Engineering Applications of Artificial Intelligence 31 (2014) 44–52

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Supervised leukocyte segmentation in tissue images using multi-objective optimization technique

Mukesh Saraswat *, K.V. Arya

ABV-Indian Institute of Information Technology and Management Gwalior, India

ARTICLE INFO

Available online 5 October 2013

Keywords:
Leukocyte segmentation
Non-dominated sorting genetic algorithm-II
Receiver operating characteristic curve
Threshold

ABSTRACT

Automated leukocytes segmentation in skin section images can be utilized by various researchers in animal experimentation for testing anti-inflammatory drugs and estimating dermatotoxicity of various toxic agents. However, complex morphological structure of skin section degrades the performance of leukocytes segmentation due to the extraction of vast number of artifacts/noise along with leukocytes. Rare works have been done to reduce such artifacts. Therefore, in this paper, a supervised methodology for leukocytes segmentation from the images of inflamed mice skin sections is introduced. The method is based on threshold based binary classifier to reduce the artifacts. The optimum values of thresholds are calculated using multi-objective optimization technique, non-dominated sorting genetic algorithm-II (NSGA-II) and receiver operating characteristic (ROC) curve. The experimental results confirm that the proposed method is prompt and precise to segment the leukocytes in highly variable images.

1. Introduction

Inflammatory cells, also known as leukocytes or white blood cells (WBCs), are cells of the immune system which defend the body against both infectious disease and foreign materials (Kumar et al., 2010). These cells are available in blood and infiltrated into tissues at the time of injury. These cells are analyzed manually by expert pathologists for disease identification and drug development which is a time consuming and biased method. Therefore, automation of leukocytes analysis can assist the pathologists in the quantification of leukocytes. Some useful algorithms have been developed for separation of leukocytes in blood smears images (Pan et al., 2011; Ko et al., 2011; Theera-Umpon and Dhompongsa, 2007; Angulo and Flandrin, 2003; Tycko et al., 1976), but rare efforts have been made to automate leukocytes segmentation in tissue section images. Leukocytes can be found in different organs but skin tissue has the most complex structure. Further, skin is the first line of defence for any infection or injury. Most of the animal models for inflammation consider skin as a major organ. Therefore, in this paper skin tissue images are considered for leukocytes segmentation.

Leukocyte can be identified with the structural and textual information of its nucleus and cytoplasm. Leukocytes identification system goes through the process of leukocytes segmentation, their feature extraction and classification. But, automatic segmentation of leukocytes from the complex morphological background of tissue section images also extracts the vast number of artifacts/noise. These artifacts are created due to improper handling prior to fixation, processing of tissue, its sectioning, and staining. Some of the artifacts appear like tiny points or Gaussian noise within skin tissue. The artifacts are identified using their morphological characteristics mainly the nuclear characteristics. In general, these structures do not contain any nucleus but, a halo dark zone misrepresent to a nucleus. Rare studies have been reported regarding the reduction of artifacts from the extracted objects of tissue section images. The research scientists used different methods for reduction of the tiny points prior to segmentation process. Most general method is median filtering (Gonzalez and Woods, 2008) in which pixel intensity is used to reduce the noise. Mohapatra et al. (2011) performed median filtering followed by unsharp masking to all the images before cell segmentation. Mathematical morphology (Gonzalez and Woods, 2008) is another method which was employed by researchers to remove small noisy areas from the images. Morphological operators work on the basis of shape characteristics of the objects. Erosion, dilation, opening and closing operations are the basic operations of mathematical morphology. Phukpattaranont and Boonyaphiphat (2006) used morphological erosion and opening operations to eliminate spike noise and to simplify shape of cells. But these methods may change the shape and size of actual leukocytes. Further, the artifacts whose shape and size may be similar to leukocytes cannot be removed using these morphological operations.

* Corresponding author. Tel.: +91 8989474947.
E-mail addresses: saraswatmukesh@gmail.com, saraswatmukesh@iiitm.ac.in (M. Saraswat), kvarya@gmail.com, kvarya@iiitm.ac.in (K.V. Arya).

0952-1976/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.engappai.2013.09.010

Recently, we have introduced a novel unsupervised method (Saraswat et al., 2013) based on differential evolution (DE) (Storn and Price, 1997) to segment the leukocytes from the images of mice skin sections stained with hematoxylin and eosin (H&E) staining and acquired at 40× magnifications. The method performs an unsupervised multilevel clustering in two phases. First phase uses pixel intensity to extract the leukocyte type objects from the complex background of skin tissue image. A vast number of artifacts are also segmented along with the leukocytes in this phase due to similarity of their intensity with leukocytes. These artifacts can only be differentiated from leukocytes on the basis of their morphological structure. Therefore, second phase is introduced which performs unsupervised multilevel clustering using the feature vector of each extracted object to reduce the artifacts. But, this process can be improved by introducing the supervised segmentation. Supervised artifacts reduction method may provide better results as they are based on learning and classification of objects.

