WAVELET-BASED STATISTICAL FEATURES FOR DISTINGUISHING MITOTIC AND NON-MITOTIC CELLS IN BREAST CANCER HISTOPATHOLOGY

Tao Wan[1,2♯], *Xu Liu*[1♯], *Jianhui Chen*[3], *Zengchang Qin*[1*]

[1] Intelligent Computing and Machine Learning Lab
[2] School of Biological Science and Medical Engineering
Beihang University, Beijing, 100191, China
[3] No 91 Central Hospital of PLA, Henan, 454003, China

ABSTRACT

To diagnose breast cancer (BCa), the number of mitotic cells present in tissue sections is an important parameter to examine and grade breast biopsy specimen. The differentiation of mitotic from non-mitotic cells in breast histopathological images is a crucial step for automatical mitosis detection. This work aims at improving the accuracy of mitosis classification by characterizing objects of interest (tissue cells) in wavelet based multi-resolution representations that better capture the statistical features having mitosis discrimination. A dual-tree complex wavelet transform (DT-CWT) is performed to decompose the image patches into multi-scale forms. Five commonly-used statistical features are extracted on each wavelet subband. Since both mitotic and non-mitotic cells appear as small objects with a large variety of shapes in the images, characterization of mitosis is a challenging problem. The inter-scale dependencies of wavelet coefficients allow extraction of important texture features within the cells that are more likely to appear at all different scales. The wavelet-based statistical features were evaluated on a dataset containing 327 mitotic and 406 non-mitotic cells via a support vector machine classifier in iterative cross-validation. The quantitative results showed that our DT-CWT based approach achieved superior classification performance with the accuracy of 87.94%, sensitivity of 86.80%, specificity of 89.89%, and the area under the curve (AUC) value of 0.94.

Index Terms— mitosis, wavelet transform, multi-resolution representation, breast cancer histopathology.

1. INTRODUCTION

According to the World Health Organization (WHO), breast cancer (BCa) is the second most lethal cancer diagnosed in women [1]. To diagnose breast cancer in histopathology, biopsy is performed and the stained histology slides are observed under microscope and graded by pathologists. Based on the Nottingham Grading System [2], a well-known international grading system for breast cancer recommended by the WHO, mitotic count is one of the main parameters in breast cancer grading as it gives an evaluation of the proliferation and aggressiveness of the tumor. Manual counting of mitosis is a tedious process and often subject to sampling bias due to massive histological images. Moreover, previous studies revealed that this process subject to considerable inter- and intra-reader variation is up to 20% between central and institutional reviewers in tumor prognosis [3].

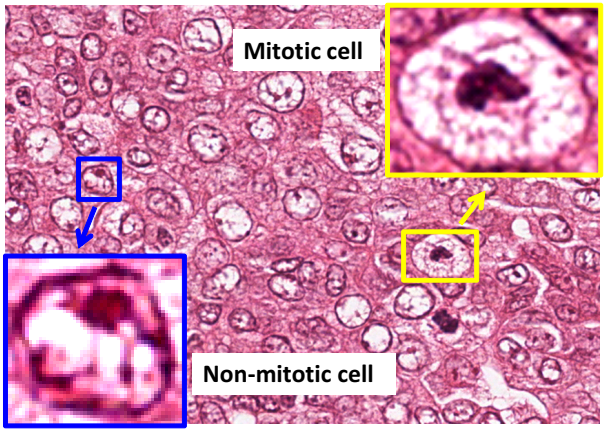

Fig. 1. An example of breast cancer histopathological image containing mitotic and non-mitotic cells. Two magnified image patches show that mitotic and non-mitotic cells may exhibit similar color and shape.

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 image analysis to facilitate BCa diagnosis and prognosis [4]. In histopathological image analysis, mitosis classification in discriminating between mitotic and non-mitotic cells is a difficult task due to the various and irregular shapes of mitotic cells under four main phases (i.e., prophase, metaphase,

* Corresponding author's email: zcqin@buaa.edu.cn. ♯ These two authors contributed equally to this paper. This work was partially funded by the National Science Foundation of China No. 61305047 and the SRF for ROCS, SEM, China.

anaphase, and telophase) [5]. It is also observed that non-mitotic cells may exhibit similar color and shape to mitotic cells as shown in Fig.1.

