Signal Processing 122 (2016) 1–13

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Automatic cell nuclei segmentation and classification of breast cancer histopathology images

Pin Wang[a,*], Xianling Hu[a], Yongming Li[a,b], Qianqian Liu[a], Xinjian Zhu[c]

[a] College of Communication Engineering of Chongqing University, Chongqing 400030, PR China
[b] Department of Medical Imaging of College of Biomedical Engineering, Third Military Medical University, Chongqing 400038, PR China
[c] Research Institute of Field Surgery, Daping Hospital, Third Military Medical University, Chongqing 400038, PR China

ARTICLE INFO

Article history:
Received 28 June 2015
Received in revised form
4 November 2015
Accepted 15 November 2015
Available online 26 November 2015

Keywords:
Breast cancer
Nuclei segmentation
Wavelet decomposition
Curvature scale space corner detection
Chain-like agent genetic algorithm
Support vector machine classification

ABSTRACT

Breast cancer is the leading type of malignant tumor observed in women and the effective treatment depends on its early diagnosis. Diagnosis from histopathological images remains the "gold standard" for breast cancer. The complexity of breast cell histopathology (BCH) images makes reliable segmentation and classification hard. In this paper, an automatic quantitative image analysis technique of BCH images is proposed. For the nuclei segmentation, top-bottom hat transform is applied to enhance image quality. Wavelet decomposition and multi-scale region-growing (WDMR) are combined to obtain regions of interest (ROIs) thereby realizing precise location. A double-strategy splitting model (DSSM) containing adaptive mathematical morphology and Curvature Scale Space (CSS) corner detection method is applied to split overlapped cells for better accuracy and robustness. For the classification of cell nuclei, 4 shape-based features and 138 textural features based on color spaces are extracted. Optimal feature set is obtained by support vector machine (SVM) with chain-like agent genetic algorithm (CAGA). The proposed method was tested on 68 BCH images containing more than 3600 cells. Experimental results show that the mean segmentation sensitivity was 91.53% (±4.05%) and specificity was 91.64% (±4.07%). The classification performance of normal and malignant cell images can achieve 96.19% (±0.31%) for accuracy, 99.05% (±0.27%) for sensitivity and 93.33% (±0.81%) for specificity.

1. Introduction

Breast cancer is the leading type of malignant tumor observed in women. Early detection of cancer is very important for successful treatment. Diagnosis from histopathological images remains the "gold standard" for breast cancer. Manually analyzing numerous biopsy slides by pathologist is labor intensive and has suboptimal reproducibility. Thanks to recent advances in digital pathology, the automatic image analysis method has the potential to overcome the subjective interpretation and reduce the workload. Computer-aided diagnosis (CADx) scheme is becoming an important tool to assist pathologist in breast cancer detection and diagnosis. The CADx scheme consists of two phases, which are segmentation phase and classification phase [1–3].

Segmentation of nuclei is an important first step towards automatic analysis of BCH images. Several algorithms for segmentation of nuclei in BCH images have been proposed. Most of them revolve around watershed segmentation, active contours, pixel classification or combination varieties, supplemented by different pre-processing and post-processing phases. Watershed segmentation is always improved by obtaining

---

[*] Corresponding author.
E-mail address: wangpin@cqu.edu.cn (P. Wang).

http://dx.doi.org/10.1016/j.sigpro.2015.11.011
0165-1684/© 2015 Elsevier B.V. All rights reserved.

locations as markers that can mark the objects of interest and the background [4–8]. But these techniques suffer from over-segmentation and did not work well for the overlapping cells. The objective of active contours is to find a minimum energy fit of the moving contours, and the algorithm is always combined with a nuclei detection method. These kinds of methods initially define a large number of candidate regions and then select the ones that can present correctly segmented nuclei [8–13]. However, these models have some limitations in convergence, the optimization problem involved leads to uncertainty and poor stability of result of segmentation. Moreover, clustering based methods such as *K*-means [14] and unsupervised or supervised machine learning [12,15] have been applied for the segmentation of the cell nuclei of breast cancer. These methods require explicit prior knowledge of the image structure and the computational complexity is relatively high. Due to the high variability of the tissue appearance, reliable cell nuclei segmentation of BCH images is still a challenging task.

