Wan, T., Zhang, W., Zhu, M., Chen, J., Achim, A., & Qin, Z. (2017). Automated mitosis detection in histopathology based on non-gaussian modeling of complex wavelet coefficients. *Neurocomputing*, *237*, 291-303. https://doi.org/10.1016/j.neucom.2017.01.008

Automated Mitosis Detection in Histopathology Based on Non-Gaussian Modeling of Complex Wavelet Coefficients

Tao Wan[a,*], Wanshu Zhang[a], Min Zhu[b], Jianhui Chen[c], Alin Achim[d], Zengchang Qin[e,*]

[a] *School of Biomedical Science and Medical Engineering, Beihang University, Beijing, 100191, China*
[b] *Pathology Department, Karamay Central Hospital, Karamay City, Xinjiang 834000, China*
[c] *No. 91 Central Hospital of PLA, Henan, 454000, China*
[d] *Visual Information Laboratory, Bristol University, Bristol BS8 1UB, UK*
[e] *Intelligent Computing and Machine Learning Lab, School of Automation Science and Electrical Engineering, Beihang University, Beijing, 100191, China*

Abstract

To diagnose breast cancer, the number of mitotic cells present in histology sections is an important indicator for examining and grading biopsy specimen. This study aims at improving the accuracy of automated mitosis detection by characterizing mitotic cells in wavelet based multi-resolution representations via a non-Gaussian modeling method. The potential mitosis candidates were decomposed into multi-scale forms by an undecimated dual-tree complex wavelet transform. Two non-Gaussian models (the generalized Gaussian distribution (GGD) and the symmetric alpha-stable (SαS) distributions) were used to accurately model the heavy-tailed behavior of wavelet marginal distributions. The method was evaluated on two independent data cohorts, including the benchmark dataset (MITOS), via a support vector machine classifier. The quantitative results shows that the bivariate SαS model achieved superior classification performance with the area under the curve value of 0.82 in comparison with 0.79 for bivariate GGD, 0.77 for

*Corresponding authors. Tel.: +86 010 82316875; fax: +86 010 82316875. E-mail address: {taowan, zcqin}@buaa.edu.cn (T. Wan, Z. Qin).
This work was partially supported by the National Natural Science Foundation of China under award Nos. 61305047 and 61401012.

univariate SαS, 0.72 for univariate GGD, and 0.59 for Gaussian model. Since both mitotic and non-mitotic cells appear as small objects with a large variety of shapes, characterization of mitosis is a hard problem. The inter-scale dependencies of wavelet coefficients allowing extraction of salient features within the cells that are more likely to appear at all different scales were captured by the bivariate non-Gaussian models, leading to more accurate detection results. The presented automated mitosis detection method might assist pathologists in enhancing the operational efficiency and productivity as well as improving diagnostic confidence.

Keywords:
histopathology; breast cancer; mitosis detection; non-Gaussian model; wavelet.

---

1. Introduction

According to the World Health Organization, breast cancer is the most frequently diagnosed cancer and is the leading cause of cancer death among women [1, 2, 3]. A breast biopsy is a diagnostic procedure that can definitely determine if the suspicious area is malignant (cancerous) or benign (non-cancerous). During biopsy, samples of tissue are removed from breast to produce stained histology slides, which are observed under microscope and graded by pathologists. Based on the Nottingham Grading (modified Bloom-Richardson grading) System [4], mitotic count is one of the main parameters in breast cancer grading as it gives an assessment of tumor proliferation and aggressiveness of breast lesions. Manual counting of mitosis is a tedious process and often prone to inter- and intra-reader variations. A recent concordance study for quantifying the magnitude of diagnostic disagreement among pathologists on breast biopsy specimens reported 27.7% overall disagreement between the individual pathologists' interpretations and the expert consensus-derived reference diagnoses [5]. A mitosis has four main phases of prophase, metaphase, anaphase and telophase, and exhibits highly variable appearance. For instance, a mitosis in telophase contains two distinct nuclei but does not yet divide into full individual cells, and must be counted as one single mitosis [6]. Most commonly in the hematoxylin and eosin (H&E) section, mitotic nuclei manifest themselves as hyperchromatic objects. However, many non-mitotic objects, such as apoptotic and necrotic nuclei, and other artifacts from the slide preparation and acquisition, can

also have a similar appearance. The major problem is that there is a low density of mitosis in a single histological image. Hence, manual identification of mitotic cells is a difficult task even for expert pathologists to make a distinction [7]. Since most current mitosis counting approaches are based on the subjective opinion of pathologists, there is clearly a need for development of an automatic mitosis detection algorithm that works with the routinely clinical practices.

With the recent advent of whole slide digital scanners and advances in computational power, it is now possible to use digitized histopathological images and computer-aided diagnosis (CAD) algorithms for disease detection, diagnosis, and prognosis prediction to complement the opinion of the pathologist [7, 8, 9, 10, 11]. Recently, a few CAD techniques have been developed to automatically detect and classify mitosis and non-mitosis using various image features [12]. For example, Sommer *et al.* [13] employed intensity, shape (e.g., circularity), and texture features (e.g., Haralick, statistical geometric features) to perform mitosis classification on a pixel basis. Tashk *et al.* [14] introduced an automatic mitosis detection method using completed local binary patterns based on a pixel-wise classification. In their recent publication [15], a CAD system was presented based on a teaching-learning-based-optimization method to reduce false positives in mitosis detection. Tek [16] employed color, binary shape-based, Laplacian, and morphological features to perform mitosis classification. Irshad *et al.* [17, 18] adopted morphological and multi-channel statistical features, or co-occurrence features, run-length features, and scale-invariant feature transform features to characterize candidate cells for mitosis detection. Although these pixel-wise textural and statistical features have been proved to be effective imaging attributes for mitosis detection, the discrimination power of these features can be degraded by artifacts present in the image due to slide preparation and acquisition. Further, Wang *et al.* [19], Ciresan *et al.* [20], Malon and Cosatto [21], and Albarqouni *et al.* [22] used convolutional neural network (CNN) to detect mitosis in breast histology images. Despite the issue of high computational complexity, these CNN-based approaches, proved to be effective and to have a high accuracy in detecting mitotic cells [12].

Wavelets have emerged as an effective tool for image processing as they provide a natural partition of the image spectrum into multiscale and oriented subbands [23]. The discrete wavelet transform (DWT) has been used extensively within various medical image processing applications, such as enhancement, fusion, compression, segmentation, and classification [24, 25].

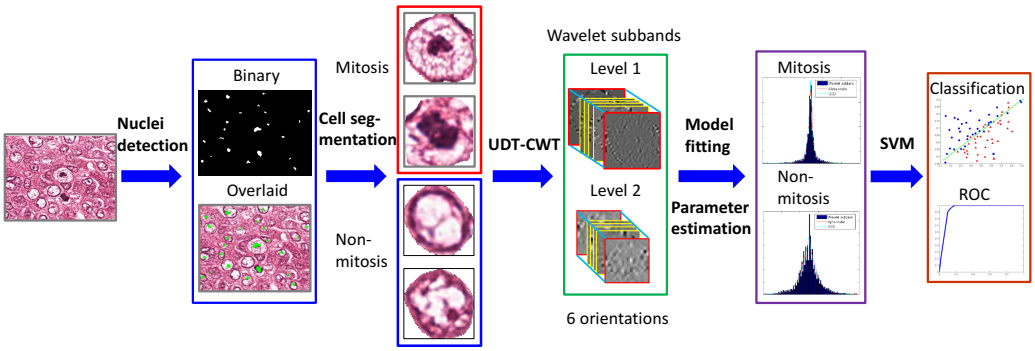


Figure 1: The flowchart illustrating that mitotic and non-mitotic cell figures segmented from original histopathological images are decomposed into multi-resolution representations via UDT-CWT. The model parameters estimated on wavelet coefficient distributions are extracted to well characterize tissue cells and evaluated by a SVM classifier.

Weyn *et al.* employed wavelets for multi-scale image analysis on the microscopic images to extract parameters for the description of chromatin texture in grading invasive breast cancers, and showed that the wavelet-based features achieved higher classification scores and recognition scores than the traditional texture features, such as densitometric and co-occurrence features [26]. Wouwer *et al.* also computed the DWT-based energy parameters for the automated identification of neoplastic nuclei in digitalized microscopic images and yielded good classification results in grading of invasive breast cancer. Tripathi *et al.* investigated the ability of multi-level Daubechies wavelets, a commonly used form of DWT, to extract texture features for detecting mitosis in breast histopathological images, and showed a remarkable increase in the sensitivity measure over previous studies [27]. Lopez and Agaian combined wavelet and fractal features for Gleason grading of prostate cancer in histopathology images [28]. Lessmann *et al.* integrated DWT-based texture features and color attributes to separate four classes of meningiomas in histopathology images. Further, Zhang *et al.* expanded 2D-DWT to 3D-DWT to handle volumetric magnetic resonance images to distinguish patients with Alzheimer's disease from mild cognitive impairment and normal controls, and showed that 3D-DWT yielded good classification results [29]. The prior work demonstrated that the wavelet-based computerized approaches offered the potential for automating classification of histological images and supporting cancer diagnosis.

Previous studies revealed that the DWT suffers from shift variance and

poor directional selectivity because its separable filters cannot distinguish between edge features on opposing diagonals [24, 25, 30]. Variations on the DWT have been developed to produce a shift invariant form [31, 32]. The dual tree complex wavelet transform (DT-CWT) [33] is a relatively recent enhancement to the DWT. The DT-CWT offered a compact representation while providing the desirable properties of approximate shift invariance, improved directional selectivity (6 directional subbands per scale), and complex valued coefficients, which are useful for magnitude/phase analysis within the transform domain. The DT-CWT has been widely applied to image modeling [34], denoising [35], enhancement [36], fusion [37, 38], and segmentation [39]. Additionally, Wang *et al.* developed an automatic classification system in detecting abnormal brains from healthy brains on MRI via DT-CWT [40]. Niwas *et al.* computed the DT-CWT based features as nucleus descriptors to automatically classify benign and malignant breast cancers in cytological Images [41]. Wan *et al.* employed DT-CWT statistical features to distinguish mitosis and non-mitosis in breast cancer histopathology, and achieved superior classification performance in terms of accuracy [42]. The undecimated dual-tree complex wavelet transform (UDT-CWT) [25, 43] is an improved version of DT-CWT combining the benefits of the undecimated DWT (exact translational invariance and a one-to-one relationship between all co-located coefficients at all scales) and the DT-CWT (improved directional selectivity and complex subbands). The DT-CWT and UDT-CWT have showed promise in various applications in medical image processing, which motivated our work in detecting mitosis and non-mitosis using wavelet-based method in histopathological images.

