Optics and Laser Technology 112 (2019) 530–538

Contents lists available at ScienceDirect

Optics and Laser Technology

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

Full length article

Leukocyte classification based on spatial and spectral features of microscopic hyperspectral images

Yifan Duan[a], Jiansheng Wang[a], Menghan Hu[a], Mei Zhou[a], Qingli Li[a,b,*], Li Sun[a], Song Qiu[a,b], Yiting Wang[a]

[a] Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200241, China
[b] Engineering Center of SHMEC for Space Information and GNSS, Shanghai 200241, China

HIGHLIGHTS

- A microscopic hyperspectral imaging system is used to identify blood cells.
- SMACC and ISODATA algorithms are combined to segment leukocytes.
- Both spectral and spatial information are used for leukocytes classification.

ARTICLE INFO

Keywords:
Microscopic hyperspectral imaging
Spectral-spatial feature
Image segmentation
Leukocytes classification

ABSTRACT

Observing and identifying blood cells is a direct way for early diagnosis of blood diseases. Traditional blood cell recognition methods are usually time-consuming and laborious tasks for medical staff. This paper proposed an efficient leukocyte recognition method based on microscopic hyperspectral imaging technology. In order to achieve better segmentation performance and further improve the representativeness of features, the sequential maximum angle convex cone algorithm and iterative self-organizing data analysis technique algorithm are combined to segment the leukocytes from microscopic hyperspectral images. In addition, the uniform and rotation invariant local binary pattern is adopted as a textural measurement of the leukocytes. Combined the texture features with shape and spectral features, support vector machine is used to classify the leukocytes into different types. Experimental results show that the proposed method provides higher segmentation and recognition accuracy compared with the existing method. Moreover, the addition of spectral features improves the recognition performance shows the potential diagnosis capacity of microscopic hyperspectral imaging technology.

1. Introduction

The immune system is the third line of defense of human beings while leukocytes play an import role in the immune system. In recent years, the incidence of many blood diseases such as leukemia, anemia and some malignant tumors is increasing [1]. The leukocyte is responsible for identifying and engulfing non-normal cells, serving as the destroyer of virus and bacteria. Usually, the number of leukocytes which is one of the main indicators in blood routine examination maintains at a certain content, thus it is often used as an important basis for the diagnosis of systemic diseases. Leukocytes are diverse in morphology and size which includes lymphocyte, neutrophil, basophil, eosinophil and monocyte in general [2]. They have their own physiological functions but also cooperate with each other to protect the human body from the invasion of the virus or bacteria. Usually, blood smears are stained before observing [3]. It's more reliable and time-saving to identify leukocytes automatically, but for most of the time, they are artificially recognized by the doctor or blood specialist through a microscope [4,5]. Besides the heavy workload and low efficiency, blood routine examination was easily affected by human factors in the past.

Nowadays, images processing of many different fields are increasingly relying on computer-aided algorithms, like image defect detection and content-based image retrieval (CBIR) [6,7 8]. Zhang, LN et al. significantly improved the performance of CBIR systems by alleviating the drawbacks of directly using SVM as a relevance feedback schemes [9]. Giacinto et al. proposed an approach to the automatic design of neural network ensembles which is tested in multisensor images [10].

* Corresponding author at: Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200241, China.
E-mail address: qlli@cs.ecnu.edu.cn (Q. Li).

https://doi.org/10.1016/j.optlastec.2018.11.057
Received 15 June 2018; Received in revised form 25 November 2018; Accepted 27 November 2018
Available online 07 December 2018
0030-3992/

With the rapid development of computer science and artificial intelligence, some computer-assisted methods have been widely used in the field of medical microscopic image processing. The analysis of blood cells based on RGB images mainly focuses on two-dimensional image processing methods such as threshold segmentation or morphological operations. Salim Arslan et al. split the image into RGB Channels and found that green band of the region of leukocytes have darker intensities compared with the other cellular regions [11]. The color transformation based on the green band is adopted to do further segmentation. Der-Chen Huang et al. proposed an unsupervised classification method to recognize leukocytes by extracting the features of leukocyte nucleus [12]. A better segmentation result is obtained by enhancing the leukocyte nucleus and suppressing other regions, which guarantees a more convincing result of shape feature extraction. Othman, Mazin et al. applied a feed forward back propagation neural network to classify white blood cells (WBC) with 16 spatial features [13]. Ghosh et al. proposed a hybrid microscopic image analyzing technique to got a classification results of WBC, where the color, size, shape and texture features are considered as the significant features [14]. More recently, Shen et al. reported the latest advances in machine learning, especially with regard to deep learning in the field of medical image analysis [15]. These studies have obtained accurate results to prove the effectiveness and robustness of the methods and provided a new idea for the diagnosis of hematological diseases, but obviously, these existing methods can only analyze blood cells on the basis of spatial information. Because of the diversity of WBC, it is difficult to get accurate segmentation results of RGB images. Moreover, WBC classification of RGB images have also shown the over-reliance on the segmentation performance. If the image segmentation is unreasonable, then the spatial feature extraction can't be achieved as well. Therefore, some new technologies are called for to solve this problem. In this way, microscopic hyperspectral imaging technology can provide a new solution.

Hyperspectral imaging is an emerging technology that combines the merits of spectral imaging with traditional imaging techniques. This technology evolved from multispectral remote sensing imaging since the early 1980s which could be described as a revolutionary leap. Initially, hyperspectral imaging was used for geological exploration [16], followed by agriculture [17], food safety [18], biology [19], and medicine [20], where dramatic progress has been made. Each single-band image generated through the hyperspectral imaging system corresponds to a specific narrow-band wavelength from which we can simultaneously obtain both spatial and spectral information. As a result, the hyperspectral image possesses richer information than gray or color image through spectral visualization. Ratle et al. proposed a general framework for semisupervised image classification on the basis of neural networks. The results showed the improved accuracy and scalability among several problems of hyperspectral image classification [21]. In recent years, the analysis of blood cells combining with hyperspectral imaging technology has become a popular trend, especially for leukocytes. Li W. et al. provided a promising classification system that can recognize hyperspectral image of blood cells using parallel computation which offered fast speed [22]. Verebes et al. probed changes in blood cell components through a novel methodology based on hyperspectral imagery with enhanced Darkfield microscopy and found it is possible to differentiate active from inactive WBC by spectra [23]. Neugebauer et al. proposed that Raman spectroscopic characterization can be found in peripheral blood [24]. Robison et al. indicated that features of white blood cells are most prominent in the 428–442 nm band over a varying range of illumination intensities in Snapshot Hyper-Spectral imaging systems [25]. Liu et al. proposed a full-automatic red blood cell counting method with high precision, but there is no verification of blood smears with WBC and this method is inapplicable in the serious cell overlapping problem [26]. Wang et al. applied mathematical morphology-based methods to extract spatial features and supervised method to classify WBCs into five types, while the spectral range of the hyperspectral imaging system failed to cover the whole spectral signatures of each kind of WBC [27]. These studies illustrated the combination of spectral and spatial information showed great prospect. However, few of the recent studies have focused on the combination of spatial and spectral features for leukocyte classification.

