and multidirectional decomposition of the Gabor orientation angle map was realized with the DT–CWT implementation of 4 scale and 6 directional subbands. General form of the transformation is

$$\Psi_s(\mathbf{r}; k) = \int G^{(s)}_i(\mathbf{r}) 2^{2s} \theta^s(2^s \mathbf{p} - \mathbf{r}, k) d\mathbf{p}$$  \hspace{1cm} (3)$$

with complex wavelet $\psi_c$, scale index $s = 1,...,4$, subband index $k = 1,...,6$ and image position $\mathbf{r} = [m,n]$. The advantages of applied image decomposition for analysis of spiculated ADs in mammograms were listed below.

- An analytically profitable concept of complex wavelets refers to the useful benefits of analytical signals and best forms of implementation of wavelet transformation. Complex wavelet $\psi_c(2^s \mathbf{r}, k) = \psi_r + \psi_i$ (and complementing, respective complex scaling function) is a superposition of two real wavelets, shifted in phase by $\pi/2$, i.e. one wavelet is approximated Hilbert transform of the other. Thus, complex wavelet is potentially analytical signal $\psi_c = \psi_r + \psi_i$ what implies that the Fourier transform of the $\psi_c$ is (almost) identical to zero at any negative frequency. According to DT realization, both wavelets $\psi_r$ and $\psi_i$ separately generates the orthonormal or biorthogonal basis. Consequently, moderately redundant (four-times in 2-D) tight frame of CWT is formed providing good directional selectivity and approximate shift invariance of the amplitude with the linear phase encoding resulting in reduced aliasing [14,34].

- The CWT implementation, based on two sets of FB for analysis/synthesis, assures limited computational complexity and perfect reconstruction with short linear-phase filters, filtered sample decimation and successive decomposition of scaled domain. Theoretical and practical issues translate to only approximately analytic form of $c$ and thus, only approximate advantageous features as above. Quality of this approximation
is fundamental criterion of optimal FB design. Applied complex wavelet decomposition was determined with selected empirically FBs: Near-Symmetric FB (13,19 tap filters) for first tree stage and Q-Shift, i.e. Quarter Sample Shift Orthogonal (14,14) for next decomposition stages. The Matlab implementation of the dual-tree complex wavelet transform [22] was used. The number of scales was experimentally adapted to size of analyzed images. A 4-level decomposition of the selected ROIs with 6 directional subbands results in 24 complex coefficient matrices \( W_z(m,n,k) \) to be analyzed.

- Analytical capabilities of complex representation determined for Gabor orientation map are mainly due to invariant expression of further clarified: i) directional characteristics of curvilinear singularities in magnitude domain \( |W_s(r,k)| \) (directional diversity) – 24 matrices (for each scale and all subbands) of the wavelet amplitude map were obtained; ii) detailed description of selected singularity orientation based on block-based division of each subband in phase transform domain \( LW_s(r,k) \) (more precise orientation correction in local space) – similarly 24 matrices of the wavelet phase map were determined.

2.3.3. Local direction deviations of potential singularities

In order to reduce false positives (more precisely, diagnostically insignificant image structures), initial selection of relevant representatives (potential orientation singularities) in each directional subband was carried out in two ways:

a) based on four binarized matrices created for each scale \( s \) by appropriate subsampling (every \( 2^s \) pixel) of Gabor magnitude map and binarization with experimentally adopted threshold \( T_1 \); as a result, only dominant coefficients (with higher amplitudes than \( 0.35 \max |G^{10}(m,n)| \)) in each subband were retained for further analysis; achieved frame of scaled spicule maps indicates the location of potential singularities for feature extraction in \( LW_s(r,k) \) domain;

b) by thresholding magnitudes of CWT, i.e. \( |W_s(r,k)| \); as result of fixed \( T_2 \) application, only significant coefficients (i.e with amplitudes higher than \( 0.3 \max |W_s(r,k)| \) in each subband) were retained and normalized; as a consequence, the map of significant directions of potential speculated structures was formulated.

