Blood Cells Classification Using Hyperspectral Imaging Technique

Chunni Dai[*1], Qingli Li[2], Jingao Liu[1, 2]

Department of Computer Science & Technology, Shanghai Jianqiao College, Shanghai, PR. China

Department of Communication Engineering, East China Normal University, Shanghai, PR. China

[*1]daicn@126.com; [2]qlli@cs.ecnu.edu.cn; [3]jgliu@ee.ecnu.edu.cn

***Abstract*-Hyperspectral imaging is a relatively new method for identifying blood cells. Except for morphological and texture information in gray images, hyperspectral data contains a lot of spectral signatures which represent chemical analysis of a sample. Therefore, hyperspectrum has an advantage over digital color images due to spectral signatures. With these spectral and spatial features blood cells can be recognized and classified. Over 40 features are extracted from hyperspectral image sequence. These features include spectral pattern traits and similarity measures. To implement blood cell discrimination, a back propagation neural network (BPNN) is proposed in this paper. The connection weights of the BPNN were fixed through the training by a genetic algorithm (GA) which employed two adaptive mechanisms during the evolutional processes. Three-fold cross validation was applied to classify blood cells of given samples. Experimental results demonstrated that the classifier using a BPNN and an adaptive GA was effective. Finally, this paper also described a cursory investigation of the effect of spectral data volume on classification accuracy. Two compressed image series which can be viewed as multispectral series were obtained by systematic sampling from two original hyperspectral series, respectively. Compared to multispectral data, the hyperspectral data with high dimensionality achieved superior accuracy in recognizing blood cells, although requiring greater processing time due to the large amount of data dimension.**

Keywords- Classification; Blood Cell; Hyperspectral Imaging; BP Neural Network; Genetic Algorithm

I. INTRODUCTION

Blood cell identification is highly informative for detection and treatment of many diseases in clinical practice. Originally, blood cell classification was performed manually under a microscope, but manual methods are not only prone to error, but also tedious and time-consuming [1]. Subsequently, automatic devices with digital imaging equipment and pattern recognition techniques were employed [2], which obtained quantitative results and therefore are more objective than manual-based methods. Blood cells are recognized by analysis of morphology, color and/or texture features of different kinds of cells using gray or color images. Liu et al. [3] proposed a cell classification method by combining the synergetic theory and the prototype-set technique to analyze morphologic features of H&E-stained cells. In their study four common types of lymphocytes were classified and the accuracies were over 90.8%. Theera-Umpon and Dhompongsa [4] applied Bayes classifiers and artificial neural networks to automatically cluster white blood cells in bone marrow, but they analyzed just four morphology-based nucleus features and got 77% classification rate on the test sets. Su et al. [5] used a pulse-coupled neural network to segment and count cells according to their morphology through digital color images. They found that the method can well eliminate disturbed objects, and segment specific isolated cell from its background. However, they also found the algorithm was only suitable under the condition that the cells had similar attributes, such as size, shape and connections.

It is a challenge for automatic color-image-based systems to meet clinical practice needs because the accuracy of discrimination is affected by the complexity of image contents, the quality of images, the wide variability in cell characteristics, and the wide variation in cell maturity [6-7]. It is extremely difficult to reproduce an identical image even with the same instrument and under the same conditions. Since image quality is critical to pattern recognition, the utility of digital color images was limited. Therefore, in recent years the potential of spectroscopic techniques has been investigated in the field of pathology as a new method for the identification of blood cells [8-10].

Spectroscopy studies how materials uniquely absorb and emit light at certain wavelengths based on the particular molecular constitution of the object [11, 12]. The physiological constitution of biological tissues may be derived according to the spectral signatures generated by the interaction of electromagnetic waves on samples of these tissues. By exploring the spectral differences of dissimilar blood cells, some chemical and physical properties that are difficult to detect under traditional imaging methods may be revealed to help categorization. The main advantage of spectral data is their capability to produce measurable spectral patterns to aid the detection, identification and quantification of tissues. Wu and Shah [13] compare segmentation results by applying a watershed method to both multispectral absorption images and color images. They found that segmentation results using both approaches yielded high detection rates. Although the increase in accuracy of target cell segmentation using multispectral images was less than 1% compared to methods using color ones, the false positive rate using spectral technique decreased up to 25.26% compared to methods using color images. Researchers at Wuhan University focused on the problem of automated analysis of white blood cells in multispectral images of bone marrow [6-7]. They introduced methods using microscopic multispectral imaging techniques to segment and recognize blood cells, and applied classifiers to the spectra of each pixel. Experimental results demonstrated that multispectral information was very helpful for cell

segmentation.