In this paper, an approach combined spectral and spatial information is introduced to classify leukocytes automatically. The leukocyte images are captured from an AOTF-based hyperspectral imaging system which contains spectral information and the application of spectral features could reduce the impact of segmentation on the feature extraction. The sequential maximum angle convex cone (SMACC) algorithm and iterative self-organizing data analysis technique algorithm (ISODATA) are applied to segmente the leukocyte. Then, the shape and texture features of single-band images are extracted based on the characteristic of different types of leukocytes. We use Support Vector Machines (SVM) in the cell classification step to divide the leukocytes into five different types. Some indicators are mainly introduced to measure the performance of image segmentation and classification. The experimental results indicate that using combined spectral features to spatial features can improve the efficiency and accuracy of leukocyte classification compared with traditional methods.

2. Methods

2.1. Hyperspectral image preprocessing

During the process of image capturing, some unwanted factors such as noise probably are introduced into the image. Due to the uneven distribution of the light source, the difference of blood smears, and other uncertain factors, the quality of the hyperspectral images are declined. In order to reduce the impact of these factors on subsequent processing, it is necessary to calibrate the original images. Let $I_B(x,y,\lambda)$ be the pixel intensity in the blank area of a blood smear and $I_s(x,y,\lambda)$ be the original pixel intensity of sample image, where x, y, λ is the three-dimensional coordinates of this pixel, respectively. Then each pixel value in the image is calibrated by the following formula:

$$I'_S(x, y, \lambda) = \frac{\sum_{x=1}^{N}\sum_{y=1}^{M}\sum_{\lambda=1}^{P} I_B(x, y, \lambda)}{N * M * P} * \frac{I_S(x, y, \lambda)}{I_B(x, y, \lambda)} \quad (1)$$

where $I'_s(x,y,\lambda)$ is pixel intensity at (x,y,λ) of the processed image; N and M represent the spatial size of a single band hyperspectral image, P is the total number of bands. After calibration, the images become clearer to recognize and the stripe noise is restrained.

2.2. Segmentation method

To pinpoint leukocyte, we first need to separate the leukocyte from the background. Some misclassification occurred due to the similarities between spectra of two different regions. In order to get more useful information to extract features, the distinctive components like the nucleus and the cytoplasm were precisely segmented for further classification.

A. Cell segmentation

The sequential maximum angle convex cone (SMACC) algorithm is used to determine the whole leukocyte [28]. SMACC is a method based on convex cone model for acquiring endmembers from original hyperspectral images and providing abundance images automatically. This model uses the pole to determine the convex cone and define the first endmember. An oblique projection with constraint condition is applied in the existing pyramid to generate the next endmember, continue the process to generate new endmembers [29]. In a mathematical sense, SMACC uses the following formula for the expansion of endmembers:

$$I(j, i) = \sum_{m}^{X} E(j, m)A(m, n) \tag{2}$$

where *A(m,n)* is the abundance image and *E(j,m)* is the spectra of the endmember. *X* indicates the maximum endmember number.

After binarizing the leukocyte abundance image, we found the largest connected region, through which some pixels that don't belong to the region of the cell could be removed. Considering that the segmented leukocyte might not be perfectly connected and some break exists along the border, morphology operation is applied to solve the problem.

B. Nucleus and cytoplasm segmentation

Usually, the characteristic of the nucleus is highlighted, so the leukocyte is further separated into the nucleus and the cytoplasm. The spectra of leukocyte nucleus are characteristic which presents more distinguishable divisibility than the rest parts of the leukocyte. After getting the spectrum of each pixel, Iterative ISODATA is applied to cluster the pixels with similar spectra because it makes the classification process more automated and faster [30]. Since the number of classes in a blood cell image is uncertain, unsupervised classification method with less human intervention is more effective. Then the clustering results are marked with different colors. Moreover, there may be some misclassifications of the nucleus region, then use morphological operation to promote the integrity of the nucleus region. Combining the results with the segmented leukocyte region can obtain a complete nucleus region and cytoplasm region.

The whole flow chart of the segmentation process is shown in Fig. 1. The SMACC algorithm was first applied to extract the region of leukocyte and the largest connected area was introduced to eliminate error segmentation. A further segmentation using ISODATA clustering algorithm is necessary to obtain the nucleus region of the leukocyte. Finally, the cytoplasm region can be got by combining the two segmentation results.

2.3. Leukocyte feature extraction

A. Texture feature

There are five common types of leukocyte, some of which contain granular and others are relatively smooth. According to this, leukocytes can be divided into granulocytes and agranulocyte, which can be reflected through the texture. The leukocyte all have their features in nucleus and cytoplasm texture which means texture features can be extracted to characterize different cells for leukocyte recognition. Local Binary Pattern (LBP) is an operator used to describe the local features of an image. It was first proposed by Ojala et al. [31,32] and it has the advantages of grayscale invariance and rotation invariance. A formal description of LBP operator used in an image was given by:

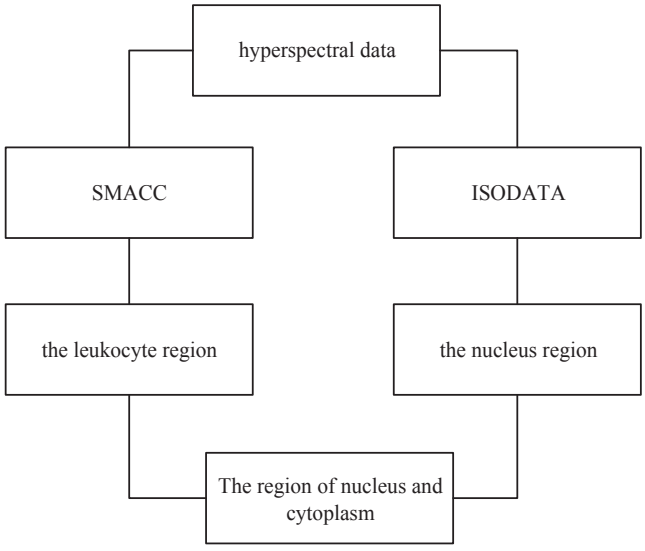

Fig. 1. Flow chart of segmentation process.

$$f(x, y) = \sum_{p=0}^{P-1} 2^p \operatorname{sgn}(I_p - I_c) \tag{3}$$

where I_c is the intensity of a central pixel; $I_p(p = 0, 1, 2,\ldots,P-1)$ is the intensity of P uniformly distributed pixels in the neighborhood of the central pixel respectively; *f(x,y)* is the encoding value of the central pixel. The LBP operator is used to characterize the texture of the leukocyte image by calculating the histogram of the LBP feature map which used as an eigenvector for classification and identification. However, a large number of eigenvalues will make the amount of data too large and the histogram too sparse. For example when $P = 8$, applying this operator will get 256 eigenvalues. Therefore, the original LBP pattern needs to be dimensionally reduced so that the image information can be expressed more precisely. Considering that the cell image itself does not have a specific directionality, we adopt an improved operator called uniform & rotation invariant LBP.

