Digital Signal Processing ••• (••••) •••–•••

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Feature-based analysis of cell nuclei structure for classification of histopathological images

Pin Wang [a,*], Sha Xu [a], Yongming Li [a,b], Lirui Wang [a], Qi Song [a]

[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

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Histopathological images
Nuclei structure
Feature analysis
Feature selection methods

ABSTRACT

Pathological examination of histopathological images remains the main method of cancer diagnosis. We proposed a novel automatic feature-based analysis scheme for classification of the histopathological images. An animal model of intestinal carcinogenesis-multiple intestinal neoplasia mouse model was used to evaluate the feasibility of the nuclear structure feature for detecting early-stage carcinogenesis and assessing cancer risk. Firstly, the cell nuclei are segmented based on collaborate cell localization and the improved morphology method. Then several types of features, including shape features, statistical features and textual features (Gabor and Markov random field features) are extracted. Feature selection methods including wrapper, filter and the maximum relevance-minimum multicollinearity are applied to obtain the optimal feature set. Experimental results show that the proposed segmentation method can automatically segment histopathological images and has effective segmentation results. The maximum relevance-minimum multicollinearity method outperformed all other methods in term of classification accuracy. The textual features can effectively improve the characterization of cell nuclei structure and feature selection methods can get better classification results.

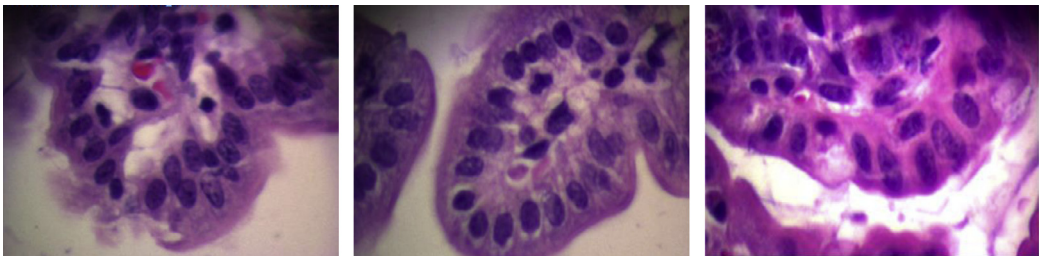
1. Introduction

Today, cancer is one of the most common diseases in the world, and the prevention, early diagnosis and treatment of cancer have attracted more attention and discussion than before. Pathology examination remains the gold-standard for cancer detection [1,2]. It is a means of inspecting characteristics at cellular level that serves as predictor for cancer risk. However, it is hard to confirm an early-stage carcinogenesis case by tracking the cellular change, so a mouse model of spontaneous carcinogenesis is applied. Furthermore, manually analyzing numerous histopathological images by pathologist is an intensive labor and the results cannot be always guaranteed, especially for the uninvolved (histological normal appearing) cell image. In this case, feature-based analysis scheme [3,4] is becoming an important tool to assist pathologist in cancer detection and diagnosis. The scheme consists two phases, which are segmentation and classification phase. The main challenge of the technique remains in accurate segmentation of the cell nuclei and obtaining optimal feature set that can better characterize cell nuclei structural alterations associated with carcinogenesis.

Several algorithms for nuclei segmentation of histopathological image have been proposed, which revolved around active contours [5,6], marker-controlled watershed segmentation [7,8], thresholding method [9,10], and pixel classification [11,12] or combination varieties [13]. The active contours can describe the contour of objects in an image using gradient information by minimizing an energy function, but this algorithm cannot get a balance between the adaptability of topological changes, the robustness of boundary gaps and image noise [14]. The process of marker-controlled watershed segmentation usually starts from specific pixels called markers [15] and gradually floods the surrounding regions. But this method cannot work well with the overlapped cells. The thresholding method converts the intensity image into a binary image by assigning all pixels to the value one or zero if the intensity is above or below the threshold. This method can only be applied to simple images. Moreover, some classification methods like K-means [16,17] or supervised or unsupervised machine learning methods [18,19] have also been applied to the segmentation of histopathological images. These techniques require explicit prior knowledge of the image structure and the computational complexity is relatively high. From earlier threshold and active contour techniques till now, various methods have been proposed for cell nuclei segmentation of histopathological image. However, the nuclei are of-

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

https://doi.org/10.1016/j.dsp.2018.03.008
1051-2004/



(a) (b) (c)

Fig. 1. Representative images of each class: (a) normal, (b) uninvolved, (c) dysplastic.

ten presenting complex and irregular visual aspect, so reliable cell segmentation of the images is still a challenging task.