Recently, Khan *et al.* [44] presented a statistical approach using a Gamma-Gaussian mixture model for mitosis detection, in which the pixel intensities were modeled as random variables via Gamma and Gaussian distributions. This method yielded 51% F-score during the ICPR 2012 Context [6]. Compared to the image domain, statistical modeling is much easier to perform in the transform space, such as the wavelet domain, because of its energy compaction property. Wan *et al.* [38] have graphically showed that histogram of wavelet coefficients has heavy tails and sharply peaked modes at zero (non-Gaussian heavy tailed behavior), which deviate significantly from the Gaussian distribution. In addition, Loza *et al.* [37] showed that wavelet subband coefficients can be better modeled by the generalized Gaussian distribution (GGD) or symmetric alpha-stable (SαS) distributions compared to the Gaussian model. Although statistical modeling (GGD and SαS models)

in wavelet domain has been widely adopted in the computer vision and recently in the medical image processing field [45, 46, 47], this method has not been fully explored in the computerized histology analysis.

In this paper, we present a wavelet based approach to automatically identify mitosis in breast histopathological images via a non-Gaussian modeling technique. This work is an extension of our previous work published in [42]. The segmented individual tissue cells are decomposed into multi-resolution representations via an undecimated dual-tree complex wavelet transform [25]. Different from the previous work [42], in which 5 statistical features (i.e., mean, median, variance, energy, and entropy) were extracted from DT-CWT subbands, the presented work utilized two non-Gaussian models, i.e., GGD and SαS distributions, to accurately model the heavy-tailed behavior of subband coefficient distributions [38, 39]. Model parameters estimated from wavelet subbands are used to represent the cell characteristics. A support vector machine (SVM) classifier is utilized in distinguishing mitosis from non-mitosis on a cell basis. Our approach was quantitatively tested on two independent data cohorts obtained from two institutions, rather than only one dataset used by the previous work [42]. The flowchart of the presented method is illustrated in Figure 1.

Mitosis detection is very challenging since the biological variability of mitotic cells and imaging artifacts make the detection extremely difficult. Our main contributions are: (1) Non-Gaussian distributions (GGD and SαS) are used to characterize texture attributes for mitosis and non-mitosis by modeling the marginal densities of wavelet subband coefficients. This is justified by recent psychological research on human texture perception suggesting that two homogeneous textures are often difficult to discriminate if they produce similar marginal distributions of responses from a bank of filters [48]. The texture difference between the mitosis and non-mitosis can be accurately captured through adaptively estimating the model parameters of wavelet coefficient distributions, thus leading to a robust detection approach which is less sensitive to the quality of mitosis segmentation. (2) The use of bivariate non-Gaussian model to capture both marginal densities of wavelet subbands and the strong interdependencies between wavelet coefficients at different scales, leads to a more accurate texture model compared to an univariate model. For instance, a coefficient with a high value appearing in the current scale which does not have a correspondingly high coefficient appearing in the lower adjacent scale is probably caused by noise. By taking into account the joint statistics of wavelet coefficients across subbands, we are able to extract

not only the salient features of mitosis and non-mitosis on multi-resolution representations, but also reduce the possibility of missing information which often happens in processing small regions of interest (ROI).

The rest of this paper is organized as follows. Section 2 provides the necessary preliminaries on generalized Gaussian and alpha-stable processes. Section 3 describes the details of the method. The experimental design and results are presented in Section 4 and Section 5, respectively. Concluding remarks are given in Section 6.

2. Non-Gaussian models for image wavelet coefficients

In this section, two non-Gaussian models, i.e., GGD and SαS, underlying our wavelet-based mitosis detection method are presented to characterize the heavy-tailed behavior of the wavelet subband marginal distribution. SαS processes include the Cauchy and Gaussian distributions as limiting cases, whereas the GGDs cover both Gaussian and Laplace.

2.1. Generalized Gaussian model

2.1.1. Univariate model

A univariate case of GGD probability density functions (PDF) can be written as [37]:

$$p(x; s, p) = \frac{p}{2s\Gamma(1/p)} e^{-(|x-\mu|/s)^p} \tag{1}$$

where $\Gamma(t) = \int_0^\infty e^{-u} u^{t-1} du$ is the gamma function, μ represents location, s and p are the scale and shape parameters, respectively. Since our developments are in the framework of wavelet analysis, only signals with zero mean ($\mu = 0$) are considered throughout this work. s models the width of the PDF peak, while p is inversely proportional to the decreasing rate of the peak. The smaller values of p ($p < 2$) reflect the heavier tails of the corresponding GG density function compared to the Gaussian distribution. Examples of GGD family of distributions illustrating the different values of the shape parameters p are shown in Figure 2(a). The GGD family includes the Gaussian model when $p = 2$ and the Laplace model when $p = 1$ as special cases. Their marginal PDFs with mean $\mu = 0$ can be written as follows:

$$p(x; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(|x|/(\sqrt{2}\sigma))^2}; p(x; \sigma) = \frac{1}{\sqrt{2}\sigma} e^{(-\sqrt{2}|x|/\sigma)} \tag{2}$$

where $\sigma = s\sqrt{\Gamma(3/p)/\Gamma(1/p)}$ is the standard deviation of the distribution.

2.1.2. Bivariate model

Univariate distributions cannot model the statistical dependencies between wavelet coefficients. Previous work revealed that there are strong dependencies between the coefficients, their parent (adjacent coarser scale locations), and their sibling coefficients across different orientations [39, 37]. In this study, we consider the problem of modeling the statistical dependencies between coefficients and corresponding parent coefficients. Let x_{l+1} represent the parent coefficient of x_l at the same position. We assumed that the variances of wavelet coefficients at child and parent levels are equal. The bivariate dependent Laplace model is given by [49]:

$$p(x_l, x_{l+1}; \sigma) = \frac{3}{2\pi\sigma^2} e^{(-\sqrt{3((\frac{x_l}{\sigma})^2 + (\frac{x_{l+1}}{\sigma})^2)})} \tag{3}$$

The reason for restricting our modelling to the Laplace case is the availability of easily derived parameter estimators.

2.2. Symmetric alpha-stable model

2.2.1. Univariate model

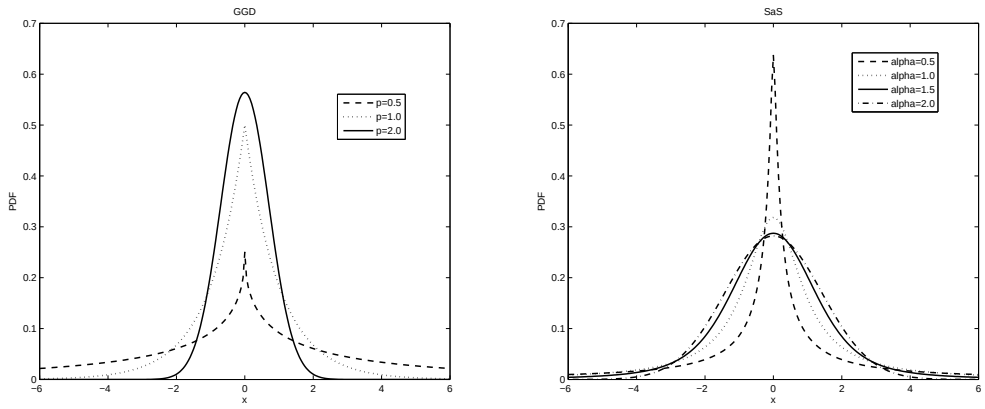
Due to lack of a compact analytical expression for their probability density function, the univariate SαS distribution can be defined by their characteristic functions (i.e., the Fourier transform of the PDF) as follows [50]:

$$\varphi(\omega) = e^{(j\delta\omega - \gamma|\omega|^\alpha)} \tag{4}$$

where ω denotes frequency, j is the imaginary unit, $\delta(-\infty < \delta < \infty)$ is the location parameter ($\delta = 0$ in our work), $\alpha(0 < \alpha \leq 2)$ is the characteristic exponent that determines the shape of the distribution, and $\gamma(\gamma > 0)$ is the dispersion of the distribution that describes the spread of the distribution around its location parameter δ.

Figure 2(b) illustrate the SαS density function taking various values of the characteristic exponent α. Note that the smaller α is, the heavier the tails of the SαS density function. This implies that random variables following distributions with small characteristic exponents are highly impulsive. Gaussian and Cauchy processes are special cases of stable processes with $\alpha = 2$ and $\alpha = 1$, respectively. The probability distribution for the Cauchy density with location parameter $\delta = 0$ is given by:

$$p(x; \gamma) = \frac{\gamma}{\pi(x^2 + \gamma^2)} \tag{5}$$



(a) (b)

Figure 2: (a) Graphs of generalized Gaussian densities corresponding to the values $p = 0.5$, $p = 1.0$ (Laplace), and $p = 2.0$ (Gaussian), for $s = 1.0$ and $\mu = 0$. (b) Graphs of symmetric alpha-stable densities corresponding to the values $\alpha = 0.5$, $\alpha = 1.0$ (Cauchy), $\alpha = 1.5$, and $\alpha = 2.0$ (Gaussian), for $\gamma = 1.0$.

Although the SαS density function behaves approximately like a Gaussian density function near the origin, the stable densities have algebraic tails (heavy tails) while the Gaussian density function has exponential tails which decay at a faster rate [37, 38]. One consequence of heavy tails is that only moments of order less than α exit for the non-Gaussian alpha-stable family members, i.e., $E|x|^m = C(m,\alpha)\gamma^{\frac{m}{\alpha}}$ for $-1 < m < \alpha$, where m is the moment order and $C(m,\alpha) = (2^{m+1}\Gamma(\frac{m+1}{2})\Gamma(-\frac{m}{\alpha}))/(\alpha\sqrt{\pi}\Gamma(-\frac{m}{2}))$.