Only after the initial selection of relevant representatives a proper 16–block analysis is possible to precise directional activity. For this purpose, a map of local direction deviations (LDD) in a form of 4x4 matrix series based of scaled subdivisions of each of 24 decomposition subbands has been calculated. It presents the distribution of the coefficient average phase separately for each subband block across directions and scales, determined as follows:

a) division of each subband coefficients for all scales and directions into 16 blocks: \( B_p \), where \( p = [i,j] \); \( i,j = 1,...,4 \) (each subband is a cell of 4x4 matrix of blocks); it corresponds with 64 x 64, 32 x 32, 16 x 16, 8 x 8 block \( B_p \) sizes in the increased scale order;

b) determining averaged phase of nonzero coefficients for each block \( B_p \) of successive directional subbands across scales as \( \alpha(s,k,p) = \text{mean}(LW_s(B_p,k)) \), where \( s = 1,...,4 \), \( k = 1,...,6 \) and \( p = [1,...,4] \), normalized to the average phase maximum of all blocks belonging to the direction \( k \) across scales;

c) determining ”active directions” by estimation of phase singularities of specified directional characteristics, approximated by thresholded subband blocks (curvilinear, with the spatial accuracy limited by block size in successive scales); for that purpose the values of \( \alpha(s,k,p)_{\text{norm}} \) were binarized with experimentally chosen global threshold \( T_3 = 0.2 \);

d) calculation of integrating 16-element matrix \( LDD_{14x4} \) by summing indicated blocks of ”active directions” across all scales and directions taking into account local structure of 4 x 4 subband subdivision.
such that indicator function $I(i,j) = 1 \leftrightarrow \alpha(s,k_i[i,j])_{\text{norm}} > T_3$ while $I(i,j) = 0$ in other cases; the possible maximum value of the matrix $LDD$ is 24, which means that the representative block is active for all the subbands of complex wavelet decomposition; the 16 values of the matrix $LDD$ form a features vector used to recognize the specific local directional activity of tissue (differentiating of directions and indirectly the radiating of piecewise linear segments); thus $LDD$ is the significant factor affecting ADs differentiation.

2.3.4. Global direction deviations of potential singularities

The gained map of local direction deviations provides information on the image angular selectivity and image directional resolution. In order to generalize the directional information, a 6-elements vector of global direction deviations (GDD) was computed as follows:

$$GDD(k) = \frac{\sum_r N_{\text{set}}(r,s,k)}{(\sum_r N(r,s,k))^2} \times \sum_r |LW_3(r,k)|$$

where $N_{\text{set}}$ = a number of selected coefficients of significant directions (taking into account spicule maps and $LDD$), $N$ = the number of all coefficients across $k$ direction subbands of all scales. This allows for an examination of the size of phase shifts and thus, determining how large are deviations of potential singularities localization (in image) in relation to the filter support. On this basis, it is possible to clarify the angular selectivity and also determine the localization of potential singularities (in this case, pathological spiculed structures) with higher directional resolution.

2.3.5. The point of convergence of spiculed structures

One of the proposed model elements is the convergence point of spiculed structures. In case of AD it is an area (not the point) and very often invisible. Clarifying information on the actual image texture orientations by adding the average value of coefficient phase to generally accepted value of characteristic angles for each subbands $\theta_k = \pm 15^\circ$, $\pm 45^\circ$, $\pm 75^\circ$ it is possible to determine quite precisely the dominant directions in image. The average value of phase coefficients that may be helpful in the convergence point estimation is calculated across $k$ directions subbands of all scales:

$$CP(k) = \frac{\sum_r N_{\text{set}}(r,s,k)}{(\sum_r N(r,s,k))^2} \times \sum_r LW_3(r,k)$$

according to the meaning of data structures and parameters defined previously.