In image processing, the vast majority of LBP patterns contain up to two transitions from 1 to 0 or from 0 to 1. So when a binary number of a corresponding LBP pattern that transforms less than twice, it is considered the corresponding binary number is called a Uniform Pattern. For example, the binary number 00000000 contains no transform, and 1001111 has twice, and both of them are called a type of uniform pattern. Maenpaa also proposed an LBP operator with rotation invariance which indicates that rotating a corresponding binary number and finds the minimum in all possible cases [33]. When the minimums of two binary numbers are equal, it is considered they are the same pattern. Then, the encoding value $f_p^{ri}(x,y)$ of the central pixel can be expressed as:

$$f_p^{ri}(x, y) = min\{ROR(f_p^{ri}(x, y), i)|p = 1, 2, \cdots, P-1\} \tag{4}$$

where ROR*(x,y)* denotes a function of moving x in a circular manner for i bits and the binary number circulate for P times. Combining the uniform pattern and the rotation invariant LBP operator, the following formula is considered:

$$f_p^{riu}(x, y) = \begin{cases} \sum_{p=0}^{P-1} \operatorname{sgn}(I_p - I_c) & \mathrm{T} \leqslant 2 \\ P + 1 & \mathrm{T} > 2 \end{cases} \tag{5}$$

where T denotes the transform times from 1 to 0 or from 0 to 1 in a binary number.

After the improvement in original LBP pattern, the later proposed uniform & rotation invariant LBP significantly reduce the dimension of features matrix to $P + 1$. Since we need the texture feature of the leukocyte, the uniform & rotation invariant LBP operator was applied to the segmented leukocyte. Considering $P = 8$ for each central pixel, there are a total of 9 patterns and the rest of the binary numbers are considered to be one another pattern. So totally we extracted 10 texture features for each hyperspectral image for further recognition.

B. Shape feature

Five types of leukocytes vary greatly in the shape of the cells and the size of the nucleus by observing the single band images. Some meaningful features were picked out to represent peculiarities of an image. Since the nucleus size differs a lot, while the overall sizes of the cells are similar, the proportion of the nucleus size to the overall cell provides an important basis for the classification. The nuclear ratio was calculated as:

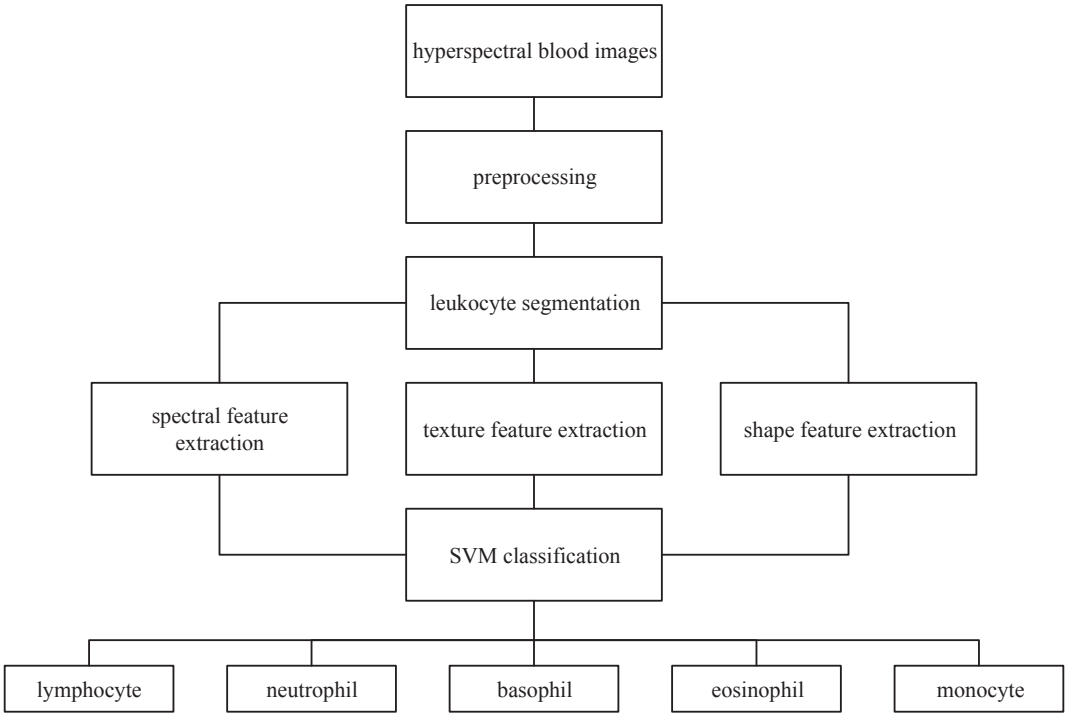

Fig. 2. The flow chart of spectral and spatial classification method.

$$R = \frac{S_{\mathrm{nu}}}{S_{\mathrm{ce}}} \tag{6}$$

where S_{nu} and S_{ce} are the nuclear and the cellular area represented by the total number of pixels.

We can see from the leukocyte images that some nuclear shape is similar to a circle, while some others can be multiple lobes. The variance is widely used to measure the deviation between the average distance and the distance from the border to the centroid, especially when the sample deviates to a great extent. The greater the variance, the greater the data fluctuates, and the nuclear shape is less similar to a circle. The nuclear variance can be described as:

$$V_{nu} = \frac{1}{(N-1)} \sum_{i} \frac{(d_i - d_{arg})^2}{d_{arg}^2} \tag{7}$$

where d_i is the distance from each boundary point to the centroid of the nucleus; d_{arg} is the average of all these distance; N is the total number of the nuclear boundary pixels.