Recently, a few computer-aided diagnosis techniques have been developed to automatically detect and classify mitosis using different features [6, 7]. For example, Sommer *et al.* [8] employed intensity, shape (e.g., circularity), and texture features (e.g., haralick, statistical geometric features) to perform a mitosis classification on a pixel basis. Tashk *et al.* [9] introduced an automatic mitosis detection method using completed local binary patterns based on a pixel-level classification. Although these pixel-wise texture 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.

In this paper, we present a wavelet based approach to distinguish mitotic from non-mitotic cells in breast cancer histopathology. The objects of interest containing tissue cells are decomposed into multi-resolution representations via a dual-tree complex wavelet transform (DT-CWT), which provides near shift invariance and good directional selectivity compared to the standard wavelet transform [10]. There is an inter-scale dependency, most notable between a wavelet coefficient at a coarse level and the four coefficients at the previous adjacent level that correspond to the same location in the image [11, 12]. This property allows extraction of important textural features that are more likely to appear at all different decomposition levels. This is particularly useful for characterizing small objects, such as mitotic and non-mitotic cells present in histopathological images. Five widely-used statistical features are then computed on each wavelet subband to better characterize tissue cells at multiple scales and orientations having mitosis discrimination from non-mitotic cells. These extracted statistical features are evaluated through a support vector machine (SVM) classifier on a cell basis. The flowchart of the presented method is illustrated in Fig.2.

The rest of the paper is organized as follows. Section 2 describes the detailed methods. The experimental results and discussion are presented in Section 3. Finally, the concluding remarks are drawn in Section 4.

2. METHODS

2.1. Data Description

The DT-CWT based approach is evaluated on the MITOS database[1], which is provided by the International Conference for Pattern Recognition 2012 contest on Mitosis Detection in Breast Cancer Histological Images. It contains a set of 5 breast cancer biopsy slides that stained by hematoxylin and eosin (H&E) and examined under a $40\times$ magnification lens. In each slide, the pathologists selected 10 images as a size of

[1]MITOS database is available at:http://ipal.cnrs.fr/ICPR2012.

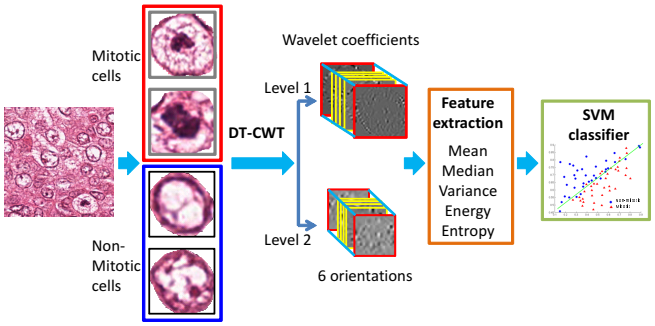

Fig. 2. The flowchart illustrating that histopathological images containing mitotic and non-mitotic cells are decomposed into multi-resolution representations via DT-CWT. The statistical features are extracted on wavelet subbands to well characterize tissue cells in the wavelet domain and evaluated by the SVM classifier.

$512 \times 512\,\mu m^2$ termed as high power fields (HPF). For each HPF image as shown in Fig.1, two experienced pathologists manually annotated mitosis as the ground truth.

2.2. Image Preprocessing

To generate mitosis and non-mitosis dataset, 327 mitotic and 406 non-mitotic cells are segmented using a semi-automatic segmentation algorithm, named distance regularized level set evolution (DRLSE) based method [13]. 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 DRLSE segmentation method requires initialization for the zero level set function, which is defined based on the ground truth for mitotic cells and the manual annotation from a pathologist for non-mitotic cells, respectively.

Before performing feature extraction, the segmented cell images are preprocessed via a contrast enhancement method to improve the image quality. In addition, to reduce the edge effect between cell boundary and background in the wavelet transform, the background of image patch is assigned a dominant color of tissue cell stained by H&E.