2.2.2. Bivariate model

Unlike univariate stable distributions, bivariate alpha-stable distributions form a nonparametric set being thus much more difficult to describe. An exception is the family of multidimensional isotropic stable distributions whose characteristic function has the form:

$$\varphi(\omega_1,\omega_2) = e^{(j(\delta_1\omega_1+\delta_2\omega_2)-\gamma|\boldsymbol{\omega}|^\alpha)} \tag{6}$$

where $\boldsymbol{\omega} = (\omega_1,\omega_2)$, and $|\boldsymbol{\omega}| = \sqrt{\omega_1^2+\omega_2^2}$. The distribution is isotropic with respect to the location point (δ_1,δ_2). The two marginal distributions of the isotropic stable distribution are SαS with parameters (δ_1,γ,α) and (δ_2,γ,α).

As before, we will assume that the location $(\delta_1, \delta_2) = (0, 0)$, while the parameters α and γ are used to characterize the distribution of wavelet subband coefficients. As in the case of the univariate SαS density function, when $\alpha = 1$ (Cauchy) and $\alpha = 2$ (Gaussian), no closed form expressions exist for the density function of bivariate alpha-stable random variables. A numerical algorithm for computing bivariate SαS densities has been developed by Nolan [51]. The bivariate Cauchy distribution as a special case of SαS is defined as:

$$p(x_l, x_{l+1}; \gamma) = \frac{\gamma}{2\pi(x_l^2 + x_{l+1}^2 + \gamma^2)^{\frac{3}{2}}} \tag{7}$$

where x_l and x_{l+1} are child and parent wavelet coefficients on the decomposition level l and $l + 1$, respectively.

3. Methods

3.1. Candidate Detection

We utilized a two-step detection method to identify mitosis candidates via a K-means clustering algorithm and a distance regularized level set evolution (DRLSE) based segmentation method [52].

3.1.1. Step 1: Nuclei detection

The breast pathological images were stained with hematoxylin and eosin (H&E) in which the nuclear and cytoplasm regions appear as hues of blue and purple while extracellular material have hues of pink. As we were interested in nuclei which appear as blue-purple regions, the input pathological images in the RGB color space were transformed into a blue-ratio space [53], in which a pixel with a high blue intensity was given a high value, whereas a pixel with a low blue intensity was given a low value. The blue-ratio representation has been proven to be able to highlight nuclei regions [17, 19]. Moreover, in the blue-ratio channel, mitotic figures usually appear to be brighter than non-mitotic ones, thus providing more discriminative information for mitosis detection in histopathological images [22]. By using the blue-ratio mapping, the nuclei of cells could be detected using a simple thresholding approach [18].

A traditional 2-class K-means clustering algorithm was employed to the blue-ratio images to produce a binary image with the high-value pixels having value of 1 as candidate nuclear regions. We repeated the clustering three

times using randomly selected initial cluster centroid positions to avoid local minima and adopted the solution with the lowest within-cluster sums of point-to-centroid distances. The squared Euclidean distance was computed as a distance measure for each centroid cluster. We then performed morphological dilation and erosion operations to divide the connected large regions into individual regions in case that the nuclei may be overlapping, clustered or tightly clumped. We finally applied a thresholding operation to eliminate anomalously small regions, which might be image artifacts due to the fact of non-standardization in histopathological work flow.

3.1.2. Step 2: Cell Segmentation

The identified nuclei candidates from Step 1 were used as initialization points for the DRLSE based segmentation method [52]. In the segmentation model, a general variational level set formulation with a distance regularization term and an external energy term drives the motion of the zero level contour toward desired locations. The distance regularization term is defined with a potential function such that it forces the gradient magnitude of the level set function to one of its minimum points, thereby maintaining a desired shape of the level set function, particularly a signed distance profile near its zero level set. The energy function $\mathcal{E}(\phi)$ is defined by [52]:

$$\mathcal{E}(\phi) = \mu\mathcal{R}_p(\phi) + \lambda\mathcal{L}_g(\phi) + \alpha\mathcal{A}_g(\phi) \tag{8}$$

where ϕ is a level set function. $\mathcal{R}_p(\phi)$ is the level set regularization term, and $\mu > 0$ is a constant. $\lambda > 0$ and α are the coefficients of the edge-based energy functions $\mathcal{L}_g(\phi)$ and $\mathcal{A}_g(\phi)$, which are defined as external energy functions to ensure that the zero level contour of ϕ is located at the object boundaries. This allows for speeding up of the motion of the zero level contour in the level set evolution process. Further, we employed the potential nuclear candidate region rather than a single pixel, which was close to the region to be segmented, as initializations of zero level set function. Thus, a small number of iterations were needed to move the zero level set from the initialized boundary to the desired object boundary. This could greatly reduce the computational cost for segmenting numerous nuclei in histopathological images.

3.2. Wavelet Representation

The undecimated dual-tree complex wavelet transform has been proved to have several advantages over the conventional discrete and complex wavelet

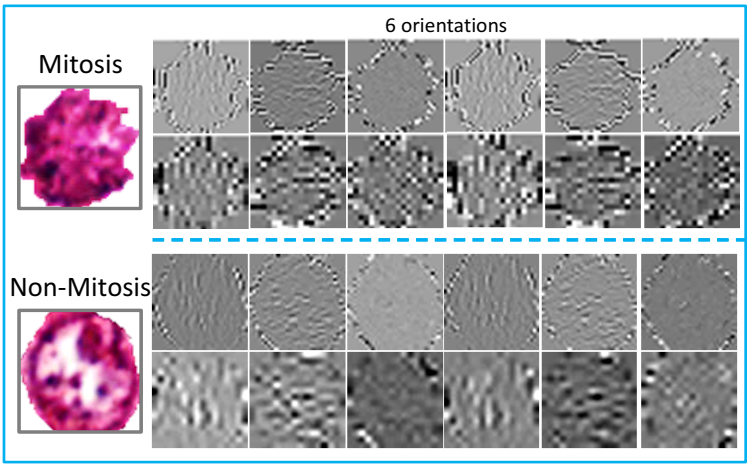


Figure 3: Two examples of UDT-CWT representations with 2 decomposition levels (rows) and 6 orientations (columns) for a mitotic cell and a non-mitotic cell. The wavelet subband images reveal that important texture features appearing in coarse level are more likely present in the adjacent level.

transforms in terms of translational invariance, improved scale-to-scale coefficient correlation and the directional selectivity [25]. The UDT-CWT is an improved form of DT-CWT [33], which is a complex-valued extension to the standard 2D DWT. The DT-CWT calculates the complex transform of a signal using two separate DWT with two decomposition trees to generate the real and imaginary parts of complex wavelet coefficients. This introduces a limited amount of redundancy and allows the transform to provide approximate shift invariance and directionally selective filters, while preserving the properties of perfect reconstruction and computational efficiency [34]. The filterbank implementation of DWT can be interpreted as computing the wavelet coefficients of a discrete set of child wavelets for a given mother wavelet. In the case of DWT, the mother wavelet ψ is defined by:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\psi(\frac{t - k2^j}{2^j}) \tag{9}$$

where j is the scale parameter and k is the shift parameter. For an image with a size of $N_w \times N_h$, the wavelet subband $W_{l,\theta}$ has a size of $N_w/2^{l-1} \times N_h/2^{l-1}$. Each subband $W_{l,\theta}$, $l \in \{1, ..., L\}$, $\theta \in \{\pm\frac{\pi}{12}, \pm\frac{3\pi}{12}, \pm\frac{5\pi}{12}\}$ contains a real and imaginary parts of the complex coefficients. The decomposition level L was chosen depending on the size of the image according to the following formula

$L = (\log_2 \min\{N_w, N_h\}) - 4$ [36].

Figure 3 shows two examples of UDT-CWT representations for a mitotic cell and a non-mitotic cell, respectively. For a better visualization, the parent subband images have been enlarged to the same size as the child subband images. The mitotic and non-mitotic cells exhibit distinctive patterns of frequency response in wavelets. Moreover, the figures revealed that important texture features appearing in coarse level are more likely present in the adjacent level, which is referred to as the inter-scale dependencies (or the persistency property of wavelet transform). This property allows better extraction of salient features having discrimination capability in distinguishing mitotic and non-mitotic cells.

3.3. Model Parameter Estimation and Feature Extraction

3.3.1. GGD

The parameters s and p of univariate GGD model can be estimated via a maximum likelihood (ML) method defined in [48], which showed that the ML estimator was superiorly performed in GGD models. Furthermore, the ML estimator has been proved to be desirable for modeling heavy-tailed distributions among classic statistical methods, which is often the case of wavelet subband coefficients [54]. Thereby, we used the ML estimator for the GGD model. Let $\boldsymbol{x_l} = (x_{1,l}, x_{2,l}, ..., x_{U,l})$ and $\boldsymbol{x_{l+1}} = (x_{1,l+1}, x_{2,l+1}, ..., x_{U,l+1})$ define U wavelet coefficients and their corresponding parent coefficients at decomposition level l and $l+1$, respectively. The likelihood function of $\boldsymbol{x_l}$ is defined as $L(\boldsymbol{x_l}; s, p) = \log \prod_{i=1}^{U} p(x_{i,l}; s, p)$.

The ML estimators are given as:

$$\frac{\partial L(\boldsymbol{x_l}; s, p)}{\partial s} = -\frac{U}{s} + \sum_{i=1}^{U} \frac{p|x_{i,l}|^p s^{-p}}{s} = 0 \tag{10}$$

$$\frac{\partial L(\boldsymbol{x_l}; s, p)}{\partial p} = \frac{U}{p} + \frac{L\Psi(1/p)}{p^2} - \sum_{i=1}^{U} (\frac{x_{i,l}}{s})^p \log(\frac{|x_{i,l}|}{s}) = 0 \tag{11}$$

where the Digamma function $\Psi(u) = \Gamma'(u)/\Gamma(u)$. $\Gamma(u)$ is the gamma function, and u is the location. The above two equations can be numerically solved by the Newton-Raphson iterative procedure [48].