A ratio to measure the shape of the nucleus based on the minimal enclosing convex polygon was employed, which is calculated as:

$$R_{sunk} = \frac{S_{nu}}{S_{cp}} * \frac{d_{nu\,\min}}{d_{cp\,\min}} \tag{8}$$

R_{sunk} is used to show the extent of nuclear shape irregularity quantitatively and the smaller the R_{sunk}, the greater the degree of sunken of the whole nucleus. $d_{nu\min}$ and $d_{cp\min}$ are the shortest distances from the nucleus boundary and the minimal enclosing convex polygon to the centroid of the nucleus respectively.

Totally we selected 3 shape features based on the size, minimal enclosing convex polygon and some other parameters. We combined them with 10 texture features to obtain more abundant information. In addition to the shape and texture features, the spectral information is also taken into account to classify leukocytes.

During the segmentation of leukocyte, we have already used the spectra to separate the nucleus and the cytoplasm. In this phase, each hyperspectral image is taken as a whole and its average spectra are considered as the feature. The precise region of nucleus and cytoplasm was obtained in advance for each image during the image segmentation. Once the nucleus and cytoplasm were selected, the average spectra for these two regions can be calculated respectively. We take the pixel value of each band as a feature, so totally we have obtained 120 spectral features for each image.

2.4. Leukocyte classification

Support vector machine (SVM) is a dichotomous model whose purpose is to construct a linear classifier to find the maximum interval in feature space. The SVM classifier is to find out an optimal hyperplane according to the training samples so as to make the distance between the hyperplane and each side of the samples as far as possible [34]. SVM efficiently deals with large-scale data, especially those with high dimensionality. Therefore, an SVM classifier is chosen to handle the problem of leukocyte classification. The objective function of SVM can be expressed as:

$$\begin{aligned} &\min_{w,b,\xi} \quad \tfrac{1}{2}\,\|\omega\|^2 \\ &s.\,t. \quad y^{(i)}(\omega^T x^{(i)} + b) \geqslant 1 \end{aligned} \tag{9}$$

where b is the bias, ω is the weight vector and $y^{(i)}$ is the label for sample i. Then the objective function can be explained with the Lagrange function:

$$\begin{aligned} L(\omega, b, \alpha) &= \tfrac{1}{2}\,\|\omega\|^2 + \sum_{i=1}^{m} \alpha^{(i)}\left[y^{(i)}(\omega^T x^{(i)} + b) - 1\right] \\ &= \tfrac{1}{2}\,\|\omega\|^2 + \sum_{i=1}^{m} \alpha^{(i)} - \sum_{i=1}^{m} \alpha^{(i)} y^{(i)}(\omega^T x^{(i)} + b) \end{aligned} \tag{10}$$

where α are the Lagrange multipliers for constraints that should be positive.

The selection of the kernel function is crucial as it determines the classification performance to some extent. In this experiment, for the integrated features are linear non-separable, the quadratic kernel is considered to classify the images. The kernel function we used can be described as:

$$k(x_i, x_j) = (x_i^T * x_j + c)^2 \tag{11}$$

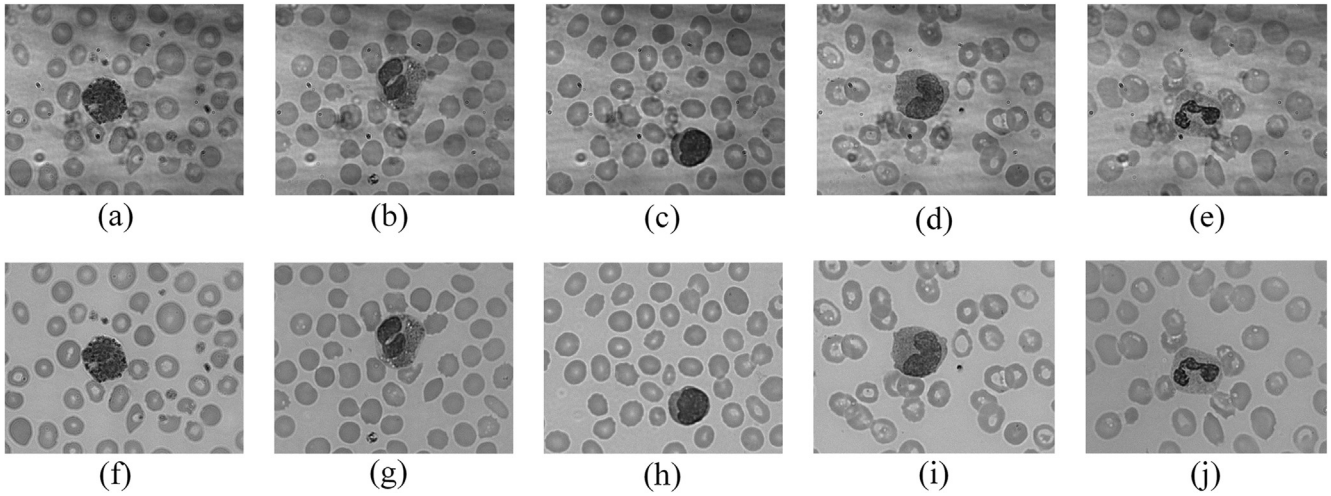

Fig. 3. (a–e) The original 30th band images of basophil, eosinophil, lymphocyte, monocyte, neutrophil (f–j) The preprocessed 30th band images of basophil, eosinophil, lymphocyte, monocyte, neutrophil.

where x_i and x_j are two samples in the train data and c is a constant.

Fig. 2 shows the overall flow chart of the proposed method. On one hand, the accurate segmentation of leukocyte guaranteed the effectiveness and feasibility of texture features and shape features extraction. On the other hand, the combination of spectral-spatial features fully described the uniqueness of the hyperspectral image.

3. Experimental results and discussions

3.1. Preprocessing results

After capturing the microscopic hyperspectral images of blood cells, the data can be preprocessed by the proposed method. Fig. 3 shows the original and the preprocessed 30th band images of five types of leukocytes extracted from the hyperspectral data. It can be seen that the background of the original data is uneven with some visible stripe noise and light spots, which will affect the recognition of leukocytes. After the preprocessing, the background of the image becomes uniform and the light spots are eliminated.

3.2. Segmentation results

The abundance images of three hyperspectral images obtained by SMACC are shown in Fig. 4. For the different types of blood cells, three abundance images represent three components of the blood cell images. It is obvious in Fig. 4 that the pure abundance image of leukocyte is well separated from the original data by SMACC. The second column in Fig. 4 represents the separated abundance images of leukocytes where leukocytes are well located.