Multispectral data series usually have tens of registered images and cover a narrow range of wavelengths, so the spectra are discrete. In comparison, hyperspectral imaging deals with more images over a continuous spectral range in a scene [14-15]. A hyperspectral image sequence is made up of hundreds of registered images over a range of wavelengths. Therefore, a hyperspectral series usually has higher spectral resolution. What's more, the use of hyperspectral imaging requires no prior knowledge of the sample since an entire spectrum is acquired at each point. Since hyperspectral features are more stable and reliable than spatial features, this method has appeared promising as a relatively new method to discriminate blood cells. Yet so far most of the studies about blood cell classification have used color images, and the limited numbers of studies using spectral imaging techniques are different in spectral range and resolutions [6, 7, 13, 14,]. In this paper, a method based on hyperspectral imaging technique is proposed to categorize blood cells. Hyperspectral imaging can make use of the spatial relationships among the different spectra in a neighborhood, allowing more elaborate spectral-spatial models for a more accurate segmentation and classification of the image.

II. MATERIALS AND METHODS

The hyperspectral image sequences in this paper were taken by a molecular hyperspectral imaging (MHI) system. This system was developed according to the principle of push-broom hyperspectral imaging commonly used in remote sensing. Its main components include a microscope, spectrometer, charge coupled device (CCD) detector, a light and a control module, as shown in Fig. 1. The light, provided by Kohler illumination, is split vertically by the spectrometer and then projected onto the CCD detector. The spectral range of this system is 400-860 nm which covers visible and near infrared wavelengths. More details about this system can be found in the literature [16].

Hyperspectral data comprises a sequence of gray images corresponding to different wavelengths. If the wavelength is recorded in three dimensions, spectral data is represented as a three-dimensional series with two dimensions recording the spatial information of the object, as shown in Fig. 2. It can be seen that each pixel in the spatial area of hyperspectral data has an observation curve representing the transmission energy spectrum of the materials. The spatial pattern for a certain pixel (x, y) in the data can also be treated as a vector in a space with the dimensionality equal to the band number. Each pattern contains many spectral signatures, with which we can identify the type of object to which the pixel corresponds.

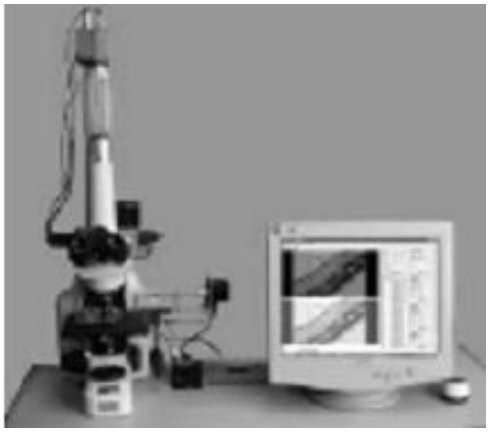


Fig.1 MHI system

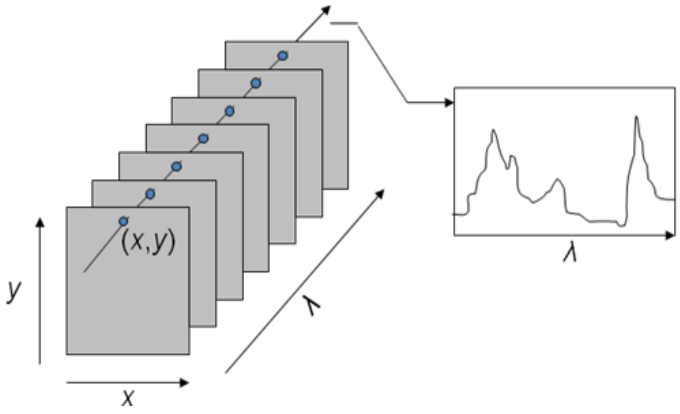
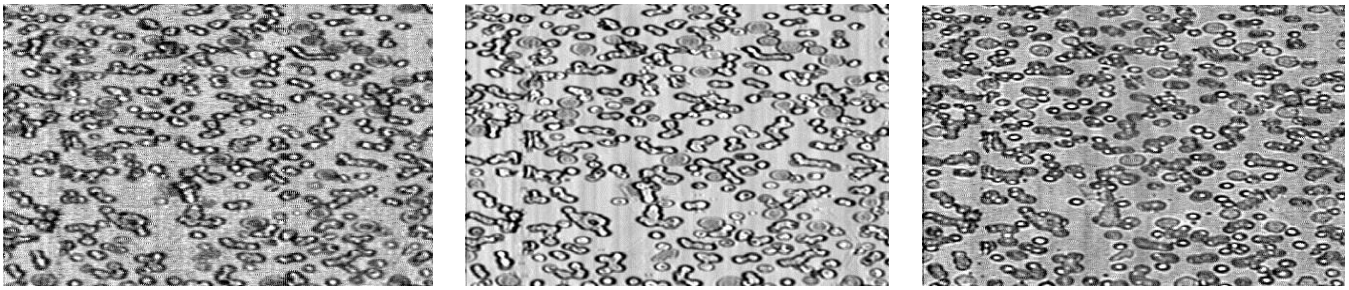


