Journal of Applied Science and Engineering Innovation, Vol.5, No.4, 2018, pp. 104-108
ISSN (Print): 2331-9062
ISSN (Online): 2331-9070

Blood Leukocyte Microscopic Image Recognition Algorithm Based on Multi-feature Fusion

WU Xiao-qiang[1], ZHANG chun-you[1], WANG Li-hua[1], LI Li[1]

[1] *College of Mechanical Engineering, Inner Mongolia University for the Nationalities, Tongliao 028000, China*

Abstract: Image analysis systems based on pattern recognition technology and artificial intelligence has gradually become one of the methods for automatic quantitative analysis and verification in the medical field. Blood leukocyte recognition can be used as an important basis for doctors to judge the type and severity of the disease, and is of great value for the study of hematological diseases in medical diagnosis. At present, the accuracy of white blood cell recognition in blood cell analyzers is low, and the clinical aids are very limited. For this reason, this paper proposes a multi-feature fusion neural network based on blood leukocyte microscopic images for white blood cell classification and identification. First, the white blood cells and the white blood cell nuclei are divided. Then, the morphological features, geometric features and texture features of the leukocyte binary images were extracted. Finally, based on the BP neural network, the classification and identification of five types of white blood cells were achieved. The experimental results show that this method can effectively identify five kinds of white blood cells, and the recognition accuracy is high. It has a high reference value for clinical decision making.

Keywords Pattern recognition; White cell recognition; Microscopic image; Feature fusion

INTRODUCTION

There is an immune system in the human body, which is a defensive structure for the body to protect itself. The immune system consists of three parts, namely, immune cells, immune organs and immune molecules. The immune cells mainly refer to the white blood cells in the blood. When the body is inflamed or infected by the pathogen, the leukocytes are concentrated to the site of the pathogenic bacteria through the deformation and encircle and devour the germs.

White blood cells have a relatively stable content in the blood of normal people. The situation of children and infants is different from that of adults. However, as their age increases, the total number of white blood cells will gradually approach adult levels [Warren A L et al., 2013]. Although the form of normal white blood cells is varied, when the body suffers from a certain blood disease, the white blood cell morphology will change significantly. Without good medical knowledge and clinical experience, it is difficult to make correct judgments on the types of bone marrow cells and thus cannot effectively diagnose blood diseases. Early leukocyte morphology analysis and classification methods rely mainly on medical experts. This traditional method is relatively inefficient, has a high demand for medical experts, and the recognition results depend on subjective judgments of human beings [Molina J R et al., 2012]. There is a great limitation in the diagnosis of clinical hematological diseases.

With the development of basic medicine and the continuous deepening of computer applications, medical examination methods applied to clinics are gradually increasing. In the aspect of hematology, various types of instruments for analyzing blood cell images have emerged, which have a high degree of automation in detection and various types of parameters [Ruberto C D et al., 2002]. These instruments and methods have a positive impact on the speed of detection. Currently available blood leukocyte recognition is mainly based on the principle of laser scattering and cytochemical flow, but due to the complexity and diversity of the target to be identified, such methods have the drawback of not having high diagnostic accuracy, and are not suitable for the transmission of a large number of medical images in hospitals.

In order to improve the accuracy of blood leukocyte recognition, this paper proposes an automatic recognition algorithm based on multi-feature fusion of leukocyte microscopic images.

WHITE BLOOD CELL MICROSCOPIC IMAGE SEGMENTATION

White blood cells are immune cells that have a fixed amount of blood. In the automatic identification of blood cells, the white blood cell segmentation effect directly influences the results of the next operation such as cell feature extraction and classification. White blood cell segmentation division is the most crucial step

White Blood Cell Nuclear Segmentation

The division of the white blood cell nucleus is the focus of the leukocyte division. Analyzing the blood sample images, we found that the white blood cells, white blood cells, red blood cells, platelets, backgrounds, and other dirty pieces interfered with in the blood cell image had very distinct color levels. No matter how the light and other conditions change, people can easily distinguish the cell type and its internal structure by naked eyes. There are many ways to divide the white blood cell nucleus. The division of the white blood cell nucleus is easier in image segmentation.

First, the original RGB color image is converted to the *rg* color space, and the converted image is converted to the HSI color space again. Then, the g-channel component of the *rg* space and the S-channel component of the HSI color space are separately extracted. It can be seen from the g component that the white blood cell and platelet pixel values are smaller, and other components have larger pixel values, and from the S component, the white blood cell and platelet pixel values are larger, and other components have smaller pixel values. Finally, the two channel components are normalized to obtain a new g component image *Ig* and S-component image *Is*

$$\begin{cases} I_g = (g - g_m) \times 255 / (g_M - g_m) \\ I_s = (S - S_m) \times 255 / (S_M - S_m) \end{cases} \quad (1)$$

where g_M and g_m represent the maximum value and the minimum value of the g component, respectively. S_M and S_m represent the maximum value and the minimum value of the S component, respectively.