We compared the proposed method with Neural Network (NN) which is a state of the art method for image segmentation and classification [13]. In the dataset, the maximum and minimum radiuses of the leukocyte are about 160 pixels and 110 pixels, respectively. The first row in Fig. 5 shows the RGB color image of five types of leukocytes, the second row shows the 30th band images of hyperspectral data, the third row shows the segmented results with the proposed method and the last row is the corresponding leukocytes segmented with NN. As it can be seen from Fig. 5, segmentation performance in our method shows a general improvement over the NN. At some regions such as the junctions of the nucleus and the cytoplasm, the segmented results of NN are not as accurate as the proposed method.

In order to quantitatively evaluate the accuracy of segmentation, some indicators are considered to measure the performance of segmentation. Due to the irregular shape of the leukocyte nucleus and the difficulty of manual labeling, the segmentation performances of the whole leukocyte are evaluated. It is unreliable to observe the segmentation results subjectively by the naked eye, thus the images segmented automatically are compared with those images segmented manually which we called GT images. Then, segmentation accuracy (SA), false positive ratio (FPR) and false negative ratio (FNR) are used to evaluate the segmentation results. Let A_{as} be the total number of pixels in the region we segmented (RS) and A_{gt} be the total number of pixels in the GT. A_{ol} represents the number of pixels in the overlapping area of RS and GT. SA indicates the percentage of correctly segmented regions to the GT image. FPR measures the proportion of extra pixels which indicates the regions included in segmentation results which should not belong to the GT, while FNR represents measures the proportion of missing pixels which indicates the regions that should have included in the GT. Then the indicators are as follows:

$$\mathrm{SA} = \left(1 - \frac{|A_{gt} - A_{as}|}{A_{gt}}\right) \times 100\% \tag{12}$$

$$\mathrm{FPR} = \frac{A_{as} - A_{ol}}{A_{gt} + A_{as} - A_{ol}} \tag{13}$$

$$\mathrm{FNR} = \frac{A_{gt} - A_{ol}}{A_{gt} + A_{as} - A_{ol}} \tag{14}$$

$$A_{ol} = RS \cap GT \tag{15}$$

Table 1 shows the segmentation results of the whole leukocyte measured with the indicators above. Comparing the method in this paper with NN, the table list three indicators calculated with five types of leukocytes. In summary, SMACC combined with ISODATA provides better performance than NN in five types of leukocytes. The average SA improved by 4.8% in the proposed method. This can be considered that the method in this paper equals to a double verification for segmentation. When using the spectra only for once, there is a high probability that applying NN for a hyperspectral image will have errors. Moreover, if using the spectral information straightly without determining the approximate location of leukocyte first, it is more likely that some other regions are misclassified as leukocytes. The indicator of FNR in the proposed method is kept below 10%, while in NN, some may be greater than 10%. On the other hand, the indicators of FPR and FNR calculated with NN in neutrophil perform as good as that with our proposed method. It can be explained that NN could still maintain the competitive edge in segmenting the images with significant differences in spectra. However, in other four types of leukocytes, the proposed method performs a significantly lower FNR and higher SA under the

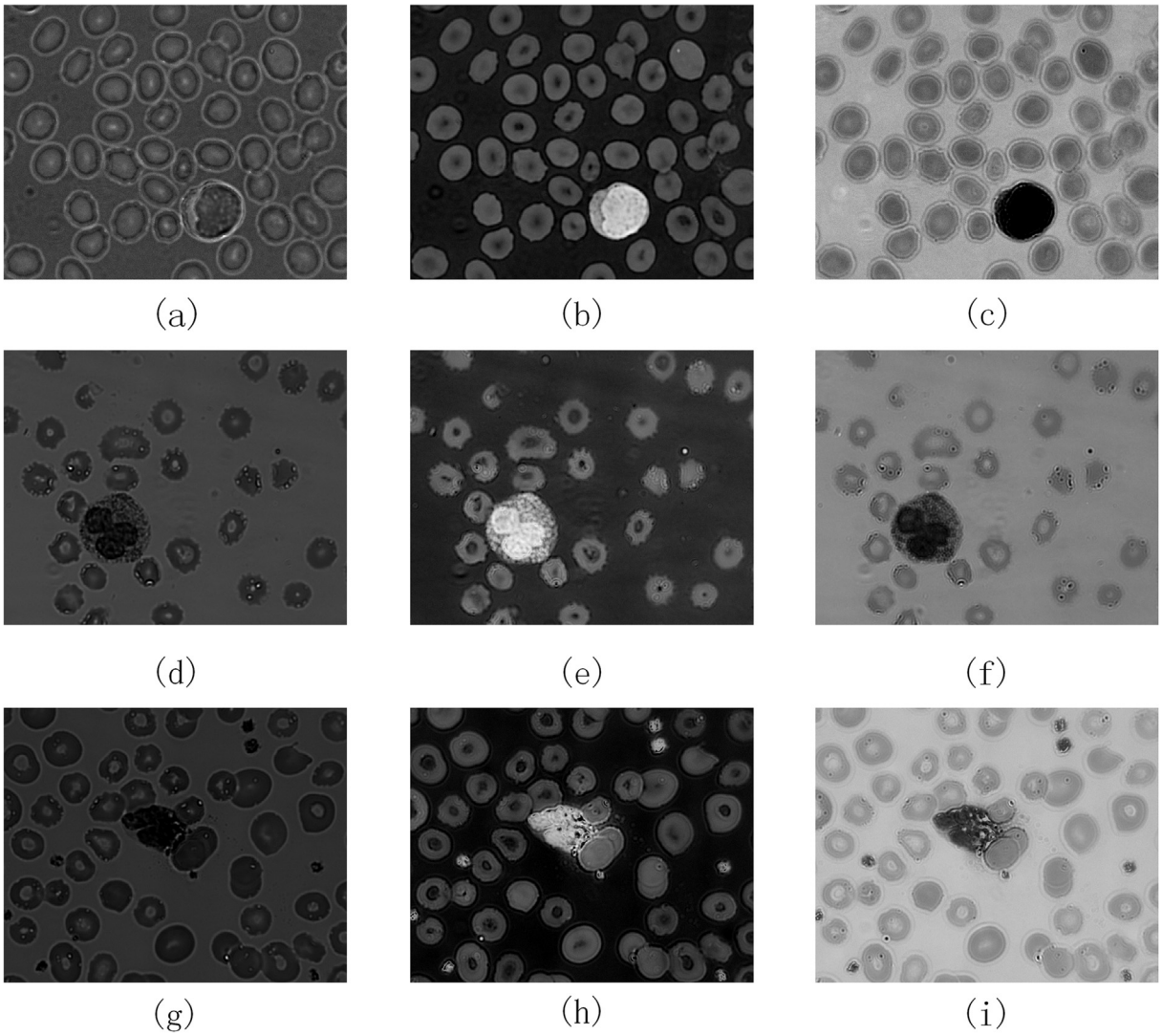

Fig. 4. The abundance images of three pure components of three hyperspectral blood cell images. (a), (d), (g) The abundance images of others endmember (b), (e), (h) The abundance images of leukocyte endmember (c), (f), (i) The abundance images of background endmember.

condition of similar FPR.