Fig. 2 Spectral data

The spatial resolution of an image made by the MHI system is related to the focus distances between the lens, the pixel pitch of the detector, and the magnification of the camera adapter coupled with the objective and CCD. In this paper, the resolution is 1.125 μm which creates an image of 460*300 pixels. As for the wavelength dimension, the spectral region of MHI is divided into 240 bands ranging from visible light to near-infrared. Since an individual image is formed in each band, the spectral data have 240 band images. Thus the spectral resolution is about 2 nm.

Blood samples in this paper were taken from hematology department of Shanghai Ruijin Hospital, PR China. Samples were obtained from first-time leukemia patients, who had not previously suffered side-effects of chemotherapy. All the images analyzed in this paper have been preprocessed, including baseline removal, radiation calibration, transmittance conversion, and noise reduction [17]. Due to the unavoidable noise interference caused by the device, some images of the head and tail end of hyperspectral sequences were discarded and only 210 images from 450nm to 758nm were analyzed. Three band images of a hyperspectral sequence made from a blood smear from a leukemia patient are shown as Fig. 3. It can be seen that even if they are of the same series in a scene, the gray images corresponding to different wavelengths still take on a dissimilar appearance.



(a) (b) (c)

Fig. 3 Three gray images of a hyperspectral sequence made by a leukemia blood smear corresponding to wavelength of (a) 778.0 nm, (b) 576.1 nm, and (c) 400.0 nm

III. FEATURE EXTRACTION

The preparation work of extracting features precedes identification. There are two major categories of hyperspectral data characteristics: spectral signatures and spatial features. Spectral signatures in this paper include spectral pattern features and similarity features.

A. Spectral Pattern Features

A simple pattern may contain useful information to help discrimination. The most important feature to be extracted in spectral data is Spectral Absorption Index (*SAI*):

$$SAI = \frac{d \times G_L + (1-d) \times G_R}{G_m}$$

where

$$d = \frac{\lambda_R - \lambda_M}{\lambda_R - \lambda_L}$$

λ_L, λ_R and λ_M are wavelengths of three stationary points, respectively, as shown in Fig. 4. Other useful features are minimum or maximum *DN*, and their corresponding wavelengths, respectively. The other features include depth (*H*), width (*W*), intensity (*P*) and Symmetry (*SYM*), which are calculated as follows:

$$H = d \cdot G_L + (1-d) \cdot G_R - G_M$$

$$W = \lambda_R - \lambda_L$$

$$P = \sqrt{H^2 + W^2}$$

$$SYM = \frac{\sum_{i=L}^{M-1} [(G_i + G_{i+1})(\lambda_{i+1} - \lambda_i)]}{\sum_{i=M}^{R-1} [(G_i + G_{i+1})(\lambda_{i+1} - \lambda_i)]}$$

B. Similarity Features

Similarity measures of two spectral patterns are based on the theory of sets and vectors. Here a spectral pattern is viewed as a vector [18]. Spectral angle (*SA*) mapper is the most widely used to extract the similarity features of two hyperspectral vectors. It is used to calculate the angle of two vectors in terms of their dot product and respective lengths:

$$\cos(\vec{x}, \vec{y}) = \frac{\vec{x}^T \vec{y}}{\|\vec{x}\| \cdot \|\vec{y}\|} = \frac{\vec{x}^T \vec{y}}{\sqrt{(\vec{x}^T \vec{x})(\vec{y}^T \vec{y})}}$$

There are other means to measure strength and direction of the relation between two spectral vectors. Two commonly used are: correlation coefficient (*r*) and covariance (*Cov*).