For the bivariate Laplace model, the parameter σ can also be estimated through the ML method published in [37] which is defined as:

$$\sigma = \frac{\sqrt{3}}{2U} \sum_{i=1}^{U} \sqrt{(|x_{i,l}|^2 + |x_{i,l+1}|^2)} \tag{12}$$

3.3.2. SαS

The model parameters α and γ of univariate SαS densities can be obtained through a log-moment method [39]. The log absolute value of wavelet coefficient $\boldsymbol{x_l}$ is give as $\boldsymbol{y_l} = log(|\boldsymbol{x_l}|)$. The mean and variance of $\boldsymbol{y_l}$ are respectively given by [37]:

$$E(\boldsymbol{y_l}) = \frac{\alpha - 1}{\alpha}\Psi(1) + \frac{\log(\gamma)}{\alpha} \tag{13}$$

and:

$$E([\boldsymbol{y_l} - E(\boldsymbol{y_l})]^2) = \frac{\pi^2(\alpha^2 + 2)}{12\alpha^2} \tag{14}$$

The estimation process simply involves solving Equation (14) for α and substituting back in Equation (13) to find the value of the dispersion parameter γ. For the univariate Cauchy density, the parameter γ can be written as $\gamma = e^{(\frac{\alpha}{U}\sum_{i=1}^{U}\log(|x_{i,l}|)+\Psi(1)\frac{1-\alpha}{\alpha})}$.

The corresponding estimation for the isotropic bivariate models can be obtained by replacing $\boldsymbol{y_l} = log(|\boldsymbol{x_l}|)$ with $\boldsymbol{y_{l,l+1}} = log(|\boldsymbol{x_l}| + |\boldsymbol{x_{l+1}}|)$. Note that we assume a constant γ in adjacent levels. Thus the estimation method applied to the univariate alpha-stable distributions can be used in the bivariate models by combining child-parent coefficients into a single data vector.

3.3.3. Feature extraction

The original image was decomposed into multi-resolution representatives via the undecimated dual-tree complex wavelet transform using six orientations $\{\pm\frac{\pi}{12}, \pm\frac{3\pi}{12}, \pm\frac{5\pi}{12}\}$, as shown in Figure 3. The decomposition level was determined based on the size of the image. The larger image intended to have more levels of decomposition. Each subband contained a real part and an imaginary part of the complex coefficients. For each wavelet subband, GGD and SαS were employed to model the heavy-tailed behavior of subband coefficient distributions. The model parameters were used to characterize the wavelet subband distribution attributes. GGD model has two parameters s and p representing the scale and shape of the distribution, respectively, while SαS model is described by two parameters α and γ denoting shape and dispersion of the distribution. For example, a 2-level based UDT-CWT decomposition produces 24 real and imaginary wavelet subbands, thus resulting in 48 model parameters in total for each non-Gaussian density. We found that real coefficient subband and its corresponding imaginary coefficient subband generated similar distributions, thereby in this work the model

parameters estimated from the real coefficient subband were utilized to form the feature set to be used in the subsequent classification.

3.4. SVM Based Classification

The model parameters estimated from wavelet coefficients were used to characterize the potential mitotic cells. Support vector machines was applied to evaluate the discrimination capability of the extracted features for distinguishing mitosis from non-mitosis. Before performing the classification, we adopted a popular feature selection method (i.e., the minimum-redundancy maximum-relevance (mRMR) feature selection method [55]) to select the features by minimizing redundancy and maximizing statistical dependency based on mutual information measure. The theoretical analysis [55] revealed that mRMR was equivalent to Max-Dependency for first-order feature selection with higher efficiency. The wavelet-based features were ranked through scoring of the most relevant features based on mRMR criterion. The optimal subset of features was generated by adding the best features with the highest ranking scores, which is known as the stepwise regression method [56].

In the SVM classification, a Laplacian radial basis function (RBF) kernel was used to map the input data into a higher dimension space where the data were supposed to have a better distribution, and then an optimal separating hyperplane was chosen. The functional form of the Laplacian RBF is given by [57]:

$$\prod(F_i, F_j) = e^{-\epsilon\|F_i - F_j\|_2}, \tag{15}$$

where F_i and F_j are the image feature matrices of nuclei ($i \neq j$). ϵ is a scaling factor which is used to project the data into the high-dimensional SVM space. A commonly used sequential minimal optimization method was employed to find the separating hyperplane.

4. Experimental Design

4.1. Data Preparation

The method was tested on two independent data cohorts. Dataset I is a publicly available dataset for the MITOS contest held in conjunction with the ICPR 2012 conference [6, 58], which is composed of a set of 50 H&E stained breast histopathological high power field (HPF) images (2084×2084 pixels) from 5 patients. They were scanned by a ScanScope XT slide scanner (Aperio Technologies Inc., Vista, USA) at $40\times$ magnification with a spatial resolution

of 0.2456 μm/pixel. Dataset II contains 20 H&E stained slides from 20 invasive breast carcinoma women patients, who underwent an excision biopsy between 2010 and 2014 at the No. 91 Central Hospital of PLA, China. All slides were digitalized via a whole slide scanner (Leica Biosystems, Wetzlar, Germany) at 40$\times$ magnification with a 0.25μ m/pixel resolution. One expert pathologist with more than 10 years of experience marked ROIs containing cancerous tissues on the digital slides and annotated the locations of mitotic figures. The areas for annotation were chosen using the standard criteria used for scoring mitotic figures [58, 59]. The ROI was divided into multiple HPF images (1200×1200 pixels) and only the HPFs ($N = 374$) that contained at least one mitotic cell were included into the dataset. A total of 1155 mitosis (Dataset I: 326; Dataset II: 829) were manually annotated by two pathologists as the ground truth. Also, 3314 non-mitotic cells (Dataset I: 1173; Dataset II: 2141) were delineated by an expert pathologist who was unaware of the ground truth. A student t test was performed to verify that there was no cell size-related bias between these two data cohorts.

We are aware that there are two other publicly available datasets, i.e., MITOS-ATYPIA-14 [59] and AMIDA13 [60] proposed during the mitosis detection contests at the conferences of ICPR 2014 and MICCAI 2013. Since the rules of the two organizers clearly stated that the datasets were limited for the contests solely and can not be used for research purpose, we did not evaluate our method on these two datasets.

4.2. Implementation Settings

In order to reduce the color variations in tissue appearance due to variability in tissue preparation and the use of scanners from different manufacturers, a nonlinear mapping based stain normalization method [61] was utilized on both Dataset I & II. A histological image from Dataset I was randomly selected as a reference (target) image. All the rest images from Dataset I and II were matched to the target image during the stain normalization.

We conducted three independent experiments. Experiment I was to distinguish mitosis and non-mitosis using a combination of Dataset I & II via an iterative 2-fold cross validation scheme. The mitotic and non-mitotic sets are equally partitioned into a training dataset $\mathcal{Z}_{tra}$ and a testing dataset $\mathcal{Z}_{tes}$ containing both mitotic and non-mitotic cells without overlapping between $\mathcal{Z}_{tra}$ and $\mathcal{Z}_{tes}$. To avoid a biased evaluation, we ensured that there were no samples (mitosis and non-mitosis) from the same patient in the training and testing sets simultaneously. Experiment II performed two detection tasks

to identify mitosis using $\{\mathcal{Z}_{tra}$ =Dataset I, $\mathcal{Z}_{tes}$ = Dataset II$\}$ and $\{\mathcal{Z}_{tra}$ = Dataset II, $\mathcal{Z}_{tes}$ = Dataset I$\}$ to independently validate our wavelet-based method on the two different datasets. Experiment III was done using the same training and testing criteria in the MITOS contest at ICPR 2012 [58] in order to compare out method with the state-of-the-art approaches of ICPR 2012 contest. In the contest, the training and testing set consisted of 35 and 15 HPFs containing 226 and 100 mitosis, respectively.

We applied candidate detection (Section 3.1) on all the training set images and considered those candidates as non-mitosis, which was not mitosis. The dataset consisting of candidate potential mitotic cells was unbalanced between the number of mitosis and non-mitosis candidates. To train a balanced classifier, we reduced non-mitosis class by replacing overlapping non-mitotic cells with their clustered center [19], and increased the number of mitotic cells by applying the synthetic minority oversampling technique [62, 63]. In Experiment I, we randomly selected a balanced mix of mitotic and non-mitotic samples to train the SVM classifier using the 2-fold cross-validation strategy. We chose stratified 2-fold cross-validation in this work rather than the higher folds of cross-validation method, to ensure that both our training and testing sets contained sufficient numbers of samples, thus reducing the variance of the classification model as well as the computational time. The disadvantage of 2-fold cross-validation is that the evaluation might be significantly different depending on the dataset division. We utilized an iterative random sampling approach to perform the 2-fold cross-validation to reduce the error rate of the predicative model.

4.3. Performance Measures

4.3.1. Segmentation performance measure

We measured the segmentation performance via both boundary-based and region-based metrics. Mean absolute difference (MAD) was calculated by evaluating the mean difference between each point on the automated cell segmentation $P^c = \{p_i^c | i \in \{1, ..., M\}\}$ and the corresponding closest point p_i^g on the ground truth manual segmentation $P^g = \{p_i^g | i \in \{1, ..., M\}\}$ delineated by the pathologists. The MAD computation is defined as $MAD = \frac{1}{M}\sum_{i=1}^{M}[min_g \|p_i^c - p_i^g\|_2]$, where M is the number of points. An MAD value of 0 reflects perfect segmentation.

The dice similarity coefficient (DSC) was used to compute the spatial overlap accuracy of the automated segmentation to the ground truth, which can be defined as $DSC = \frac{2|S^c \cap S^g|}{|S^c|+|S^g|}$, where S^c and S^g are the areas enclosed by

the automated segmentation P^c and manual segmentation P^g, respectively. $|\cdot|$ is the cardinality of set. An $DSC = 1$ is indicative of perfect segmentation.

4.3.2. Modeling performance measure

The Kullback-Leibler distance (KLD), a statistical measure of the difference between the data and a model, was used to quantify similarity between the discrete distribution of the wavelet coefficients ω_c and a fitting model ω_m. The KLD computation is given as:

$$KLD(\omega_c \| \omega_m) = \sum_i \ln(\frac{\omega_c(i)}{\omega_m(i)})\omega_c(i) \tag{16}$$

The smaller value of KLD indicates a better approximation of the model to the wavelet coefficient distribution.

