Micron 79 (2015) 29–35

Contents lists available at ScienceDirect

Micron

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

Performance evaluation of maximal separation techniques in immunohistochemical scoring of tissue images

K.A. Shahul Hameed [a,*], A. Banumathi [b], G. Ulaganathan [c]

[a] Research Scholar, Department of ECE, Thiagarajar College of Engineering, Madurai, Tamil Nadu, India.
[b] Department of ECE, Thiagarajar College of Engineering, Madurai, Tamil Nadu, India
[c] Department of Oral & Maxillofacial Surgery, Best Dental Science College, Madurai, Tamil Nadu, India

ARTICLE INFO

Article history:
Received 7 January 2015
Received in revised form 28 July 2015
Accepted 28 July 2015
Available online 5 August 2015

Keywords:
p53 immunostain
Gray-level co-occurrence matrix
Global entropy
Thresholding
Watershed transform
Automatic scoring

ABSTRACT

This paper presents an automatic scoring method for p53 immunostained tissue images of oral cancer that consist of tissue image segmentation, splitting of clustered nuclei, feature extraction and classification. The tissue images are segmented using entropy thresholding technique in which the optimum threshold value to each color component is obtained by maximizing the global entropy of its gray-level co-occurrence matrix and clustered cells are separated by selectively applying marker-controlled watershed transform. Cell nuclei feature is extracted by maximal separation technique (MS) based on blue component of tissue image and subsequently, each cell is classified into one of four categories using multi-level thresholding. Finally, IHC score of tissue images have been determined using Allred method. A statistical analysis is performed between immuno-score of manual and automatic method, and compared with the scores that have obtained using other MS techniques. According to the performance evaluation, IHC score based on blue component that has high correlation coefficients (CC) of 0.95, low mean difference (MD) of 0.15, and a very close range of 95% confidence interval with manual scores. Therefore, automatic scoring method presented in this paper has high potential to help the pathologist in IHC scoring of tissue images.

1. Introduction

Immunohistochemical (IHC) analysis plays an important role in early diagnosis and design of targeted therapies for specific tumors, and also research works related to computer-aided diagnosis (CAD) in order to improve the survival rate of cancer patients (Di Cataldo et al., 2012). This process became popular after the introduction of high-resolution microscopic scanners, which capture whole-slide of a tissue image under study (Fuchs and Buhmann, 2011; Bueno et al., 2012). In general, manual process of extracting information form a tissue section is tedious and also consume more time due to the presence of several tens of thousand cell nuclei, lymphocytes, blood vessels, etc. Moreover, the outcomes are purely subjective and found variation in the agreement of inter/intra observers (Gavrielides et al., 2011). Hence, CAD methods have been proposed to help the pathologist during IHC image analysis. Several works had been reported in the literatures about CAD systems (Krishnan et al., 2009, 2010, 2012; Huang and Lee, 2009; Sertel et al., 2010) including detection and grading of cancers such as breast, prostate, head and neck, etc., that were focused on variation on architectural patterns of histopathology and comparatively few methods were reported at IHC based cell-by-cell analysis. Generally, automatic IHC analysis is categorized according to the basic processing element namely pixels or objects. Pixel based image analysis is widely used in identifying immuno-positivity of tissue section over a specified area, in which RGB color component intensity of each pixel is considered. For example, Brey et al. (2003), have identified immuno-positive tissues by thresholding the normalized blue component of bright field diaminobenzidine (DAB) labeled tissue images, and compared the outcome to manual and other similar works reported in Ruifrok and Johnston (2001), Ma and Lozanoff (1999) and King et al. (2002).The selection of threshold value was considered as an important part of all these works and however, an inappropriate threshold could increase the probability of misclassification of pixels. In order to solve this problem, Choudhury et al. (2010), proposed averaged threshold measure (ATM), in which tissue score had been determined by thresholding the brown channel over all possible threshold value. ATM could not work well in few situations such as tissue folding, staining of non-cancerous cells, uneven

* Corresponding author.
E-mail address: ashkars@yahoo.com (K.A.S. Hameed).

http://dx.doi.org/10.1016/j.micron.2015.07.013
0968-4328/

fixation and particularly, tissue image scanned at high magnification. Moreover, pixel based approaches are mainly used in determining immuno-positive or immuno-ratio of tissue images and may not be suitable for cell-by-cell analysis.

The main objective of this work is to develop a simple automatic IHC scoring of tissue image, in which the characteristics of each cell and also group of cells need to be analyzed in the process of determining tissue scores. Therefore, object based approaches are suitable for cell-by-cell analysis based tissue image scoring. For example, Primkhajeepong et al. (2010), proposed a simple computer aided cell counting algorithm for tissue images of breast cancer that classified each cell into positive or negative according to feature extracted from b* channel of CIEL*a*b*. Similarly, Wienert et al. (2013), developed Cognition Master that had classified ki-67 cell nuclei into immuno-positive or negative using features such as size, color and shape localization. Kalinli et al. (2013), compared the performance machine learning approaches in IHC scoring of hormone receptor status of breast cancer tissue section. In all these works, segmentation and feature extraction play an important role in order to separate cell nuclei and also classify each nuclei into either one or more categories as per standard scoring methods, respectively. Many works related to tissue image segmentation had been reported in the literature include thresholding, edge detection, active contours, pixel classifiers, etc. Generally, thresholding based segmentation works well only on uniform image with bimodal histogram and moreover, histogram based thresholding techniques did not consider spatial correlation among pixel of an image. Due to these facts, two entirely different images with similar histogram pattern could produce same threshold value, which result in segmentation error. In most of the segmentation process, the conversion of color to gray scale image has produced improper threshold value due to decoupling of color information (Du et al., 2004). Moreover, edge based and region based active contour based segmentation methods were limited by contour initialization, convergence and overlapped nuclei structures. More specifically, these methods reportedly suffers due to various factors such as selection of threshold value, quality of training samples, overlapping nuclei structures, noise, an so on. In order to minimize the effect of these factors and improve the accuracy of tissue image segmentation, entropy thresholding and watershed transform has been adopted in this paper. After segmentation, feature of each cell nuclei has been extracted using MS techniques; classified into one of four categories and finally, Allred score is determined. This paper is organized as follows: Section 2, describes methodology which include tissue image segmentation, splitting of clustered cell nuclei, feature extraction and classification; Section 3, discuss the performance evaluation results and Section 4 gives conclusion.