3. Results

In order to investigate the potential of complex wavelets to represent meaningful image directional structures, the appropriate regions of interest were indicated by hand in mammograms available in the Digital Database for Screening Mammography (image resolution 43.5 – 50µm). The criteria used to form training and testing datasets of mammogram ROIs (512x512pixels) were the capture of whole diagnostically representative lesions and even surrounding tissue. Only partially represented findings lose complete characteristics included in proposed model. The set of test ROIs contains 164 cases with ADs and 116 normal cases. The ROIs were analyzed according to procedures defined in section 2.3 to extract the features optimized with criteria of well abnormal tissue differentiating. The calculated feature vectors was classified with SVM using standard linear kernel and regularization (PRTools – Matlab Pattern Recognition Toolbox, ver. 4.1.4 11-Oct-2008). To reduce an influence of training–testing data separation on the classification results, one-to-all cross-validation procedure was used. The summary results of conducted experiments were presented in Table 1. The highest
sensitivity and accuracy have been achieved for complete numerical model of malignant spicules designed according to proposed diagnostic model of complete AD description.

<table>
<thead>
<tr>
<th>Features set</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD method</td>
<td>65.85%</td>
<td>63.79%</td>
<td>65.00%</td>
</tr>
<tr>
<td>DD_CWT method</td>
<td>68.29%</td>
<td>55.17%</td>
<td>62.85%</td>
</tr>
<tr>
<td>LDD</td>
<td>72.56%</td>
<td>70.69%</td>
<td>71.79%</td>
</tr>
<tr>
<td>LDD + GDD</td>
<td>76.22%</td>
<td>66.38%</td>
<td>72.14%</td>
</tr>
<tr>
<td>LDD + CP</td>
<td>72.56%</td>
<td>70.68%</td>
<td>71.78%</td>
</tr>
<tr>
<td>LDD + GDD + CP</td>
<td>75.61%</td>
<td>66.38%</td>
<td>71.78%</td>
</tr>
<tr>
<td>LDD + PVD</td>
<td>73.78%</td>
<td>71.55%</td>
<td>72.86%</td>
</tr>
<tr>
<td>LDD + GDD + CP + PVD</td>
<td>76.83%</td>
<td>68.96%</td>
<td>73.57%</td>
</tr>
</tbody>
</table>

Table 1. The final results of spicule recognition method based on complex wavelet-based descriptors (LDD) defined in section 2.3. This results are compared to the recognition efficiency of the set of differentiating features proposed in previous studies [16,18] and based on the dataset of 280 mammographic ROIs 512x512 pixeles. The input image in DD method – the mammographic image; in DD_CWT method and LDD – the Gabor orientation map.

The mammographic images are very noisy and preprocessing is very helpful to remove diagnostically irrelevant information hindering recognition. Particularly, Gabor filtering which leads to enhance directional structures in image was found extremely useful. The best result of mammo-spicule recognition has been obtained for directional activity matrix calculated on phase map, taking into account the information from the spicule map and significant directions map. The accuracy of pathological mammo-spicule recognition using the LDD is about 72% (the sensitivity – about 72.6% whereas the specificity - about 70.7%). This is noticeably higher efficiency in comparison to the results obtained for the proposed descriptors sets in previous studies.

Extending the vector of well-differentiating features by adding descriptor PVD, specifying the point dynamic variety, is possible only by approximation of the texture characteristics from the center of mammographic findings and their background. For such constructed feature vector the sensitivity of abnormal cases detection has increased to about 74%, whereas the specificity to 71.5%. With this approach, the accuracy of ADs recognition was improved by more than 1%.

It is worth to underline the fact, that clarified description of the directional characteristics of analyzed pathological findings has allowed in most cases to obtain even better recognition sensitivity. The analysis of directional activity in small blocks, additional estimate of the spicule convergence point CP and the global direction deviation GDD define higher precision directional resolution in image and contribute to increased accuracy (more than 73.5%). This allows more accurate recognition of architectural distortions. However, this allows to simultaneously detect more false positives, which was confirmed by the obtained results – the sensitivity has significantly increased to 76.83% but with noticeable decrease in specificity - by nearly 2%.