4.3.3. Classification performance measure

The classification performance was quantitatively measured by the recall or true positive rate (TPR), precision or positive predictive value (PPV), and F-score. These three measures were utilized in the MITOS contest of ICPR 2012 [58], in which true positive (TP) was defined as detected mitosis whose centroid was localized within a range of 5μm (20.4 pixels) of the centroid of a ground truth mitosis. Nuclei that did not meet this criteria were defined as false positive (FP) or false negative (FN). In addition, a receiver operating characteristic (ROC) analysis and the area under the curve (AUC) were utilized to measure the wavelet-based features' ability in distinguishing between mitotic and non-mitotic cells. The SVM classification was repeated for 100 trials via a 2-fold cross-validation scheme and reported the mean and standard deviation values in the results.

4.4. Comparative Strategies

We compared our approach with seven state-of-the-art approaches using the same MITOS dataset, developed by the research teams who participated in the ICPR 2012 [58] or/and MICCAI 2013 [60, 64] competitions on mitosis detection and ranked as the top 7 best performing methods. The first approach (IDSIA) developed a pixel-based deep neural network framework to identify mitosis [20], which won the first place in both ICPR 2012 and MICCAI 2013 contests. The second approach (IPAL) computed co-occurrence, run-length and scale-invariant feature transform features for mitosis and non-mitosis patches [18]. The third approach (SUTECK) used completed local

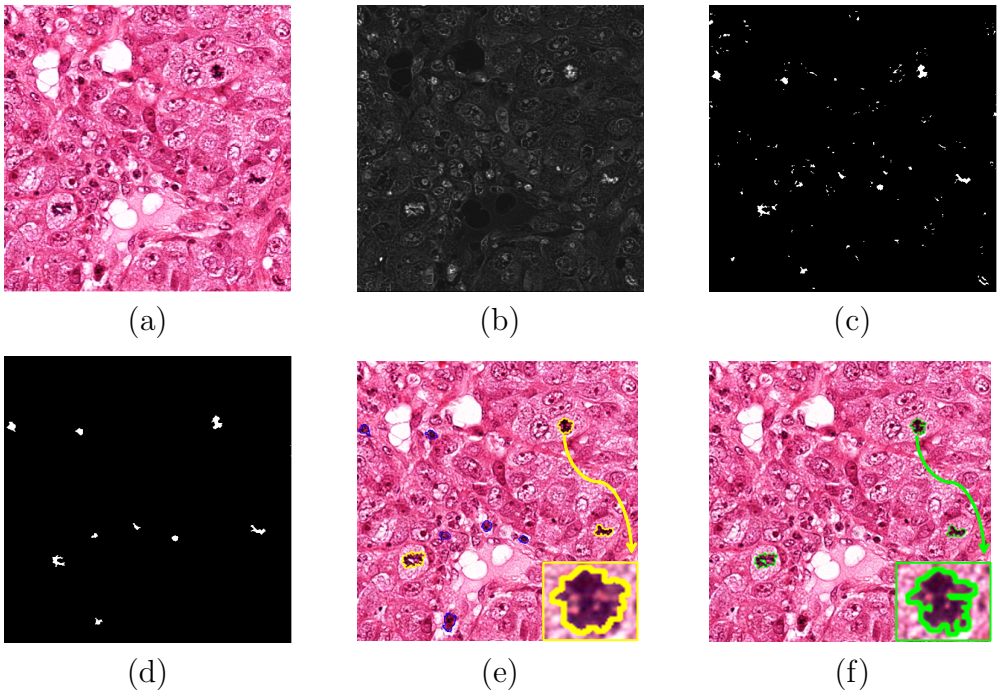


Figure 4: Visual results of potential mitosis candidate detection in a sample image. (a) Original image; (b) Blue-ratio image; (c) Nuclei detection before morphological operators and thresholding; (d) Nuclei detection after morphological operators and thresholding; (e) Results of cell segmentation marked mitotic and non-mitotic cells in yellow and blue color, respectively; (f) Original image with ground truth marked mitotic cells shown in green color. Magnified figures for segmented mitosis cell and ground truth are provided for a better comparison.

binary patterns pixel-wise SVM classification in mitosis detection [65, 14]. The NEC team [21] and CCIPD/MINDLAB team [19] employed the learned CNN-derived features for mitosis detection. The UTRECHT team extracted size, shape, color and texture features of candidate objects for automatically detecting mitotic figures [66]. The WARWICK approach modeled the pixel intensities of mitosis by a Gamma-Gaussian mixture model in conjunction with the SVM classifier [44].

Table 1: The candidate detection performance ($\mu \pm \sigma$) measured by MAD and DSC. The number of the samples (N) is given in the parentheses.

Dataset I	Mitosis ($N = 317$)	Non-mitosis ($N = 1173$)	Total ($N = 1490$)
MAD	8.81 ± 7.17	7.42 ± 11.84	7.72 ± 10.26
DSC	0.71 ± 0.23	0.72 ± 0.26	0.72 ± 0.25
Dataset II	Mitosis ($N = 801$)	Non-mitosis ($N = 2141$)	Total ($N = 2942$)
MAD	11.39 ± 10.68	9.49 ± 14.81	10.01 ± 13.87
DSC	0.66 ± 0.26	0.71 ± 0.23	0.70 ± 0.24

5. Results and Discussion

5.1. Cell Segmentation

The potential mitosis candidates were identified and segmented via our candidate detection method. Figure 4(e) shows the visual results of cell segmentation for both mitotic (in yellow color) and non-mitotic cells (in blue color), suggesting that the 2-step candidate detection method performs well in segmenting mitotic cells compared to the ground truth (Figure 4(f)). The candidate detection phase identified 40970 mitosis candidates (Dataset I: 12267; Dataset II: 28703), containing 1118 mitosis out of 1155 ground truth (Dataset I: 317 out of 326 ground truth; Dataset II: 801 out of 829 ground truth).

The MAD and DSC were used to quantitatively evaluate the segmentation performance shown in Table 1. The MAD and DSC calculations were limited to the detected mitotic ($N = 1118$) and non-mitotic cells ($N = 3314$) with available ground truth. We noted that Dataset II had slightly worse performance of segmentation than Dataset I due to the inherent diversity of tissue appearance. Also we applied stain normalization to Dataset II to be aligned to Dataset I, which might induce artifacts in the images of Dataset II. The quantitative results showed that our candidate detection method yielded a good cell segmentation, which can provide accurate coefficient distributions to ensure a more reliable modeling.

5.2. Non-Gaussian modeling

We computed KLD as a similarity metric to measure the accuracy of model matching between the statistical models (GGD, SαS, Gaussian) and

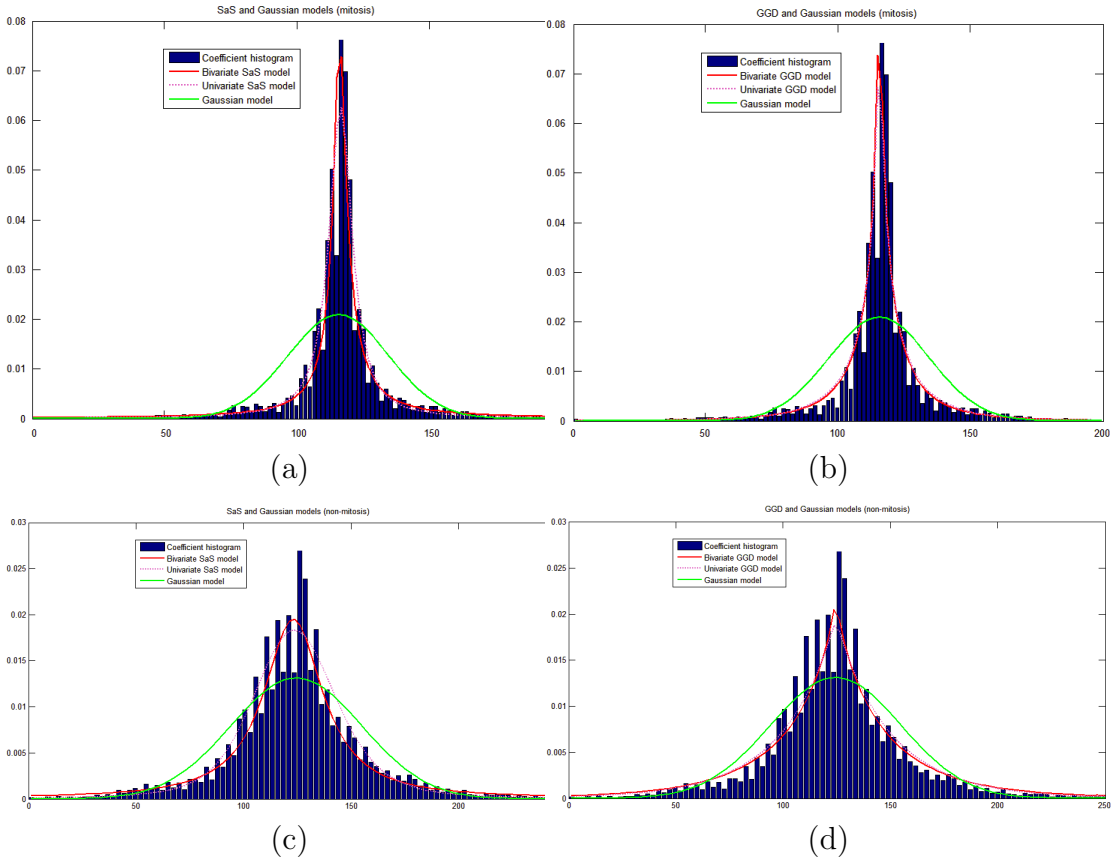


Figure 5: Two examples of model fitting for a mitotic cell (first line) and a non-mitotic cell (second line). (a) SαS (Univariate: $\alpha = 1.04$, $\gamma = 4.98$, $KLD = 0.032$; Bivariate: $\alpha = 0.83$, $\gamma = 4.87$, $KLD = 0.024$); (b) GGD (Univariate: $s = 2.87$, $p = 0.36$, $KLD = 0.037$; Bivariate:$s = 2.24$, $p = 0.52$, $KLD = 0.029$) with univariate and bivariate models in comparison with the estimated Gaussian distribution ($\mu = 1.65.$, $\sigma = 0.19$, $KLD = 0.139$). (c) SαS (Univariate: $\alpha = 1.52$, $\gamma = 5.66$, $KLD = 0.042$; Bivariate: $\alpha = 1.67$, $\gamma = 6.01$, $KLD = 0.036$); (d) GGD (Univariate: $s = 2.03$, $p = 0.96$, $KLD = 0.049$; Bivariate:$s = 1.66$, $p = 0.81$, $KLD = 0.44$) with univariate and bivariate models in comparison with the estimated Gaussian distribution ($\mu = 1.24.$, $\sigma = 0.30$, $KLD = 0.092$).

wavelet coefficient distributions. The KLD values are listed in Table 3. Figure 5 shows two typical examples of a histogram of wavelet coefficients in a particular subband for a mitotic cell and a non-mitotic cell, together with a plot of the fitted estimated SαS model in comparison with the fitted GGD model. The model parameters was obtained via the log-moment method [39]

Table 2: Numbers of mitosis and non-mitosis in the training and testing sets for three experiments.