C. Spatial Features

Spatial features are referred to as the characteristics taken from wavelength images. In this article, average value and variance are calculated with 8-nearest neighborhood.

The above three kinds of features are rather important since they are the key components to identify different kinds of items. Fig. 5 shows the typical spectra of six kinds of objects in gray image: background, red cell, lapped/connected cells, lymphocyte, nucleu and plasma of tumor cell. As shown in this figure, the pattern of background pixels (dark solid line) is almost straight and thus can be easily distinguished from the other five kinds. As for the others, the principal differences appear between bands from 15 to 160. Therefore, more attention should be paid to these middle bands among the feature extraction. From Fig. 5 it can be seen that the troughs and crests of the spectra show great disparity in size, location, amplitude, etc. The pattern of red cells (red dots line) has two small crests which fluctuate narrowly above and below the background pattern. The pattern of connected cells (green dashes line) appears two big crests, which apparently differs from the others. The primary cause may be that it is overlapped by variant cells and thus changes the expression of chemical attribute. The curve of tumor nucleu (purple dash dot dot line) has a rather deep trough in Band 40 nearby. The pattern shape of tumor plasma (yellow solid line) is alike with that of lymphocyte (gray dash dot line) with higher value.

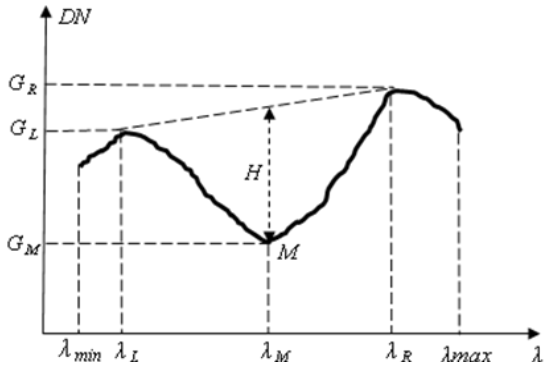


Fig. 4 Analysis of spectral pattern signatures

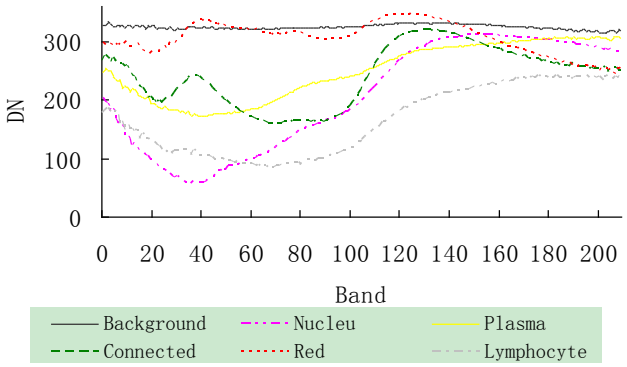


Fig. 5 Typical spectra of six kinds of pixels

IV. CLASSIFICATION ALGORITHM

A neural network provides parameterized, non-linear mapping between inputs and outputs. It has the inherent capability to deal with fuzzy information whose functional relations are not clear. Therefore, artificial neural networks are useful tools in recognizing patterns in complex data. Related literatures have showed they are powerful tools to categorize blood cells [19-20]. The commonly used neural networks, BPNNs, are multi-layered dynamic systems that propagate errors from the output nodes. During the BP network learning process, the error is propagated backward through the network to adjust the weights using the minimum error criterion. The framework of the BPNN used is shown in Fig. 6. The inputs include spectral pattern features, similarity features, and spatial features extracted by the above means. To identify the blood cells in a hyperspectral data, each pixel's pattern is investigated one by one to decide which kind of object it belongs to. The judge criterion is the spectral traits and the similarity measure mentioned in the above part. To measure the relationship of two patterns, the six typical spectra are the reference vectors. Therefore there are 4*6=24 similarity attributes. Adding the extracted traits of spectral pattern, the pixel can then be determined. The outputs are indexed from 1 to 6 to represent the six kinds of objects to be discriminated.