It is very likely that increased number of false positives was caused by proposed, more sensitive description of directional texture characteristics to increase spicule detection sensitivity. However, it contributes also to the capture of other piecewise linear structures – being not only the imaged actual spiculated structures but also the imaged ligaments, ducts, blood vessels and others (Fig. 3).
Fig. 3. Selected results of spicule recognition – ROIs with normal breast tissue: a) four false positive cases, b) indicated causes of incorrect structure recognition shown in Fig.3a, c) four cases of well classified normal tissue which mimics certain tissue properties of ADs.

Additional difficulty in recognition process of AD lesions is locally dominant dense tissue – in contrast to the fatty tissue context of findings, in which some pathological findings are much easier to recognize. The dense (connective) breast tissue sometimes appears as a solid, irregular white area which mimics the lesions looking a lot like irregular dense tissue with accidentally distributed directions of randomly formed quasi-linear segments (Fig. 3). Otherwise, the pieces of dense tissue overlapping spicules of AD could deform findings, increase its width or generally distort directional characteristics of radiating spicules (Fig. 4). Thus, optimized representations and proposed descriptor parameters (e.g. shape and size of Gabor filter support) adjusted to assume diagnostic model and mammogram specificity could not capture such deformed real structures or differentiate artificial and real findings.

Fig. 4. Selected results of spicule recognition – ROIs with abnormal breast tissue: a) four false negative cases, b) indicated causes of unrecognized spicules shown in Fig.4a, c) four cases of true positives.

Moreover, it seems important to clarify the descriptors defining nondirectional elements of proposed model of spiculated lesions. Precise investigation of convergence lines area, and accurate characteristic of background surrounding analyzed pathological structures in mammogram should lead to better results.
Next, proposed model optimization should be extended to more precise analysis of directional description to differentiate probable spicules more accurately. According to our research experience, complex wavelet domain tends to be useful in further adjusting of the AD model features, especially considering phase and energy distribution across scale and location. Phase information indicates the offset of directional features within the support region of a wavelet coefficient. Phase linearity is used to infer the offset of the edge/ridge, relative to the coefficient location, at given scale. Such information is useful to calculate more precisely the angle of directional energy in the vicinity of the coefficient. Although magnitude responses of each these subbands can be used to infer feature orientations, complex phase representation allows increase angular resolution of the singularity distribution.

4. Discussion

The subject of the study was to improve the efficiency of pathological spicules description and consequently recognition by using the complex wavelets as an effective tool to reliably analyze the directional orientation of piecewise structures. The proposed scheme assumes:

a) effective use of phase information for more precise and invariant characteristics of spicule recognition,
b) narrowing the area of analysis to directional activity centers, designated based on Gabor amplitude map and CWT magnitude map,
c) increasing the recognition efficiency by using supplementary information, describing the distribution of spicules and characteristics of their background in relation to the formulated pathological model-spicule.

The obtained results confirm the usefulness of the complex wavelet domain, and the adopted solutions tend to be an effective tool to verify the suspected findings (abnormalities) of active local directionality. This tool may be used to optimize the mammographic CAD systems. However, due to the complicated, nonstationary and relative expression of noisy mammogram piecewise structure, and equivocal projection of real diagnostic structures to 2D image plane the resulting effect of ADs recognition in mammograms is still unsolved problem. The pragmatic difficulty lies in increased number of indicated false positives, while the sensitivity of spicule detection is rising. Further development should address both aspects: a more objectified and complete – at the level of details – diagnostic model and its effective numerical representation based on spatial, scalable and multidirectional model of findings and analysis of their planar projections to 2D image plane. Thus, 3D complex wavelets are welcomed to be applied.

Acknowledgments. This publication was funded by the National Science Centre (Poland) based on the decision DEC-2011/03/B/ST7/03649.

References