Experiment	$\mathcal{Z}_{tra}$	Before sampling		After sampling		$\mathcal{Z}_{tes}$	Mitosis	Non-mitosis
		Mitosis	Non-mitosis	Mitosis	Non-mitosis			
Exp. I	I & II	1155	39852	1923	27220	I & II	1923	27220
Exp. II	I	326	11950	539	8266	II	829	-
Exp. II	II	829	27902	1406	18954	I	326	-
Exp. III	I	226	7365	459	5047	I	100	-

I & II denote the Dataset I and Dataset II.

for SαS model and the maximum likelihood method [48] for GGD model. The SαS model provides a better fit than the GGD in terms of modeling the non-Gaussianity inherent in wavelet subbands. In addition, by incorporating the dependencies between parent and child coefficients across scales, the bivariate models can accurately describe the persistency property of wavelet coefficients, which is consistent with the KLD results (Table 3). By examining the figures, we noted that mitosis and non-mitosis have different subband marginal distributions, which can be closely matched by non-Gaussian models.

5.3. Mitosis Classification and Detection

We conducted three experiments on two datasets using GGD and SαS models. Experiment I was done to distinguish mitotic and non-mitotic cells via a 2-fold cross-validation scheme. Experiment II and III were performed to detect mitosis using two independent data cohorts for training and testing. Table 2 shows the numbers of mitosis and non-mitosis in training and testing sets. Over-sampling and down-sampling techniques were used to correct the classification bias due to the unbalanced datasets. The performance metrics are reported in Table 3.

5.3.1. Experiment I

The classification results in Table 3 demonstrated that the bivariate non-Gaussian distributions outperformed the univariate distributions due to the fact that the property of inter-scale dependencies across wavelet subbands could be more accurately captured by the bivariate models than the univariate models. These inter-dependencies of wavelet coefficients allow extraction of discriminative features within the cells that are more likely to appear at

Table 3: The classification ($\mu \pm \sigma$) and detection performance measured by TP, FP, FN, TPR, PPV, F-score, and AUC. The KLD values were computed using different training datasets.

Method	TP	FP	FN	TPR	PPV	F-score	AUC	KLD*
Experiment I: $\{\mathcal{Z}_{tra}, \mathcal{Z}_{tes}\}$ = Dataset I & Dataset II								
Uni-SαS (15)**	–	–	–	0.76 ± 0.14	0.71 ± 0.20	0.73 ± 0.16	0.77 ± 0.17	0.090 ± 0.056
Bi-SαS (11)	–	–	–	0.79 ± 0.10	0.75 ± 0.18	0.77 ± 0.13	0.82 ± 0.15	0.074 ± 0.049
Uni-GGD (14)	–	–	–	0.70 ± 0.15	0.69 ± 0.12	0.70 ± 0.13	0.72 ± 0.05	0.102 ± 0.050
Bi-GGD (12)	–	–	–	0.76 ± 0.13	0.72 ± 0.16	0.74 ± 0.14	0.79 ± 0.08	0.079 ± 0.063
Gaussian (11)	–	–	–	0.58 ± 0.23	0.48 ± 0.27	0.53 ± 0.24	0.59 ± 0.19	0.346 ± 0.178
Experiment II: $\{\mathcal{Z}_{tra}$ = Dataset I, $\mathcal{Z}_{tes}$ = Dataset II$\}$; GT = 829								
Uni-SαS (13)	555	355	274	0.67	0.61	0.64	–	0.102 ± 0.068
Bi-SαS (12)	580	341	249	0.70	0.63	0.66	–	0.082 ± 0.061
Uni-GGD (14)	531	400	298	0.64	0.57	0.60	–	0.113 ± 0.047
Bi-GGD (13)	564	392	265	0.68	0.59	0.63	–	0.087 ± 0.073
Gaussian (10)	415	572	414	0.50	0.42	0.46	–	0.387 ± 0.187
Experiment II: $\{\mathcal{Z}_{tra}$ = Dataset II, $\mathcal{Z}_{tes}$ = Dataset I$\}$; GT = 326								
Uni-SαS (12)	228	140	98	0.70	0.62	0.66	–	0.061 ± 0.043
Bi-SαS (10)	235	121	91	0.72	0.66	0.69	–	0.052 ± 0.037
Uni-GGD (13)	222	148	104	0.68	0.60	0.64	–	0.073 ± 0.055
Bi-GGD (11)	228	123	98	0.70	0.65	0.67	–	0.059 ± 0.042
Gaussian (10)	166	142	160	0.51	0.54	0.52	–	0.243 ± 0.166
Experiment III: $\{\mathcal{Z}_{tra}, \mathcal{Z}_{tes}\}$ = Dataset I; GT = 100								
Uni-SαS (11)	75	35	25	0.75	0.68	0.71	–	–
Bi-SαS (10)	76	31	24	0.76	0.71	0.73	–	–
Uni-GGD (14)	70	36	30	0.70	0.66	0.68	–	–
Bi-GGD (11)	72	32	28	0.72	0.69	0.70	–	–
Gaussian (9)	61	63	39	0.61	0.49	0.54	–	–
IDSIA [20]	70	9	30	0.70	0.88	0.78	–	–
IPAL [17]	74	32	26	0.74	0.70	0.72	–	–
SUTECH [65]	72	31	28	0.72	0.70	0.71	–	–
NEC [21]	59	20	41	0.59	0.75	0.66	–	–
CCIPD/MINDLAB [19]	65	12	35	0.65	0.84	0.73	–	–
UTRECHT [66]	68	65	32	0.68	0.51	0.58	–	–
WARWICK [44]	57	65	43	0.57	0.47	0.51	–	–

* KLD = Kullback-Leibler distance. KLD was computed using training dataset $\mathcal{Z}_{tra}$.
** The optimal number of features used in the classification chosen by the mRMR feature selection method.
Uni- = univariate; Bi- = bivariate; GT = ground truth; TP = truth positive;
FP = false positive; FN = false negative.
TPR = true positive rate; PPV = positive predictive value; AUC = area under the curve.

all different scales. This is particularly useful in characterizing small objects with a large variety of appearances, such as mitotic and non-mitotic cells present in histopathological images, in which salient information might be corrupted by image artifacts. We noted that the SαS distributions achieve better classification performance compared to the univariate and bivariate GGD models, suggesting that non-Gaussian heavy-tailed behavior of wavelet subband marginal distribution can be best modeled by SαS models.

5.3.2. Experiment II

In order to validate the presented method to the choice of whole slide scanners and clinical sites, we performed two detection tasks using $\{\mathcal{Z}_{tra} =$ Dataset I, $\mathcal{Z}_{tes} =$ Dataset II$\}$ and $\{\mathcal{Z}_{tra} =$ Dataset II, $\mathcal{Z}_{tes} =$ Dataset I$\}$, respectively, for automatically identify mitotic cells. The detection results are shown in Table 3. Consistently, bivariate SαS distributions yielded the best performances (F-score= 0.66 and 0.69) among all the models. By comparing

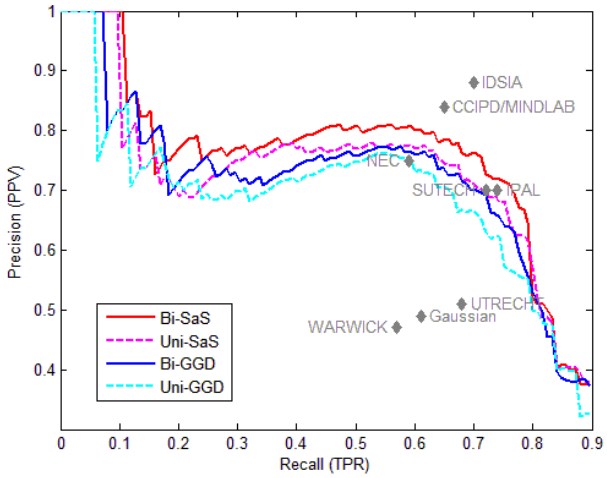


Figure 6: Performance of our approach compared to other methods in the Precision-Recall plane (Experiment III). It can be seen that the bivariate SαS model outperformed GGD and univariate SαS model for distinguishing mitotic from non-mitotic cells.

these two detection tasks, we can see that our method yielded better classification performance in terms of F-score by increasing the training sample size. It seemed that a larger training set did enhanced the SVM classifier in the generalization ability to distinguish mitosis and non-mitosis. We noted that the classifiers were biased toward non-mitosis, which resulted in high numbers of false positive due to the unbalanced training sets. This issue could be improved using a balanced cross-validation scheme as shown in PPV values of Experiment I (Table 3).

5.3.3. Experiment III

Further, we compared our method with seven state-of-the-art methods using Dataset I and the best results reported in their original publications [6]. Our method achieved a higher F-score (0.73) than all other methods except the IDSIA method (0.78) and the CCIPD/MINDLAB method (0.73). Compared with the IDSIA [20], NEC [21] and CCIPD/MINDLAB [19] methods, our method did not require CNN model training which could take a large amount of time ($>$ 10 hours). It can be seen that the TPR values associated with SαS models were higher than the other methods. However, we also noted that our method yielded higher numbers of false positive compared to the IDSIA, NEC, and CCIPD/MINDLAB methods. In addition, our model-based approach greatly outperformed the WARWICK method [44] (F-score=0.51),

which also used statistical modeling and SVM. The same experimental results using different statistical models are plotted in the Precision-Recall plane in Figure 6. The Precision-Recall values of the other methods are also plotted for comparison. Figure 6 shows that the performances of other methods (except IDSIA and CCIPD/MINDLAB) lie in the interior of the areas spanned by the curve of bivariate SαS models.