After nuclei are precisely segmented, classification phase is implemented. The most important aspects of the classification performance are the features extracted and the classification algorithms. Some researchers have studied the analysis of BCH images by proposing new features or considering different classification algorithms. Most of the extracted features are morphology-based and texture-based features. Some new features were extracted such as textural features using a critical exponent analysis (CEA) [16] and complex Daubechies wavelets [13,17], distribution-based features of nuclei [18], etc. Most classification algorithms utilized in literature revolve around Support Vector Machine (SVM) [19], *k*-Nearest Neighbor (*k*-NN) [13], Naive Bayes (NB) [18], fuzzy *c*-means (FCM) [20], neural network [19] or a combination of the above mentioned algorithm [18]. Some other classification approaches such as Decision Tree (DT) and partial least squares regression [21] have been applied for breast cell classification as well. Almost all the literature applied all extracted feature as the input of classifiers. However, feature selection can improve the classification accuracy and the reliability [18,22].

In this paper, an automatic CADx scheme of BCH images is proposed. For the nuclei segmentation of BCH images, top-bottom hat transform is applied to enhance grayscale image. Wavelet decomposition and multi-scale region-growing (WDMR) are combined to obtain regions of interest (ROIs), a double strategy splitting model (DSSM) containing adaptive mathematical morphology and Curvature Scale Space (CSS) corner detection method is applied to split overlapped cells for better accuracy and robustness. For the classification of the cell nuclei, 4 shape-based features and 138 textural

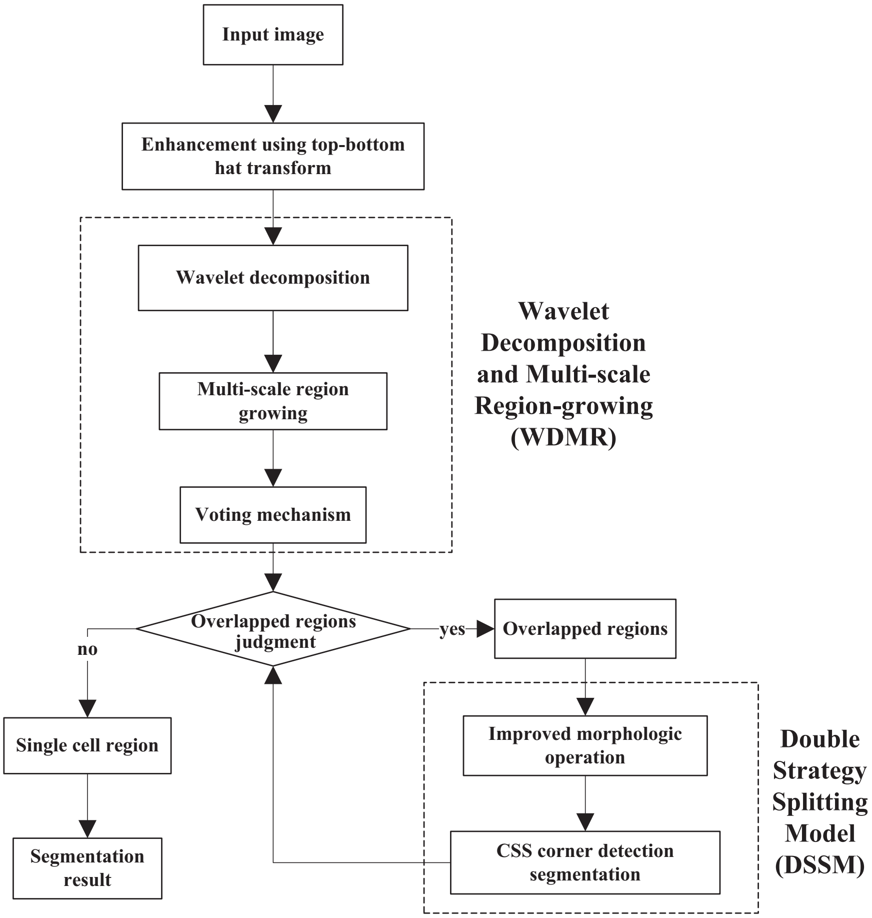

Fig.1. Block-diagram of the proposed segmentation method.