2.3. Feature Extraction

2.3.1. Multi-resolution Representations via DT-CWT

The dual-tree complex wavelet transform [10] has been proved to have several advantages over the traditional wavelet transform in terms of shift invariance and directional selectivity. The segmented cell images can be decomposed into multi-resolution representations in different spatial scales and orientations. In this work, we adopt a 2-level DT-CWT with 6 orientations of $[\pm 15°, \pm 45°, \pm 75°]$.

For an input image with a size of $m \times n$, we obtain a set of real and imaginary coefficients at different decomposition

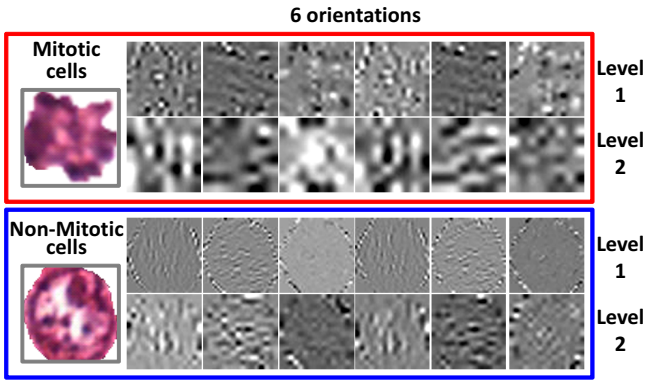

Fig. 3. Two examples of DT-CWT representations with 2 decomposition levels and 6 orientations for a mitotic cell and a non-mitotic cell. The wavelet subband images revealed that important texture features appearing in coarse level are more likely present in the adjacent level. This inter-scale dependencies allow better extraction of salient features having discrimination capability in distinguishing mitotic and non-mitotic cells.

levels and orientations. As shown in Fig.2, the subband image of level l has a size of $m/2^{l-1} \times n/2^{l-1}$. For each level, the DT-CWT produces 6 wavelet subband images reflecting the high frequency response on the corresponding 6 orientations.

Fig.3 shows two examples of DT-CWT representations for a mitotic cell and a non-mitotic cell, respectively. For a better visualization, the subband images for the level 2 have been enlarged to the same size as the subband images of level 1. 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. This property allows better extraction of salient features having discrimination capability in distinguishing mitotic and non-mitotic cells.

2.3.2. Statistical Features

In this work, we extract five popular statistical features, including mean (F_1), median (F_2), variance (F_3), energy (F_4), and entropy (F_5), on each wavelet subband $W_{l,\theta}$, $l \in \{1, ..., L\}$, $\theta \in \{\pm 15°, \pm 45°, \pm 75°\}$. The energy and entropy are computed as:

$$F_4 = \sum_{i=1}^{m} \sum_{j=1}^{n} g^2(i,j) \tag{1}$$

$$F_5 = \sum_{i=1}^{m} \sum_{j=1}^{n} g(i,j) \times \log[g(i,j)] \tag{2}$$

where $g(i,j) \in W_{l,\theta}$ represents the wavelet coefficient at location (i,j) in wavelet subband image with a size of $m \times n$. For each wavelet subband, the features are calculated for real and imaginary part of coefficients, respectively. All the features across wavelet subbands are combined to form a feature vector $\mathcal{F}$ to characterize mitotic and non-mitotic cells. Before performing a cell level classification, we rescale the range of features in order to make the features independent to each other. The feature vector is rescaled to the range of $[0, 1]$. The objective of this work is to investigate the potential application of multi-resolution approach to improve the accuracy of mitosis classification in breast cancer histopathology. Thus, we only consider five commonly used statistical features. Other statistical features, such as second or higher order statistics, can also be included in this framework.

2.3.3. SVM-based Classification via Cross-validation

Support vector machines, a well-known classifier, is applied to evaluate the performance of the DT-CWT based method. In SVM, kernel functions are used to map the input data into a higher dimension space where the data are supposed to have a better distribution, and then an optimal separating hyperplane is chosen. Here we utilize the radial kernel and tune the parameters to yield the best classification results.