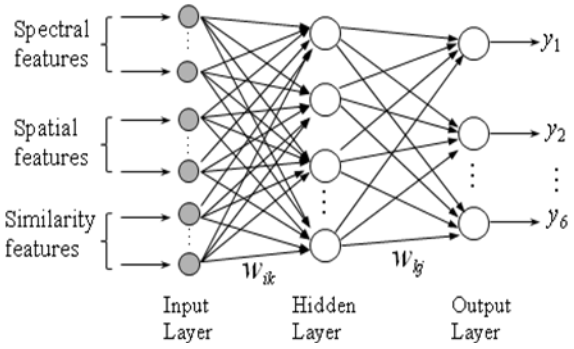


Fig. 6 Classifier architecture

BPNN is an adaptive system which changes its structure based on external or internal information that flows through the network. The key to back propagation is a method for calculating the gradient of the error with respect to the weights for a given input. The traditional network weight training generally uses a gradient descent method, which makes it easily stuck to local optimum. In this paper, an adaptive GA is used to search for a set of optimal or approximately optimal weights without computing gradient information and initializing connection weights [21-23]. Next, the procedure is completed by applying BP training with the inherent learning rule of neural network to obtain the final weights.

The fitness function of GA is to assign to each individual in the population a fitness score which depends on how well its

chromosome solves a given objective. In this paper it can be described as:

$$y = \frac{1}{2}\sum_{j=1}^{6}(\hat{y}_j - y_j)^2$$

where $\hat{y}_j$ and y_j represent actual output and desired one, respectively. In general, GAs are effective global optimal algorithms, but in some cases they will be easily trapped into premature convergence. Therefore an adaptive strategy is often needed for the process of an optimal search. In this paper two adaptive mechanisms were applied to crossover and mutation processes, respectively [8]:

$$P_c = \begin{cases} P_{c0}, & N \le N_f \\ P_{c0} + (\alpha - P_{c0}) \cdot \dfrac{N - N_f}{N}, & N > N_f \end{cases}$$

$$P_m = \begin{cases} P_{m0}, & N \le N_f \\ P_{m0} + (\beta - P_{m0}) \cdot \dfrac{N - N_f}{N}, & N > N_f \end{cases}$$

where α and β are real constants, N represents the lasting generation number having the same best solution, N_f is a positive integer constant, and P_{c0} and P_{m0} are initial crossover probability and mutation probability, respectively.

Table 1 shows the GA settings for this classifier. As for the genetic operators, one-point crossover and uniform mutation were utilized. The initial values of P_c and P_m were 0.7 and 0.01, while the maximums were 1.0 and 0.2, respectively.

TABLE 1 GA SETTINGS

Items	Settings
Population size	40
Max evolution	200
P_{c0}	0.7
P_{m0}	0.01
α	1.0
β	0.2
N_f	5

V. RESULTS AND DISCUSSION

Three-fold cross validation was employed to do the classification. The original sample was randomly and equally divided into three subsamples. Every time, a single subsample was retained as the validation data, while the remaining two were used as training data. The process was then repeated three times, with each of the three subsamples used exactly once as the validation data. Finally the validation results were averaged over the rounds, as shown in Table 2. It can be clearly seen from the table that the identification accuracy of background pixel is as high as 95.0%, while the overlapped/connected cells is just 84.8%. Fig. 7 compares the classification accuracies of different items in a chart. The high background recognition should attribute to their stable and plain spectral pattern, which makes it quite distinct from other kinds of objects. It is difficult to identify overlapped cells, especially when the cells belong to more than one category. In this case the spectral signatures contain mixed information from two or more object classes. As for the recognition of lymphocytes, the accuracy is relatively low. The main reason is the insufficient training data due to the low number of red cells in the sample images. Compared to the Nucleus, it is found that the plasma of a tumor cell is difficult to identify, likely due to the variety of cell states developed with the disease.

The disadvantages of the hyperspectral imaging technique are the significant storage capacity re-quirements and the greater processing time due to the huge volume of data dimension. In exchange, the hyperspectral approach achieves superior identification and is an effective means for blood cell classification.