2. Methodology

2.1. Dataset description

Oral cancer is frequently occurring common cancer worldwide, among 90% of which were diagnosed as oral squamous cell carcinoma (Ali, 2008; Das and Nagpal, 2002) and particularly, the incident rate was reported as 40–50% of all cancer in India (Sathiyapriya et al., 2013; Balakrishnan et al., 2010). Early diagnosis and timely treatment can improve the chance of cure and also the survival rate of oral cavity cancer patients. The chronic use of alcohol, tobacco and radiations are the main causes of oral cancer carcinogens that may alter the structure of gene and chromosomes. The most common of these alterations in human cancer are due to mutation in biomarkers, for example, p53 protein is regarded as an early carcinogenesis event of OSCC. Generally, p53 protein has a very short half-life in normal tissues that are normally undetectable. However, the mutant p53 gene has prolonged half-life, i.e., it can remain longer in the tissues and accumulate in cell nuclei to the levels that can be easily detectable using p53 immunostaining procedure (Mehrotra et al., 2006; Polanska et al., 2014; Swaminathan et al., 2012). Hence, p53 protein expression deserves particular attention in early detection of oral cancer. The presence of p53 protein can be generally identified when cell nuclei is stained with brown color (Reibel, 2003). In this paper, a total of 12 tissue sections are collected from the Department of Oro-Maxillofacial Surgical & Medical Science, Faculty of Dentistry, University of Malaya, Kuala Lumpur, Malaysia that were stained with p53 immunohistochemistry and scanned using whole-slide scanner with a resolution of 128,653 × 288,108 and an average memory size of 300 MB. For experimental purpose, 20 representative regions are selected using a panoramic viewer at 40× magnification of size 975 × 620 pixels, without image artifacts such tissue folds, pen-marks and shadows that consist of an average of 266 cells/image.

2.2. Tissue image segmentation

Algorithm 1. Tissue Image Segmentation.

Input: p53 immunostained Image
Output: Segmented Image
Step 1: **Preprocessing of input image**
For each color component R,G, & B
Step 2: **Compute co-occurrence matrix** (P_{ij})
Step 3: **To obtain threshold value** (t_R, t_G, t_B) **for each color component - do** *step 3.1 to 3.4*
For $t = 0$ *to* 255
Step 3.1: **Divide** P_{ij} **into four quadrants and compute their probabilities** $P_K(t)$, $K = 1, 2, 3 \& 4$
Step 3.2: **Normalize the probability of pixel intensity transition within quadrants** $P_{ij/K}(t)$
Step 3.3: **Determine local, joint, and global entropy**
Step 3.4: **Find the threshold value** t, **which gives maximum global entropy**
End for loop
Step 4: **Thresholding each color component using threshold value obtained in** *step 3*
Step 5: **Combine the output of** *step 4*

In this paper, entropy thresholding (Du et al., 2004; Shannon, 2001; Chang et al., 1994) technique has been adopted and various steps involved in segmentation of tissue image are given in Algorithm 1. Prior to segmentation, each color component of the tissue images are separately preprocessed using Gaussian smoothing and principal component analysis (PCA) (Sertel et al., 2010); Gaussian smoothing removes noise and minimizes the variance among the pixels of cell nuclei and background; PCA increases the dynamic range of color components by projecting each one onto the first principal component associated with the highest variance. Hence, the resulting image had the highest contrast, which could enhance the process of obtaining the optimum threshold value. Subsequently, the co-occurrence matrix (P_{ij}) of each color component of dimension $M \times N$ has been computed as per the following equation

$$P_{ij} = \frac{1}{M \times N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \begin{cases} 1 & \text{if } f_z(x,y) = i \,\&\, f_z(x+\delta_x, y+\delta_y) = j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $f_z(x, y)$ is gray-level value of the color component $z = R, G, B$, and (δ_x, δ_y) is the distance vector pair of two arbitrary pixels i & j, For example, $(0, 1)$, $(-1, 1)$, $(-1, 0)$, and $(-1, -1)$ are the pair of pixels of eight neighbors along the direction of 0°, 45°, 90° and 135°. In this study, P_{ij} has been determined for the distance vectors $(0, 1)$ and $(-1, 1)$, which had been chosen by extensive experimental procedure of comparing segmentation outputs for different possible distance pairs with manual method. Later, color components are independently thresholded, hole-filling operation is performed to