The mitotic and non-mitotic datasets 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 $\mathcal{Z}_{tes}$. An iterative 2-fold cross validation scheme is utilized to train a SVM classifier and evaluate the performance of classification using the feature set $\mathcal{F}$.

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1. Experimental Design

In the SVM classification, the cross-validation process was repeated 5000 trials to reduce random errors. The classification performance was quantitatively measured by the classification accuracy (AC), sensitivity (SN), and specificity (SP). In addition, a receive operating characteristics (ROC) analysis was utilized to evaluate the performance of the SVM classifier. When using normalized units, the area under the curve (AUC) was calculated to measure the wavelet-based features' ability in distinguishing between mitotic and non-mitotic cells. The associated mean μ and standard deviation σ were computed for each metric.

3.2. Classification Results

Table 1 lists the SVM classification results measured by AC, SN, SP, and AUC. To evaluate the effect of DT-CWT in the classification performance, we compared the wavelet-based statistical features to the the same features directly computed on the image intensities. The ROC curves for these two methods are illustrated in Fig.4. To further evaluate the classification performance, we compared the DT-CWT based method

Table 1. The classification results ($\mu \pm \sigma$) measured by classification accuracy (AC), sensitivity (SN), specificity (SP), and the area under the curve (AUC). The comparison results suggested that the DT-CWT based statistical features are able to better distinguish texture patterns between mitotic and non-mitotic cells compared to the same features computed on image intensities.

Measure	With DT-CWT	Without DT-CWT
AC	**87.94**$\% \pm 1.09\%$	$83.08\% \pm 3.59\%$
SN	**86.80**$\% \pm 4.78\%$	$69.47\% \pm 4.26\%$
SP	**89.89**$\% \pm 1.42\%$	$87.74\% \pm 1.96\%$
AUC	**0.94** ± 0.02	0.83 ± 0.01

Table 2. Comparison between methods.

Measure	DT-CWT	XICA[14]	HLW[8]	SMOTE[5]
AUC	**0.94**	0.84	0.91	0.74

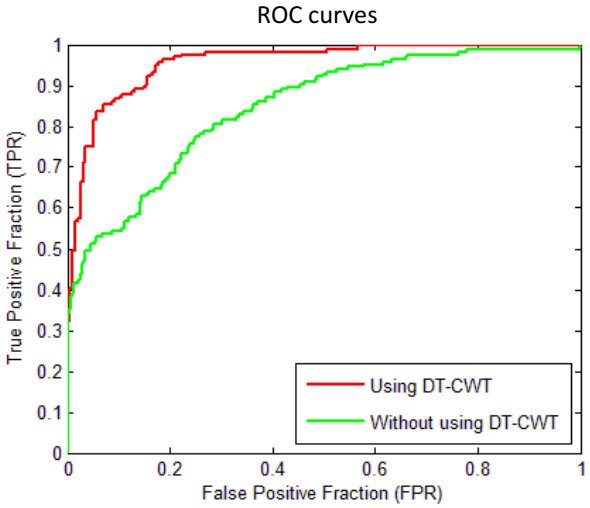

Fig. 4. ROC curves illustrating the comparison of classification performance using the wavelet-based statistical features and the same features computed on image intensities.

to the other three approaches using the same MITOS dataset. The first approach [14] utilized the eXclusive independent component analysis (XICA), an extension of a generic ICA, to automatically detect mitosis based on the components of differences between two classes of positive and negative patterns. The second approach [8] employed a hierarchical learning workflow (HLW) in conjunction with SVM to distinguish between mitotic and non-mitotic cells using object shape and texture features. The third approach [5] combined statistical and morphological features extracted from selected multiple color channels and applied a synthetic minority oversampling technique (SMOTE) [15] to reduce mitosis classification bias. The comparison results using the AUC measure are listed in Table 2.