Table 2 further compares the segmentation results of the two methods with the nucleus region. In general, the proposed method keeps the SA larger than 95%. One reason can be that the nucleus is one of the most distinguishable parts of the leukocyte. The calculated values show no significant differences in lymphocyte and monocyte between the two compared methods. One thing worth noting is that the results of FNR in the neutrophil and basophil are larger than 10% with NN, while these errors are lower than 5% with ISODATA. It shows that our method has done more accurate segmentation on these two types of leukocytes. Therefore, it can be concluded that with the applying of SMACC-ISODATA, lower FNR and FPR with higher SA mean less misclassified pixels, which is a better result. That is, the combination of SMACC and ISODATA can reduce the error probability and shows its effectiveness.

3.3. Classification results

In this experiment, shape, texture and spectral features are selected to represent the hyperspectral leukocyte images. For convenience, we call the shape and texture features as spatial features. In order to verify the validity of the spectral features, we chose the combination of spatial-spectral features, and spatial features only to use as the input of SVM. Table 3 illustrated the classification accuracy of hyperspectral images with spatial-spectral features and spatial features only. From the table, we can see that overall accuracy has a great improvement in lymphocyte, eosinophil and monocyte. The texture and shape features can be used to identify a certain type of leukocyte while adding the spectral features will make the classification results better.

Table 4 shows the impact of different SVM kernel functions on classification results. The quadratic SVM performs best among five types of leukocytes, which was chosen as the kernel function in this paper. From this we can see that using the same features as the input of SVM with different kernel functions may obtain completely different results.

The final classification accuracy results are shown in Table 5. Taking into account the number of sample images, we used 5-fold cross-validation for classification. All the features are considered as the input of the classifier. The accuracy in the table is the average value of multiple calculations. We used a type of k-Nearest Neighbor (KNN) classifier that makes finely-detailed distinctions between classes [35], with the number of neighbors set to 1 to compare with SVM. Linear Discriminate Analysis (LDA) is also used to classify the leukocytes [36]. The results show that with cross-validation, SVM shows the highest accuracy, followed by KNN. It can be concluded that in this experiment, the hyperspectral image with SVM is superior to some other methods.

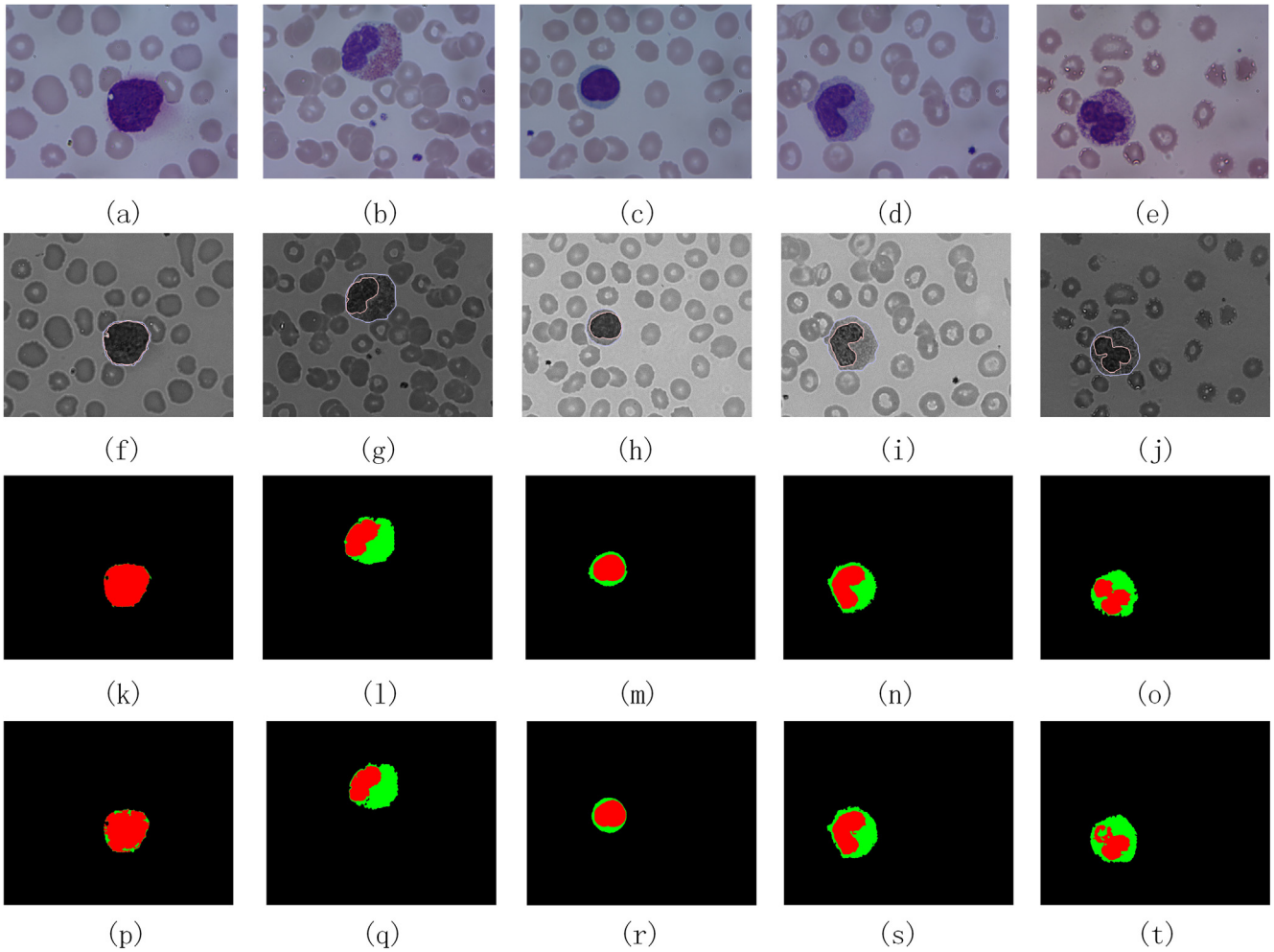

Fig. 5. (a–e) RGB color images of basophil, eosinophil, lymphocyte, monocyte, neutrophil (f–j) Ground Truth (GT) images of basophil, eosinophil, lymphocyte, monocyte, neutrophil (k–o) Segmented results of basophil, eosinophil, lymphocyte, monocyte, neutrophil with proposed method (p–t) Segmented results of basophil, eosinophil, lymphocyte, monocyte, neutrophil with NN.