TABLE 2 BLOOD CELL CLASSIFICATION ACCURACIES WITH HYPERSPECTRAL IMAGES SEQUENCE

Items (%)	1	2	3	Avg.
Background	92.4	95.6	97.1	95.0
Nucleu	92.5	88.7	93.4	91.5
Plasma	89.7	86.3	88.4	88.1
Overlapped	87.8	84.6	81.9	84.8
Red	94.1	91.4	91.9	92.5
Lymphocyte	87.9	88.6	86.0	87.5

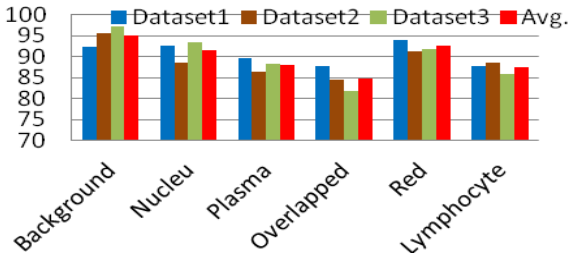


Fig. 7 Classification accuracies compared in a chart

TABLE 3 COMPARISON OF CLASSIFICATION RESULTS OF THREE SPECTRAL SEQUENCES

Items	Data10	Data50	Data210
Band number	10	50	210
Nucleu (%)	87.2	90.2	91.5
Plasma (%)	84.1	86.9	88.1
Overlapped (%)	78.3	83.3	84.8
Red (%)	87.5	91.0	92.5
Lymphocyte (%)	82.8	85.9	87.5
Time (s)	5.16	7.25	10.32

VI. CONCLUSIONS

The hyperspectral imaging technique was applied to the blood cell classification through microscopy in this paper. The first step was to extract object features, including spectral features, similarity features and spatial features. Next, a BP neural network and an adaptive genetic algorithm were employed to perform the categorization. The connection weights of the BPNN were fixed by an adaptive GA. The GA implied two adaptive mechanisms during the evolutional processes of processes of crossover and mutation to avoid being trapped into a local optimum. Results demonstrated that hyperspectral imaging techniques work for blood cell classification and that the classifier using a BPNN and an adaptive GA is effective. Finally, two compressed series of multispectral data were applied to the identification. The comparative analysis of spectral series with different band numbers showed that classification using hyperspectral image sequences is more accurate than classification using multispectral ones. As high-performance as hyperspectral data is, there are still disadvantages, which include significant storage capacity and greater processing time due to the huge volume of data dimension.

Our research on blood cell identification using hyperspectral imaging techniques is still preliminary. Employment of the process for blood cell recognition in clinical practice will require further enhancement. The main factors influencing accuracy include the extraction of value features, and the performance of the classifier, both of which need to be improved in the future.

ACKNOWLEDGMENTS

This work is funded by Shanghai Municipal Educational Commission (NO. 12YZ190). Many thanks are given to Ms. D. Chang, Dr. G. Schellhammer and Ms. Ruth for their kind help.

REFERENCES

[1] S. Nilufar, N. Ray, H. Zhang, “Automatic blood cell classification based on joint histogram based feature and bhattacharya kernel”, in: Proceedings of Asilomar Conference on Signal, System and Computers, 2008, pp. 26-29.

[2] R. Adollah, M. Y. Mashor, N. F. Mohd Nasir, “Blood cell image segementation: a review”, in: Proceedings of International Conference on Biomedical Engineering, 2008, pp. 141-144.

[3] B. Liu, Y. Liu, J. Zhang, Y. Zeng, W. Wang, “Application of the synergetic algorithm on the classification of lymph tissue cells”, Computers in Biology and Medicine 38, 2008, pp. 650-658.

[4] N. Theera-Umpon and S. Dhompongsa, “Morphological granulometric feature of nucleus in automatic bone marrow white blood cell classification”, IEEE Transactions on Information Technology in Biomedicine 11 (3), 2007, pp. 353-359.

[5] M. Su, Z. Wang, H. Zhang, Y. Ma, “A new method for blood cell image segmentation and counting based on PCNN and autowave”, in: Proceedings of IEEE International Symposium on Control, Communications and Signal Processing, 2008, pp. 12-14.

[6] N. Guo, L. Zeng, Q. Wu, “A method based on multispectral imaging technique for white blood cell segmentation”, Computers in Biology and Medicine 37, 2006, pp. 70-76.

[7] Q. Wu, L. Zeng, H. Zheng, and N. Guo, “Precise segmentation of white blood cells by using multispectral imaging analysis techniques”, in: Proceedings of IEEE International Conference on Intelligent Networks and Intelligent Systems, 2008, pp. 491-494.