The quantitative results showed that the DT-CWT based method achieved better classification performance in distinguishing between mitotic and non-mitotic cells compared to the the same features computed on the image intensities. Owing to the irregular shape and various texture of mitosis in different phases, a DT-CWT based approach to analyzing mitotic and non-mitotic images is able to capture the important texture features and architectural arrangement of individual histological structures (glands and nuclei) present in different resolution levels and orientations. Further, the inter-scale dependencies between wavelet subbands allow extraction of discriminative features of mitosis appearing at all resolution levels.

The experiments were performed in the Matlab R2031a platform using an Intel Duo E7500 2.94GHz machine with a 4GB RAM. The average running time per cell is 0.046 and 0.051 second(s) for feature extraction and classification, respectively.

4. CONCLUDING REMARKS

We presented a multi-resolution based approach to discriminate mitotic from non-mitotic cells using breast cancer histopathological images in a complex wavelet domain. Since mitotic cells have different shape and texture for each main evolution phase, there is no simple way to characterize mitosis based on shape and pixel intensities. The property of inter-scale dependencies across wavelet subbands allows to capture the discriminative features that appear at both coarse level and previous adjacent level. The statistical features extracted from wavelet coefficients at multiple scales and orientations are able to capture the salient information within tissue cells to distinguish mitotic and non-mitotic cells.

The DT-CWT based method has been evaluated on the MITOS dataset used in the ICPR 2012 contest via a SVM classifier. The method achieved an improved accuracy of 87.94%, sensitivity of 86.80%, specificity of 89.89%, and the AUC values of 0.94 compared to the same features computed from image intensities. The quantitative results suggested that the wavelet-based statistical features are able to capture important texture attributes related to mitosis development and differentiate mitotic from non-mitotic cells.

5. REFERENCES

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

[2] C.W.Elston and 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*, vol. 19, pp. 403–410, 1991.

[3] H.A.Teot, R.Sposto, A.Khayat, S.Qualman, G.Reaman, and D.Parham, "The problems and promise of central pathology review development of a standardized procedure for the children oncology group," *Pediatr. Dev. Pathol.*, vol. 10, pp. 199–207, 2007.

[4] M.N.Gurcan, L.Boucheron, A.Can, A.Madabhushi, N.Rajpoot, and B.Yener, "Histopathological image analysis: A review," *IEEE Rev. Biomed Eng.*, vol. 2, pp. 147–171, 2009.

[5] H.Irshad, L.Roux, and D.Racoceanu, "Multi-channels statistical and morphological features based mitosis detection in breast cancer histopathology," in *EMBS*. IEEE, 2013, pp. 6091–6094.

[6] A.Albayrak and G.Bilgin, "Breast cancer mitosis detection in histopathological images with spatial feature extraction," in *SPIE*, 2013, pp. 1–6.

[7] H.Irshad, "Automated mitosis detection in histopathology using morphological and multi-channel statistics features," *J Pathol. Inform.*, vol. 4, pp. 1–10, 2013.

[8] C.Sommer, F.A.Hamprecht, L.Fiaschi, and D.W.Gerlich, "Learning-based mitotic cell detction in histopathological images," in *ICPR*, 2012, pp. 2306–2309.

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

[10] N.Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Appl. Computt. Harmon. Anal.*, vol. 10, pp. 234–253, 2001.

[11] T.Wan, N.Canagarajah, and A.Achim, "Segmentation-driven image fusion based on alpha-stable modeling of wavelet coefficients," *IEEE Trans. on Multimedia*, vol. 11, pp. 624–633, 2009.

[12] T.Wan, N.Canagarajah, and A.Achim, "Segmentation of noisy colour images using Cauchy distribution in the complex wavelet domain," *IET Image Processing*, vol. 5, pp. 159–170, 2011.

[13] C.Li, C.Xu, C.Gui, and M.D.Fox, "Distance regularized level set evolution and its application to image segmentation," *IEEE Trans. Image Processing*, vol. 19, pp. 3243–3254, 2010.

[14] C.H.Huang and H.K.Lee, "Automated mitosis detection based on eXclusive independent component analysis," in *ICPR*, 2012, pp. 1856–1859.

[15] N.V.Chawla, K.W.Bowyer, L.O.Hall, and W.P.Kegelmeyer, "Smote: Synthetic minority oversampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.