Table 1
The segmentation results of the whole leukocyte.

Ratio (%)	Proposed method			NN		
	SA	FPR	FNR	SA	FPR	FNR
Lymphocyte	95.19	3.12	7.78	90.96	0.18	9.20
Neutrophil	95.67	0.70	5.00	94.05	0.83	6.72
Basophil	93.56	0.27	6.70	90.01	0	9.99
Eosinophil	94.54	0.56	5.99	88.17	0.17	11.97
Monocyte	94.64	1.13	6.42	86.39	0	13.62

Table 2
The segmentation results of the nucleus.

Ratio (%)	Proposed method			NN		
	SA	FPR	FNR	SA	FPR	FNR
Lymphocyte	99.80	2.31	2.50	99.58	2.22	2.63
Neutrophil	98.46	3.29	4.78	87.13	2.79	15.30
Basophil	99.73	2.31	2.05	90.14	0.67	10.46
Eosinophil	96.77	5.22	2.16	91.06	0.53	9.43
Monocyte	96.77	1.28	4.47	97.83	2.16	4.29

Table 3
The comparison of the classification results with spatial-spectral features and spatial features only.

	Lymphocyte	Neutrophil	Basophil	Eosinophil	Monocyte
Spatial-spectral features	28/28	18/19	20/21	23/23	29/29
Spatial features only	24/28	18/19	19/21	17/23	24/29

Table 4
The classification results using different kernel function of SVM.

	Lymphocyte	Neutrophil	Basophil	Eosinophil	Monocyte
Linear SVM	27/28	17/19	20/21	22/23	28/29
Quadratic SVM	28/28	18/19	20/21	23/23	29/29
Fine Gaussian SVM	21/28	3/19	0/21	4/23	28/29
Coarse Gaussian SVM	28/28	9/19	11/21	22/23	23/29

The experimental results have shown the effectiveness of the proposed method compared with traditional methods. However, there are still some limits on the experiment of leukocyte classification. The current dataset does not involve the circumstance that more than one leukocyte coexists in one image. In our follow-up studies on cell

Table 5
The comparison of the classification accuracy with SVM and other methods.

	SVM	KNN	LDA
Accuracy (%)	98.3	94.2	84.2

classification, further research will be done under the presence of overlapped or adjacent cells in the hyperspectral images. For the leukocytes adjacent to each other, it is feasible to firstly segment cells based on some morphological contour features [5]. Also, it has been proved that segmentation of adjacent cells can be achieved based on the slope difference of histogram [37]. As for the overlapped areas, due to the rich spectral information of hyperspectral images, SMACC or other spectral unmixing methods can be applied to separate the mixed endmembers [28]. It may suggest that the overlapping cells can first be separated by spectral unmixing. One thing noteworthy is that the texture features are no longer available after the spectral unmixing. Once the leukocytes are separated into individuals, we can try to find more appropriate features and classifiers for subsequent cell classification. We hope to improve our approach to leukocyte classification based on microscopic hyperspectral images for more comprehensive situation as soon as possible.

4. Conclusion

The proposed leukocyte classification method in this paper can be used for the preliminary diagnosis of hematological diseases. In the past, leukocyte recognition has largely relied on the morphology of the cell or the color channel of the image while the proposed method can effectively eliminate the impact of cell segmentation on subsequent classification. In this paper, to make the classification result more reliable, we have further improved the accuracy of the segmentation with the combination of SMACC and ISODATA. The spatial features are extracted using morphological expression and the uniform & rotation invariant LBP. The addition of spectral features in hyperspectral images makes the final classification results of SVM more accurate.

The experimental results show that using the proposed segmentation method, the segmentation accuracy of the nucleus and the whole leukocyte can reach 93%, and the results of some types of leukocytes are obviously better than NN. By cross-validation, the final accuracy of classification can reach 98.3% using the combination of spectral and spatial features. At present, our proposed method has been validated only in normal mature leukocytes. And the current dataset does not involve the circumstance that more than one leukocyte coexists in one image. For diseased leukocytes or other types of cells, we also need to select more appropriate features according to the characteristics of different cells. In future work, more research will be explored under the presence of overlapped or adjacent cells in the images and we plan to increase the amount of leukocyte samples which can further improve the reliability of the experimental results, and achieve the classification of more types of cells.

Acknowledgment

The authors have no relevant financial interests or conflicts of interest to disclose. This work is supported by the National Natural Science Foundation of China (Grant No. 61377107), the Science and Technology Commission of Shanghai Municipality (Grant No. 18511102500).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.optlastec.2018.11.057.