In summary, we demonstrated an effective technique through a framework of combination of wavelet decomposition and non-Gaussian modeling to identify mitotic figures with fair performance results. Compared to the CNN-based methods, such as IDSIA and CCIPD/MINDLAB methods, our approach was faster, requiring far less computing resources. Further, our wavelet-based method provided more accurate results in detecting mitosis in comparison with statistical modeling based WARWICK method. The disadvantage is that our method obtained high numbers of false positive due to a great amount of potential mitosis candidates identified by the nuclei detection phase, which can be improved via a more accurate candidate detection algorithm by considering the texture features within the neighborhood of nuclei.

Our work did have its limitations, and as such, it is important to acknowledge that this is a preliminary work with need for additional independent validation of our initial findings. Additionally, we did not provide the characteristics of the patients due to lack of patient information for Dataset I. Owing to the limited size of the dataset considered in this study, we did not conduct multiple statistical tests of comparisons on the features. We also did not explicitly quantify the inter-observer variability in segmentation of mitotic and non-mitotic cells between multiple pathologists.

5.4. Parameter Setting

In the nuclei detection, we computed a ratio of region area and the mean area of detected large region candidates (areas are larger than the mean area of all the regions) in the image as a thresholding to remove anomalously small regions resulting from image artifacts. The thresholding values were automatically selected from 0.2 to 0.5 and found that 0.35 gave the best performance in removing the small regions, meanwhile preserving the nuclear regions. Additionally, there were three parameters $\{\mu, \lambda, \alpha\}$ defined in the DRLSE model (Equation (8)) for segmenting nuclei. The model was not sensitive to the choice of μ and λ, which were fixed as $\mu = 0.02$, and $\lambda = 5.0$ reported as default values in [52]. The parameter α controlled additional

external force to drive the motion of contour towards the true object boundary. A large value of α would cause boundary leakage, which was undesirable for small objects with week boundaries, such as mitotic and non-mitotic figures. Therefore, α was empirically tuned to yield the best segmentation results by choosing relatively small values to avoid boundary leakage. In all the experiments, we used $\alpha = 1.5$ for Dataset I and $\alpha = 2.0$ for Dataset II. In the SVM classification, we used the Laplacian RBF kernel instead of the standard Gaussian RBF kernels and linear kernel because its nonlinear projection creates additional separation between the data points in the high-dimensional SVM space, and hence simplifies the classification task [57, 67]. The classification parameters were tuned via a 10-fold cross-validation over the training dataset to yield the best classification results.

5.5. Computational Time

In order to increase the computational efficiency and reduce the processing time, we employed an image level parallelization with a bag-of-tasks execution model on a multi-core CPU equipped machine. Each CPU core was assigned an image and the tasks of nuclei detection, cell segmentation, and non-Gaussian modeling in that image. Multiple images were processed concurrently on multiple CPU cores. This parallelization strategy has been utilized to conduct cancer grading [9] and large-scale image retrieval in histopathology [68].

The execution time with respect to candidate detection, wavelet decomposition, model parameter estimation, and classification was measured using the Matlab R2013a platform on a 2.67 GHz Intel Core 4 processor with 10GB of RAM. Since the mitosis candidates used in the experiments have various sizes containing different numbers of pixels, the computational times reported here are the average values across the entire dataset. The running time for segmenting a single mitosis candidate is 0.16 second(s). The UDT-CWT decomposition and model fitting take 0.14s and 0.33s per cell, respectively. For a classification task in Experiment I using all the model parameters, the average computational time is 5.54s for one cross-validation. The running time (including segmentation and feature extraction) for Experiment II (two detection tasks) and Experiment III took about 3.5 hours, 8 hours and 2 hour, respectively, for the training stage without an optimized GPU implementation. At the detection stage, three tasks took about 1.2 mins to process each H&E image.

6. Concluding Remarks

We presented a wavelet based non-Gaussian modeling approach for automated detection of mitosis in breast cancer histopathological images. Since mitotic cells exhibit highly various appearances and textures in the images, there is no simple way to characterize mitosis based on shapes and pixel intensities. In this study, bivariate non-Gaussian models of GGD and SαS were employed to characterize both non-Gaussian heavy-tailed behavior of wavelet coefficient marginal distribution and strong inter-scale dependencies across wavelet subbands. The properties allow to capture the discriminative features that appear at both coarse scale and previous adjacent scale in the wavelet domain. The model parameters estimated from wavelet coefficient distributions were used as cell attributes to train a SVM classifier. The wavelet based non-Gaussian modeling method was evaluated on both the benchmark MITOS dataset and an independent dataset from two clinical sites. Compared to the univariate models and Gaussian model, the bivariate SαS model achieved superior classification performance with the AUC values of 0.82 in distinguishing mitosis from non-mitosis. The automated mitosis detection method might assist pathologists in enhancing the operational efficiency and productivity as well as improving diagnostic confidence during clinical practice. Indeed, the mitosis detection of critical importance in breast cancer grading is a challenging task. In future work, we aim at validating the approach on a larger dataset and extending to different types of cancers. Furthermore, we also plan to investigate other statistical models to improve the accuracy of mitosis detection.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China under award Nos. 61305047 and 61401012.

References

[1] J.Ma, A.Jemal, Breast cancer statistics, breast cancer metastasis and drug resistance, Springer New York (2013) 1–18.

[2] Y. Zhang, S. Wang, G. Liu, J. Yang, Computer-aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional fourier transform, Advances in Mechanical Engineering 8 (2) (2016) 1–11.

[3] Y. Zhang, X. Wu, S. Lu, H. Wang, P. Phillips, S. Wang, Smart detection on abnormal breasts in digital mammography based on contrast-limited adaptive histogram equalization and chaotic adaptive real-coded biogeography-based optimization, Simulation 92 (9) (2016) 873–885.

[4] C.W.Elston, I.O.Ellis, Pathological prognostic factors in breast cancer I, the value of histological grade in breast cancer: Experience from a large study with long-term follow-up, Histopathology 19 (1991) 403–410.

[5] J. Elmore, G. Longton, P. Carney, B. Geller, T. Onega, A. Tosteson, H. Nelson, M. Pepe, K. Allison, S. Schnitt, F. O'Malley, D. Weaver, Diagnostic concordance among pathologists interpreting breast biopsy specimens, JAMA 313 (11) (2015) 1122–1132.

[6] L. Roux, D. Racoceanu, N. Lomenie, M. Kulikova, H. Irshad, J. Klossa, F. Capron, C. Genestie, G. L. Naour, M. Gurcan, Mitosis detection in breast cancer histological images An ICPR 2012 contest, J. Pathol. Inform. 2 (8), doi:10.4103/2153-3539.

[7] M. Veta, J. pluim, P. van Diest, M. Viergever, Breast cancer histopathology image analysis: A review, IEEE Trans. Biomed. Eng. 61 (5) (2014) 1400–1411.

[8] A. Madabhushi, S. Agner, A. Basavanhally, S. Doyle, G. Lee, Computer-aided prognosis: predicting patient and disease outcome via quantittive fusion of multi-scale, multi-modal data, Comput. Med. Imag. Graph. 35 (7-8) (2011) 506–514.

[9] T. Wan, J. Cao, J. Chen, Z. Qin, Automated grading of breast cancer histopathology using cascaded ensemble with combinatin of multi-level image features, Neurocomputing to appear.

[10] J. Xu, X. Lou, G. Wang, H. Gilmore, A. Madabhushi, A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images, Neurocomputing 191 (2016) 214–223.

[11] X. Zhang, F. Xing, H. Su, L. Yang, S. Zhang, High-throughput histopathological image analysis via robust cell segmentation and hashing, Med. Image Anal. 26 (1) (2015) 306–315.

[12] H. Irshad, A. Veillard, L. Roux, D. Racoceanu, Methods for nuclei detection, segmentation, and classification in digital histopathology: A reveiw - current status and future potential, IEEE Rev. Biomed. Eng. 7 (2014) 97–114.

[13] C. Sommer, L. Fiaschi, F. Hamprecht, D. Gerlich, Learning-based mitotic cell detction in histopathological images, in: Proc. ICPR, 2012, pp. 2306–2309.

[14] A. Tashk, M. Helfroush, H. Danyali, M. Akbarzadeh, A novel CAD system for mitosis detection using histopathology slide images, J. Med. Signals Sens. 4 (2) (2014) 139–149.

[15] R. Nateghi, H. Danyali, M. Helfroush, A. Tashk, Intelligent CAD system for automatic detection of mitotic cells from breast cancer histology slide images based on teaching-learning-based optimization, Computational Biology Journal 2014, doi:10.1155/2014/970898.

[16] F. B. Tek, Mitosis detection using generic features and an ensemble of cascade adaboosts, J. Pathol. Inform. 4 (1), doi:10.4103/2153-3539.112697.

[17] H.Irshad, Automated mitosis detection in histopathology using morphological and multi-channel statistics features, J. Pathol. Inform. 4, doi:10.4103/2153-3539.112695.

[18] H. Irshad, S. Jalali, L. Roux, D. Racoceanu, L. Hwee, G. L. Naour, F. Capron, Automated mitosis detection using texture, SIFT features and HMAX biologically inspired approach, J. Pathol. Inform. 4 (12), doi:10.4103/2153-3539.109870.

[19] H. Wang, A. Cruz-Roa, A. Basavanhally, H. Gilmore, N. Shih, M. Feldman, J. Tomaszewski, F. Gonzalez, A. Madabhushi, Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features, J. Med. Imaging 1 (3) (2014) 034003.

[20] D. Ciresan, A. Giusti, L. Gambardella, J. Schmidhuber, Mitosis detection in breast cancer histology images with deep neural networks, in: Proc. MICCAI, 2013, pp. 441–418.

[21] C.D.Malon, E.Cosatto, Classification of mitotic figures with convolutional neural networks and seeded blob features, J. Pathol. Inform. 4 (9), doi:10.4103/2153-3539.112694.