Thus, keeping in view the need of drug research scientists and availability of limited number of algorithms for reducing the artifacts in leukocytes segmentation from tissue section images, this study was planned to design a supervised and efficient system for reduction of artifacts/noise in leukocytes segmentation. The proposed system modifies the second phase of our earlier unsupervised method (Saraswat et al., 2013) through threshold based supervised binary classifier to classify the extracted objects into noise and leukocytes. The classifier compares the feature vector of an object with the optimal threshold values for its classification. To find the optimal threshold values for each feature, non-dominated sorting genetic algorithm II (NSGA-II) (Deb et al., 2002) is used. It is one of the most efficient multi-objective evolutionary algorithms derived from its predecessor NSGA (Srinivas and Deb, 1994). NSGA-II is used extensively in many applications due to its fast non-dominated sorting procedure, an elitist strategy, and parameter less approach. Coello Coello (2006) presented state-of-the-art of NSGA-II and other multi-objective algorithms along with its applicability to a larger set of applications. Salazar et al. (2006) have applied and compared the efficiency of NSGA-II with existing methods for reliability optimization problems. NSGA-II has also been successfully applied to the applications of pattern recognition (Tekguc et al., 2009).

Therefore, in this paper NSGA-II (Deb et al., 2002) has been used to calculate the optimum thresholds. Since, NSGA-II is a multi-objective technique, two objective functions are generated using receiver operating characteristic (ROC) curve (Fawcett, 2003; Qin, 2005). ROC curves are utilized to evaluate the classifier performance in case of skewed class distribution and uneven classification error. The performance of the proposed method is compared with Saraswat et al. (2013) and other supervised classification methods (support vector machine (SVM) Vapnik, 1998, k-nearest neighbor (kNN) Duda and Hart, 1973, artificial neural network (ANN) Hagan et al., 1996, linear discriminant analysis (LDA) McLachlan, 2004).

Rest of the paper is organized as follows. The problem formulations and its description are described in Section 2. Section 3 presents the proposed segmentation method. The considered performance measures for this study are explained in Section 4. In Section 5, experiments are performed and the results are compared. Finally, paper is concluded in Section 6.

2. Problem description

The supervised segmentation is a process of labeling the objects to their respective class based on their properties like color, area, shape, etc. This paper deals with the problem of reducing the artifacts/noise in an automated leukocytes segmentation process using a threshold based supervised binary classifier. The proposed method calculates the optimal thresholds, to be used by binary classifier for the identification of noise and leukocytes, using NSGA-II and ROC curve.

The problem of finding the optimal thresholds can be treated as an optimization problem defined by an appropriate objective function. Let there be N number of objects which are to be classified into their respective class. Each object is represented by a features vector $\mathbf{f}$, having M number of features $[x_1, x_2, \ldots x_M]$. The classifier separates the noisy elements from required leukocytes using the feature vector $\mathbf{f}$. For each feature value, two optimum thresholds are calculated which represent the boundary wall between the two classes. The binary classifier compares a feature value with the optimum threshold values to take the decision. Therefore, total $2M$ number of optimum thresholds is to be calculated for comparison in classifier. Various combinations of the threshold values can generate a set of optimal solutions, known as pareto-optimal solutions instead of a single optimal solution. For obtaining pareto-optimal solutions, a number of multi-objective evolutionary algorithms have been proposed (Coello Coello, 2006). In this paper, non-dominated sorting genetic algorithm II (NSGA-II) has been used for obtaining the pareto-optimal solutions where $2M$ will be the number of decision variables in a chromosome. The objective function values for each of the chromosome are found using ROC curve which is described in the following section.

2.1. Objective function

To find the objective functions for NSGA-II, the concept of ROC curve is implemented. An ROC curve is a plot of sensitivity or true positive rate (TPR) versus false positive rate (FPR). The value of TPR and FPR can be calculated using Eqs. (1) and (2) respectively (Qin, 2005).

$$TPR = \frac{TP}{TP+FN} \tag{1}$$

$$FPR = \frac{FP}{FP+TN} \tag{2}$$

where TP, FN, FP, and TN are true positive, false negative, false positive, and true negative value of a classifier respectively. TP represents the number of correctly classified leukocytes, FN is the number of objects which are classified as noise but actually are leukocytes, FP represents the number of objects classified as leukocytes but actually are noise and TN is the number of correctly classified noisy elements. FPR can be represented in terms of specificity as shown in Eqs. (3) and (4) (Qin, 2005).

$$FPR = 1 - specificity \tag{3}$$

$$specificity = \frac{TN}{FP+TN} \tag{4}$$

A numeric value on each of the parameters, TPR and FPR, represents the degree to which a region belongs to noise or cell. An ROC curve gives a comparative view of classifier for all possible decision thresholds. Optimum classifier shows higher value of TPR and lower value FPR. On this basis, two objective functions, TPR and FPR, are proposed for selecting the thresholds for these features. One will calculate the TPR which is to be maximized and other will calculate the FPR which is to be minimized. To calculate the TPR and FPR for the selected decision variables, training data set (gold standard) is used which is the set of labeled objects along with their feature values. The process of calculating objective function values are depicted in Algorithm 1.