Next, perform a mathematical operation as shown in Equation (2) on the two-component image to get a new image I_E

$$I_E = \begin{cases} I_s & if\ I_g < T_1 \\ I_s / I_g & others \end{cases} \quad (2)$$

where T_1 is empirical value. After the operation, the brightness of the nuclear region in the image I_E is enhanced, and the brightness of other regions is suppressed. Finally, image I_E is binarized and morphologically processed to obtain a binary image of the nuclear region X.

White Blood Cell Segmentation

The division of white blood cells has always been a difficult point in blood cell image segmentation. Different physical conditions, lighting conditions, and staining conditions can all result in different cell image quality. In addition to these external conditions, different types of leukocytes also have different characteristics, such as neutrophil neutrality, acidic eosinophils and alkaline basophils. These differences in the characteristics of different types of white blood cells in the Different colors appear in the same sample image. Neutral leukocyte cytoplasm is mostly transparent, similar to or the same as the background color, which undoubtedly makes the division of white blood cells more difficult. For a variety of blood sample images, there are currently few white blood cell segmentation methods that can simultaneously segment the cytoplasm of different types of white blood cells. This section proposes a white blood cell segmentation method based on color space and mathematical operations. Based on the characteristics of the sample image in each color channel, combined with a variety of commonly used simple and easy-to-operate algorithms in the field of image processing, a white blood cell segmentation method based on color space and mathematical operations is proposed. This algorithm is easy to understand, easy to operate, fast and high segmentation rate.

First, convert the original RGB image to grayscale space. From the grayscale image and its histogram, it can be known that the gray value of the background part of the grayscale image is large, while the white blood cell and the red blood cell have a relatively small gray value. Therefore, a global threshold segmentation of the grayscale image can be used to obtain a binary map of the leukocyte and red blood cell regions.

Observing the B channel of the RGB image, it can be found that the red blood cells and the white blood cell nuclei exhibit relatively small pixel values, and the white blood cell cytoplasm and the background have larger pixel values. Therefore, the global closed-band segmentation of the B-channel component can yield a binary map of the red blood cells and white blood cell nuclei.

Assume that the binary image of leukocyte nucleus is represented by *I*, the binary chart of leukocyte nucleus and erythrocyte is represented by *II*, and the binary chart of leukocyte and erythrocyte is represented by *III*. The nucleus and erythrocyte binogram *II2* can be obtained by logical OR of *I* and *II* The binogram *IV* of the erythrocyte region can be derived from the binogram *II2* minus the binogram. The effect of the global threshold segmentation is related to the proportion of the target object to the full-scale image. The number of white blood cells in the image may be one, and it may be two or four. Therefore, the leukocyte and erythrocyte region binograms obtained by the global threshold segmentation algorithm are not strictly accurate in the erythrocyte and leukocyte nucleus region binograms. However, the effect of the threshold setting on the final processing result of the image is very small.

We use a few mathematical operations to find the leukocyte binary map. The purpose of the operation is to obtain a complete leukocyte region. Through morphological operations including logic or, denoising, filling holes, morphological reconstruction and other modifications of the leukocyte binary image

is not strictly accurate, you can get the final desired white blood cell binary image. The fact that more than 50 samples are controlled within a certain range of illumination is segmented, and the experimental results show that this algorithm is particularly fast and has a good leukocyte-division effect on cell images within a certain range of illumination.

WHITE BLOOD CELL MULTI-DIMENSIONAL FEATURE EXTRACTION

Morphological Feature Extraction

Blood is essentially a liquid tissue composed of suspended cells such as plasma and cell debris. Red blood cells are the most common cells in the blood circulation and are normally about 5 billion cells per milliliter in humans. The mature erythrocytes have no cell nucleus and no organelles. Their main function is to transport and exchange oxygen and carbon dioxide. On the contrary, the number of white blood cells in the blood is much less. For the average person, there are about 4 to 10 million per milliliter. Leukocytes can be further subdivided into granulocytes, lymphocytes and monocytes. Among them, there are three types of granulocytes: neutrophils, basophils, and eosinophils. These cells are multinuclear and are often referred to as polymorphonuclear leukocytes.

Different types of white blood cells have different morphological characteristics. In this paper, 10 morphological features information were extracted, and 40 white blood cells of each type were selected as examples. The feature extraction was averaged to obtain the final morphological features.

Geometric Feature Extraction

Binary image chain codes are very suitable for representing binary images consisting of straight lines and curves, as well as describing the boundary contours of the image. Specifying the starting point coordinate of the chain and the sequence of the slope of the chain can describe the curve or straight line. We chose the nuclear perimeter, the area of the nucleus and the cell, and the circularity of the nucleus as the geometric features to identify the white blood cells.

A. nuclear perimeter

When the pixel is considered as a point, the perimeter is represented by a chain code, and the perimeter is calculated by calculating the chain code length.

$$L = \sum_{i=1}^{n} \Delta l_i = N_e + \sqrt{2} N_o \tag{3}$$

where N_e is the number of even-numbered segments in the chain code sequence, N_o is the number of odd-numbered segments in the chain code sequence.