[8] C. Dai, Q. Li, J. Liu, “Band selection for biomedical hyperspectral data studies using genetic algorithms”, in: Proceedings of International Conference of Bioinformatics and Biomedical Engineering, 2009, pp. 11-13.

[9] Q. Li, Y. Wang, J. Zhang, “Quantitative Analysis of Protective Effect of Erythropoietin on Diabetic Retinal Cells Using Molecular

Hyperspectral Imaging Technology", IEEE transactions on biomedical engineering 57, 2010, pp. 1699-1706.

[10] Q. Li, C. Dai, H. Liu and J. Liu, "Leukemic cells segmentation algorithm based on molecular spectral imaging technology", SPIE, 7383, 2009.

[11] A. T. Harris, "Spectral mapping tools from the earth sciences applied to spectral microscopy data", Cytometry Part A 69, 2006, pp. 872-879.

[12] N. J. Martin, J. Bunch, H. J. Cooper, "Dried blood spot proteomics: surface extraction of endogenous proteins coupled with automated sample preparation and mass spectrometry", Journal of the American Society for Mass Spectrometry, 24, 2013, pp. 1242-1249.

[13] X. Wu and S. Shah, "Comparative analysis of cell segmentation using absorption and color images in fine needle aspiration cytology", in: Proceedings of International Conference on System, Man and Cybernetics, 2008.

[14] M. Li, W. Wang, D. Yang, S. Wang, "Support vector machines for multispectral microscopic cell image segmentation", Computer Engineering and applications 8, 2006, pp. 37-43.

[15] J. Deglon, A. Thomas, P. Mangin, S. Staub, "Direct analysis of dried blood spots coupled with mass spectrometry: concepts and biomedical applications, Anal. Bioanal. Chem. 402, 2012, pp. 2485-2498.

[16] G. Xiao, R. Shu, Y. Xue, "Design of microscopic hyperspectral imaging system", Optics and Precision Engineering 12, 2004, pp. 367-372.

[17] H. Liu, Y. Guan, Q. Li, J. Liu, Y. Xue, "Radiometric correction of hyperspectral imaging data in spacial dimension and spectral dimension", in: Proceedings of International Conference on Mechanic Automation and Control Engineering, 2011, pp. 4265-4268.

[18] M. Borengasser, W. Hungate, R. Watkins, "Hyperspectral Remote Sensing: Principles and Applications", CRC Press, Boca Raton, 2008.

[19] A. Khashman, "Blood cell identification using emotional neural networks", Journal of information science and engineering 25, 2009, pp. 1737-1751.

[20] M. Adjouadi, M. Ayala, M. Cabrerizo, N. Zong, G. Lizarraga, M. Rossman, "Classification of leukemia blood samples using neural networks", Annals of biomedical engineering, 38, 2010, pp. 1473-1482.

[21] D. Venkatesan, K. Kannan, R. Saravanan, "A genetic algorithm-based artificial neural network model for the optimization of machining processes", Neural Computing & Applications 18, 2009, pp. 135-140.

[22] C. W. M.Yuen, W. K. Wong, S. Q. Qian, L. K. Chan, E. H. K. Fung, "A hybrid model using genetic algorithm and neural network for classifying garment defects", Expert Systems with Applications 36, 2009, pp. 2037-2047.

[23] Z. Fu, J. Mo, L. Chen, W. Chen, "Using genetic algorithm-back propagation neural network prediction and finite-element model simulation to optimize the process of multiple-step incremental air-banding forming of sheet metal", Meterials and Design 31, 2010, pp. 267-277.

Chunni Dai received the M.S and Ph.D degrees from East China Normal University, Shanghai, China, in 2006 and 2009, respectively.

She was first with Guangxi University for Nationalities from July, 2000 to July, 2010. Currently she is an Associate Professor at the Department of Computer Science and Technology. Her research interest spans several areas include multimedia communication, intelligent video surveillance, and pattern recognition. She has been served as reviewer for some journal/conference and published 20 papers on journals or conferences

Qingli Li received the B.S. and M.S. degrees in computer science and engineering from Shandong University, Jinan, China, in 2000 and 2003, respectively, and the Ph.D. degree in pattern recognition and intelligent system from Shanghai Jiaotong University, Shanghai, China, in 2006.

He is currently with the Key Laboratory of Polor Materials and Devices, East China Normal University, Shanghai, China, where he is engaged in research on medical imaging, pattern recognition, and image processing.