[22] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, N. Navab, AggNet: Deep learning from crowds for mitosis detection in breast cancer histology images, IEEE Trans. Med. Imaging. 35 (5) (2016) 1313–1321.

[23] S. Mallat (Ed.), A Wavelet Tour of Signal Processing, 3rd Edition,The Sparse Way, Academic Press, Cambridge, Massachusetts, 2008.

[24] A. N. Akansu, W. A. Serdijn, I. W. Selesnick, Emerging applications of wavelets: A review, Physical Communication 3 (1) (2010) 1–18.

[25] P. Hill, N. Anantrasirichai, A. Achim, M. Al-Mualla, D. Bull, Undecimated dual-tree complex wavelet transforms, Signal Processing: Image Communication 35 (2015) 61–70.

[26] B. Weyn, G. van de Wouwer, A. van Daele, P. Scheunders, D. van Dyck, E. van Marck, W. Jocob, Automated breast tumor diagnosis and grading based on wavelet chromatin texture description, Cytometry 33 (1998) 32–40.

[27] A. Tripathi, A. Mathur, M. Daga, M. Kuse, O. Au, 2-SiMDoM: A 2-sieve model for detection of mitosis in multispectral breast cancer imagery, in: Proc. ICIP, 2013, pp. 611–615.

[28] C. M. Lopez, S. Agaian, A new set of wavelet- and fractals-based features for Gleason grading of prostate cancer histopathology images, in: Proc. SPIE, 2013, p. 865516.

[29] Y. Zhang, S. Wang, P. Phillips, Z. Dong, G. Ji, J. Yang, Detection of Alzheimer's disease and mild cognitive impairment based on structural volumetric MR images using 3D-DWT and WTA-KSVM trained by PSOTVAC, Biomedical Signal Processing and Control 21 (2015) 58–73.

[30] Y. Zhang, Z. Dong, A. Liu, S. Wang, G. Ji, Z. Zhang, J. Yang, Magnetic resonance brain image classification via stationary wavelet transform

and generalized eigenvalue proximal support vector machine, Journal of Medical Imaging and Health Informatics 5 (7) (2015) 1395–1403.

[31] A. Akansu, W. Serdijn, I. Selesnick, Emerging applications of wavelets: A review, Physical Communication 3 (1) (2010) 1–18.

[32] S. Wang, X. Yang, Y. Zhang, P. Philips, J. Yang, T. Yuan, Identification of green, oolong and black teas in china via wavelet packet entropy and fuzzy support vector machine, Entropy 17 (10) (2015) 6663–6682.

[33] I. W. Selesnick, R. G. Baraniuk, N. G. Kingsbury, The dual-tree complex wavelet transform, IEEE Signal Process. Mag. 22 (2005) 123–151.

[34] M. A. Miller, N. G. Kingsbury, Image modeling using interscale phase properties of complex wavelet coefficients, IEEE Trans. Image Process. 17 (6) (2008) 1491–1499.

[35] M. A. Miller, N. G. Kingsbury, Image denoising using derotated complex wavelet coefficients, IEEE Trans. Image Process. 17 (9) (2008) 1500–1511.

[36] A. Loza, D. Bull, A. Achim, Automatic contrast enhancement of low-light images based on local statistics of wavelet coefficients, Digital Signal Processing 23 (6) (2013) 1856–1866.

[37] A. Loza, D. Bull, N. Canagarajah, A. Achim, Non-Gaussian model-based fusion of noisy images in the wavelet domain, Comput. Vis. Image Underst. 114 (1) (2010) 54–65.

[38] T. Wan, N. Canagarajah, A. Achim, Segmentation-driven image fusion based on alpha-stable modeling of wavelet coefficients, IEEE Trans. Multimedia 11 (4) (2009) 624–633.

[39] T. Wan, N. Canagarajah, A. Achim, Segmentation of noisy colour images using Cauchy distribution in the complex wavelet domain, IET Image Process. 5 (2) (2011) 159–170.

[40] S. Wang, S. Lu, Z. Dong, J. Yang, M. Yang, Y. Zhang, Dual-tree complex wavelet transform and twin support vector machine for pathological brain detection, Appl. Sci. 6 (6) (2016) 169–187.

[41] S. I. Niwas, P. Palanisamy, K. Sujathan, Complex wavelet as nucleus descriptors for automated cancer cyology classifier system using ANN, in: Proc. IEEE Conf. on Computational Intelligence and Computing Research, 2010, p. 11822740.

[42] T. Wan, X. Liu, J. Chen, Z. Qin, Wavelet-based statistical features for distinguishing mitotic and non-mitotic cells in breast cancer histopathology, in: Proc. ICIP, 2014, pp. 2290–2294.

[43] N. Anantrasirichai, J. Burn, D. R. Bull, Robust texture features for blurred images using undecimated dual-tree compolex wavelets, in: Proc. ICIP, 2014, pp. 1–4.

[44] A. Khan, H. Eldaly, N. Rajpoot, A gamma-gaussian mixture model for detection of mitotic cells in breast cancer histopathology images, J. Pathol. Inform. 4 (1), doi:10.4103/2153-3539.112696.

[45] A. Achim, A. Bezerianos, P. Tsakalides, Novel bayesian multiscale method for speckle removal in medical ultrasound images, IEEE Trans. Image Process. 20 (8) (2001) 772–783.

[46] A. Achim, B. Buxton, G. Tzaqkarakis, P. Tsakalides, Compressive sensing for ultrasound RF echoes using a-stable distributions, in: Proc. IEEE Eng. Med. Biol. Soc., 2010, pp. 1–4.

[47] T. Wan, Z. Qin, Statistical multisensor image segmentation in complex wavelet domains, in: Proc. ICIAP, 2011, pp. 60–68.

[48] M. N. Do, M. Vetterli, Wavelet-based texture retrieval using generalised Gaussian density and Kullback-Leibler distance, IEEE Trans. Imag. Proc. 11 (2) (2002) 146–158.

[49] L. Sendur, I.W.Selesnick, Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency, IEEE Trans. Sign. Proc. 50 (11) (2002) 2744–2756.

[50] C. L. Nikias, M. Shao (Eds.), Signal Processing with Alpha-Stable Distributions and Applications, Wiley, New York, 1995.

[51] J.P.Nolan (Ed.), Stable Distributions: Models for Heavy Tailed Data, Birkhauser, Boston, 2005.

[52] C. Li, C. Xu, C. Gui, M. Fox, Distance regularized level set evolution and its application to image segmentation, IEEE Trans. Image Process. 19 (2010) 3243–3254.

[53] H. Chang, L. Loss, B. Parvin, Nuclear segmentation in H and E sections via multi-reference graph-cut (MRGC), in: Proc. ISBI, 2012, pp. 611–615.

[54] M. Varanasi, B. Aazhang, Parametric generalized Gaussian density estimation, J. Acoust. Soc. Am. 86 (4) (1989) 1404–1415.

[55] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.

[56] N. R. Draper, H. Smith (Eds.), Applied Regression Analysis, John Wiley & Sons, Inc., New York, 1998.

[57] S. Agner, S. Soman, E. Libfeld, M. McDonald, K. Thomas, S. Englander, M. Rosen, D. Chin, J. Nosher, A. Madabhushi, Textural kinetics: A novel dynamic contrast-enhanced (DCE)-MRI feature for breast lesion classification, J. Digit. Imaging 24 (3) (2011) 446–463.

[58] The International Conference for Pattern Recognition (ICPR), The mitosis Detection in Breast Cancer Histological Images (MITOS) database (2012).
URL `http://ipal.cnrs.fr/ICPR2012/.`

[59] The International Conference for Pattern Recognition (ICPR), Detection of mitosis and evaluation of nuclear atypia score in Breast Cancer Histological Images (MITOS & ATYPIA) dataset (2014).
URL `http://grand-challenge.org/site/mitos-atypia-14/home/.`

[60] The International Conference on Medical Image Computiing and Computer Assisted Intervention (MICCAI), Assessment of mitosis detection algorithms (AMIDA) dataset (2013).
URL `http://amida13.isi.uu.nl/.`

[61] A. Khan, N. Rajpoot, D. Treanor, D. Magee, A nonlinear mapping approach to stain normalization in digital histopathology images using

image-specific color decomvolution, IEEE Trans. Biomed. Eng. 61 (6) (2014) 1729–1738.

[62] N. Chawla, K. Bowyer, L. Hall, W. Kegelmeyer, SMOTE:synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (1) (2002) 321–357.

[63] A. Pozzolo, O. Caelen, S. Waterschoot, G. Bontempi, Racing for unbalanced methods selection, in: Proc. IDEAL, 2013, pp. 24–31.

[64] M. Veta, P. van Diest, S. Willems, H. Wang, A. Madabhushi, A. Cruz-Roa, F. Gonzalez, A. Larsen, J. Vestergaard, A. Dahl, D. Ciresan, J. Schmidhuber, A. Giusti, L. Gambardella, F. Tek, T. Walter, C. Wang, S. Kondo, B. Matuszewski, F. Precioso, V. Snell, J. Kittler, T. de Campos, A. Khan, N. Rajpoot, E. Arkoumani, M. Lacle, M. Viergever, J. Pluim, Assessment of algorithms for mitosis detection in breast cancer histopathology images, Med. Image Anal. 20 (1) (2015) 237–248.

[65] A. Tashk, M. Helfroush, H. Danyali, M. Akbarzadeh, An automatic mitosis detection method for breast cancer histopathology slide images based on objective and pixel-wise textural features classification, in: Proc. IKT, IEEE, 2013, pp. 406–410.

[66] M. Veta, P. van Diest, J. Pluim, Detecting mitotic figures in breast cancer histopathology images, in: Proc. SPIE, 2013, p. 867607.

[67] A. Basavanhally, S. Ganesan, S. Agner, J. Monaco, M. Feldman, J. Tomaszewski, G. Bhanot, A. Madabhushi, Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology, IEEE Trans. Biomed. Eng. 57 (3) (2010) 642–653.

[68] T. Kurc, X. Qi, D. Wang, F. Wang, G. Teodoro, L. Cooper, M. Nalisnik, L. Yang, J. Saltz, D. Foran, Scalable analysis of big pathology image data cohorts using efficient methods and high-performance computing strategies, BMC Bioinformatics 16 (399) (2015) doi: 10.1186/s12859–015–0831–6.