After the cell nuclei are accurately segmented, the classification phase is implemented. The extracted features and the classification method will greatly affect the classification performance. Most extracted features are morphology-based [20] and statistical-based [21] features. The 2D Gabor filters [22] can provide good resolution both in temporal and frequency domain, and provide the optimal basis to extract local features. The Markov random field (MRF) [23] model combines the spatial correlation of each pixel and describe texture features in different orientations and forms. In this paper, Gabor filters and MRF model are also applied on the image to extract the features of the nuclei. At first, researches only used one texture feature for classification like co-occurrence matrix [24], but it can only reflect one aspect of characteristics. In order to represent the structure characteristics of the cell nuclei better, the multiple-features has been a trend [25,26]. As for the multi-features classification, the feature selection [27–29] step plays an important role in the whole procedure because the original dataset may contain a large number of redundant and irrelevant features. The feature selection method can be categorized into wrapper and filter method by considering the classifiers used in the feature selecting process. The method based on chain-like agent genetic algorithm (CAGA) [30] is a typical wrapper feature selection method and the individuals within the population are the feature subset candidates. This method relied much on the pre-defined parameters like the initial population, and the wrapper methods always tend to be time-consuming. The filter methods use some measurement criteria like correlation thresholding [31], entropy [32], separability distance (SD) [33] and so on to find the optimal subset. The filter model is also considered as having simple search structure. Among the numerous measurement criteria, the correlation coefficient is more appropriate for the multi-feature classification since this measurement can evaluate the specific correlation of each type of feature.

In the present study, we evaluate the feasibility of structure features of cell nuclei extracted from the histopathological images for detecting early-stage cancer. For this purpose, an animal model of intestinal carcinogenesis – Apc^{Min} mouse model is used for analysis. This model is chosen for this analysis because it has been shown to be a robust indicator of carcinogenesis with changes in cell structure [34]. The model has *adenomatous polyposis coil* (*Apc*) gene mutation that causes spontaneous development of multiple intestinal neoplasia (Min) [35], primarily in the small intestine. An automatic feature-based analysis scheme of microscopy images from histological specimens is proposed. For the nuclei segmentation, top-bottom hat transformation is applied to enhance the gray scale image. An automatic segmentation method based on collaborate regions of interest localization and the improved morphology method (CRL-IMM) is presented. Firstly, the collaborate ROI localization algorithm based on Wavelet transform and Mean shifting clustering is used to obtain regions of interest (ROIs) because the gray-level variations along different directions and the density of the nuclei and background is different. Then in order to split overlapped cells, the improved morphology method is applied. After segmentation, four different types of features are extracted from the cell nuclei, they are 4 shape-based features, 144 statistical features based on color spaces, 60 Gabor features based on the square energy and mean amplitude of the different direction and different scale, and 12 Gaussian MRF features. From it, we intend to provide a better understanding about how these features relate to the underlying physiology. In order to improve the classification performance, a feature selection method of maximum relevance-minimum multicollinearity (MRmMC) [36] is applied to obtain the optimal feature set. The major contribution of this paper is listed as following: 1) In order to investigate the feasibility of the scheme for detecting early-stage cancer, Apc^{Min} mouse model is used for the feature analysis; 2) An automatic segmentation technique based on collaborate ROI localization and the improved morphology method is proposed to extract the cell nuclei; 3) Different types of features are studied for detecting the dysplastic changes including shape-based feature, statistical features, Gabor features and Gaussian Markov features; 4) Feature selection method MRmMC is used to improve the classification performance of histopathological images.

2. Materials

Three wide type mice at the age of six weeks, three age-matched Min mice, and three Apc^{Min} mice at the age of 4.5 months were sacrificed. Those Apc^{Min} mice at the age of six weeks had only a few microscopically-visible micro-adenomas; whereas those Min mice at the age of 4.5 months had developed visible multiple adenomatous polyps in their small intestine. The histology slides were made using the small intestine of the mice. The cell images were captured by a digital camera mounted on a microscope with 40× magnification. The images were reviewed by expert cytopathologists to classify. We have obtained 55 normal cell images (normal cells from wide type mice), 54 uninvolved cell images (histological normal appearing cells from 6 weeks Apc^{Min} mice) and 17 dysplastic cell images (dysplastic cells from 4.5 month Apc^{Min} mice), respectively. The representative pictures of each class are shown in Fig. 1. Compared with Fig. 1(a) and (b), the cell nuclei of dysplastic image (c) is larger and the shape is more irregular. It can be observed that the difference between normal image (a) and uninvolved image (b) is hard to be recognized and the classification of the two classes remains a challenge. The size of all images is 591×832.

3. Method

3.1. Segmentation

While all the subsequent processes are based on the segmentation results, it is very important that all the cell nuclei were segmented accurately. An automatic segmentation algorithm CRL-IMM

Please cite this article in press as: P. Wang et al., Feature-based analysis of cell nuclei structure for classification of histopathological images, Digit. Signal Process. (2018), https://doi.org/10.1016/j.dsp.2018.03.008