References

[1] L.B. Dorini, R. Minetto, N.J. Leite, Semiautomatic white blood cell segmentation based on multiscale analysis, IEEE J. Biomed. Health. Inf. 17 (2013) 250–256.
[2] J. Prinyakupt, C. Pluempitiwiriyawej, Segmentation of white blood cells and comparison of cell morphology by linear and naive Bayes classifiers, Biomed. Eng. Online 14 (2015).
[3] F.L. Cao, M.M. Cai, J.J. Chu, J.W. Zhao, Z.H. Zhou, A novel segmentation algorithm for nucleus in white blood cells based on low-rank representation, Neural Comput. Appl. 28 (2017) S503–S511.
[4] M. Sajjad, S. Khan, Z. Jan, K. Muhammad, H. Moons, J.T. Kwak, S. Rho, S.W. Baik, I. Mehmood, Leukocytes classification and segmentation in microscopic blood smear: a resource-aware healthcare service in smart cities, IEEE Access 5 (2017) 3475–3489.
[5] C. Di Ruberto, A. Loddo, L. Putzu, A leucocytes count system from blood smear images segmentation and counting of white blood cells based on learning by sampling, Mach. Vis. Appl. 27 (2016) 1151–1160.
[6] X.L. Bai, Y.M. Fang, W.S. Lin, L.P. Wang, B.F. Ju, Saliency-based defect detection in industrial images by using phase spectrum, IEEE Trans. Ind. Inform. 10 (2014) 2135–2145.
[7] L.N. Zhang, L.P. Wang, W.S. Lin, S.C. Yan, Geometric optimum experimental design for collaborative image retrieval, IEEE Trans. Circuits Syst. Video Technol. 24 (2014) 346–359.
[8] L.N. Zhang, L.P. Wang, W.S. Lin, Conjunctive patches subspace learning with side information for collaborative image retrieval, IEEE Trans. Image Process. 21 (2012) 3707–3720.
[9] L.N. Zhang, L.P. Wang, W.S. Lin, Semisupervised biased maximum margin analysis for interactive image retrieval, IEEE Trans. Image Process. 21 (2012) 2294–2308.
[10] G. Giacinto, F. Roli, Design of effective neural network ensembles for image classification purposes, Image Vis. Comput. 19 (2001) 699–707.
[11] S. Arslan, E. Ozyurek, C. Gunduz-Demir, A color and shape based algorithm for segmentation of white blood cells in peripheral blood and bone marrow images, Cytom Part A 85A (2014) 480–490.
[12] D.C. Huang, K.D. Hung, Y.K. Chan, A computer assisted method for leukocyte nucleus segmentation and recognition in blood smear images, J. Syst. Softw. 85 (2012) 2104–2118.
[13] M.Z. Othman, T.S. Mohammed, A.B. Ali, Neural network classification of white blood cell using microscopic images, Int. J. Adv. Comput. Sci. Appl. 8 (2017) 99–104.
[14] P. Ghosh, D. Bhattacharjee, M. Nasipuri, Blood smear analyzer for white blood cell counting: a hybrid microscopic image analyzing technique, Appl. Soft. Comput. 46 (2016) 629–638.
[15] D.G. Shen, G.R. Wu, H.I. Suk, Deep learning in medical image analysis, in: M.L. Yarmush (Ed.), Annual Review of Biomedical Engineering, vol. 192017, pp. 221–248.
[16] P. Hartzell, C. Glennie, S. Khan, Terrestrial hyperspectral image shadow restoration through lidar fusion, Remote Sens. 9 (2017) 20.
[17] F. Vanegas, D. Bratanov, K. Powell, J. Weiss, F. Gonzalez, A Novel methodology for improving plant pest surveillance in vineyards and crops using UAV-based hyperspectral and spatial data, Sensors 18 (2018) 21.
[18] H. Lee, M.S. Kim, J.W. Qin, E. Park, Y.R. Song, C.S. Oh, B.K. Cho, Raman hyperspectral imaging for detection of watermelon seeds infected with Acidovorax citrulli, Sensors 17 (2017) 11.
[19] H. Feng, Z.L. Guo, W.N. Yang, C.L. Huang, G.X. Chen, W. Fang, X. Xiong, H.Y. Zhang, G.W. Wang, L.Z. Xiong, Q. Liu, An integrated hyperspectral imaging and genome-wide association analysis platform provides spectral and genetic insights into the natural variation in rice, Sci. Rep. 7 (2017) 10.
[20] S. Junaid, P. Tidemand-Lichtenberg, C. Pedersen, Upconversion based MIR hyperspectral imaging, in: 2017 Conference on Lasers and Electro-Optics Europe & European Quantum Electronics Conference (CLEO/Europe-EQEC), 2017, p. 1.
[21] F. Ratle, G. Camps-Valls, J. Weston, Semisupervised neural networks for efficient hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 48 (2010) 2271–2282.
[22] W. Li, L.C. Wu, X.B. Qiu, Q. Ran, X.M. Xie, Parallel computation for blood cell classification in medical hyperspectral imagery, Meas. Sci. Technol. 27 (2016) 10.
[23] G.S. Verebes, M. Melchiorre, A. Garcia-Leis, C. Ferreri, C. Marzetti, A. Torreggiani, Hyperspectral enhanced dark field microscopy for imaging blood cells, J. Biophotonics 6 (2013) 960–967.
[24] U. Neugebauer, J.H. Clement, T. Bocklitz, C. Krafft, J. Popp, Identification and differentiation of single cells from peripheral blood by Raman spectroscopic imaging, J. Biophotonics 3 (2010) 579–587.
[25] C.J. Robison, C. Kolanko, T. Bourlai, J.M. Dawson, Imaging of blood cells based on snapshot hyper-spectral imaging systems, in: Proc. SPIE (USA), 9472, 2015, 94721L, p. 94726.
[26] X. Liu, M. Zhou, S. Qiu, L. Sun, H.Y. Liu, Q.L. Li, Y.T. Wang, Adaptive and automatic red blood cell counting method based on microscopic hyperspectral imaging technology, J. Opt. 19 (2017).
[27] Q. Wang, L. Chang, M. Zhou, Q.L. Li, H.Y. Liu, F.M. Guo, A spectral and morphologic method for white blood cell classification, Opt. Laser Technol. 84 (2016) 144–148.
[28] A. Aggarwal, R.D. Garg, Systematic approach towards extracting endmember spectra from hyperspectral image using PPI and SMACC and its evaluation using spectral library, Appl. Geomat. 7 (2015) 37–48.
[29] A. Ray, R. Kopelman, B. Chon, K. Briggman, J. Hwang, Scattering based hyperspectral imaging of plasmonic nanoplate clusters towards biomedical applications,

J. Biophotonics 9 (2016) 721–729.
[30] S.A. El Rahman, Hyperspectral image classification using unsupervised algorithms, Int. J. Adv. Comput. Sci. Appl. 7 (2016) 198–205.
[31] T. Ojala, M. Pietikainen, D. Harwood, Performance evaluation of texture measures with classification based on Kullback discrimination of distributions, Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing, Proceedings of the 12th IAPR International Conference on, IEEE, 1994, pp. 582–585.
[32] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recogn. 29 (1996) 51–59.
[33] T. Mäenpää, The local binary pattern approach to texture analysis: extensions and applications, Oulun yliopisto, Oulu (2003).
[34] K. Zhan, H.B. Wang, Y.G. Xie, C.T. Zhang, Y.F. Min, Albedo recovery for hyperspectral image classification, J. Electron. Imaging 26 (2017) 12.
[35] S.X. Zhang, C.B. Yan, H. Murat, J. Yao, Y. Pazilya, Classification of normal esophagus and early esophageal disease based on SVM and KNN classifiers, Basic Clin. Pharmacol. Toxicol. 121 (2017) 36.
[36] C. Huang, W. Luo, Scene classification using class-supervised local-space-constraint latent Dirichlet allocation, Multimed. Tools Appl. 75 (2016) 10227–10240.
[37] W. Zhenzhou, A generalized method for automatic segmentation of neighboring cells, IECON 2015–41st, Ann. Conf. IEEE Ind. Electron. Soc. (2015) 004235–004240.