B. the area of the nucleus and the cell

The area enclosed by the chain code sequence is

$$S = \sum_{i=1}^{n} a_{ix} (y_{i-1} + a_{iy} / 2) \tag{4}$$

$$y = \sum_{k=1}^{i} a_{ky} + y_o \tag{5}$$

where a_{ix} and a_{iy} represent the components of a_i on the x-axis and y-axis, respectively.

C. the circularity of the nucleus

Circularity is defined as

$$C = \frac{L^2}{4\pi S} \tag{6}$$

where L denotes the perimeter of the enclosed area, S denotes the area of the area.

Texture Feature Extraction

The two constituent elements of a texture feature are texture primitives and their permutations and combinations. The primitives of a certain size and shape, together with the density and periodicity and directionality of their arrangement, constitute the texture.

The texture reflects the grayscale statistics, spatial distribution and structure information of the image, and the texture feature has a good discrimination for the target. Selecting the appropriate data change range to extract the texture features can make the texture features of the image stand out, which is beneficial to the extraction of the required information. The texture features are mainly reflected in the nucleus and cytoplasm on the measurement of white blood cells. Texture features are used to describe or measure the size and uniformity of particles in leukocytes. In terms of leukocyte cytoplasmic texture, texture can be used as a feature to distinguish various types of leukocytes.

Gray level co-occurrence matrix is an analysis method based on statistical characteristics of gray distribution. The texture is formed by the repeated appearance of the gray distribution in the spatial position, so there is a certain relationship between the two pixels at every certain distance in the image space.

Gray-level co-occurrence matrix for texture analysis uses the following four main parameters to extract the following characteristic parameters that can reflect the different textures of the nuclear texture.

A. energy

$$E_1 = \sum_{i,j=0}^{n-1} p_{i,j}^2 \tag{7}$$

B. entropy

$$E_2 = -\sum_{i,j=0}^{n-1} p_{i,j} (\log p_{i,j}) \tag{8}$$

C. moment of inertia

$$I = -\sum_{i,j=0}^{n-1} p_{i,j} (i-j)^2 \tag{9}$$

D. correlation

$$C = \frac{\sum_{i=0}^{n-1}\sum_{i=0}^{n-1} ijp_{i,j} - \mu_x\mu_y}{\sigma_x\sigma_y} \quad (10)$$

WHITE BLOOD CELL RECOGNITION BASED ON BP NEURAL NETWORK

An artificial neural network is a mathematical or computational model that mimics the structure and function of a biological neural network. Artificial neural networks learn and process information through the human brain, with the ability to process nonlinear data and predict data. The basic information processing unit that the neural network operates is a neuron. Complex and massive neural networks consist of a large number of neurons. The characteristics of neurons determine the overall characteristics of the neural network to some extent. The neuron model is shown in figure 1.

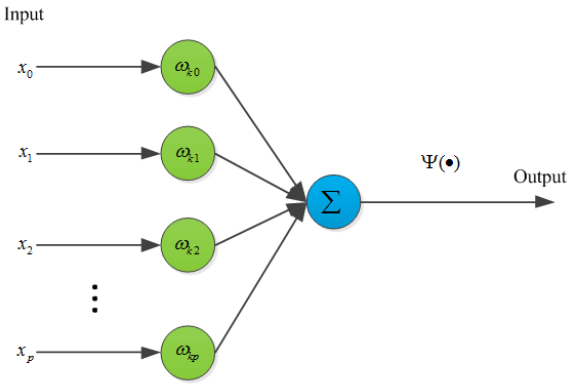

Figure 1 Neuron model

Artificial neural network topology is shown in figure 2. A neural network is an adaptive processing unit (neuron) that contains a large number of parallel computations. The learned information is stored in a variable weighted interconnected model. A multi-layer neural network has one or more hidden layer neurons so that more complex tasks can be learned by gradually extracting meaningful features from the input pattern. Neural networks are calculated by linking a large number of artificial neurons. In most cases, an artificial neural network is an adaptive system that can change the internal structure based on external information. Modern neural networks are nonlinear statistical data modeling tools that are used to model complex relationships between inputs and outputs or to explore patterns of data.

BP neural network consists of three parts: input layer, hidden layer and output layer. The hidden layer of the BP neural network can have many layers. The BP neural network is a multi-layer feed-forward network that relies on error reverse propagation training. First, the input signal enters the neural network from the input layer, passes through the hidden layer and finally reaches the output layer, and gets an output value. If the output value is consistent with the expected value, the algorithm ends; otherwise, the network returns along the original path of the network through the error between the output value and the expected value, and the gradient descent method adjusts the weight of each layer network to make the error signal smaller. As the error signal continuously propagates back, the weights of the network layers will be constantly revised, and the correctness of the network input mode will be correspondingly improved. When the network converges to have a smaller mean squared error parameter, it also completes the network training and learning process. Figure 3 is a white blood cell recognition framework based on BP neural network written in blood micrographs.

EXPERIMENT ANALYSIS

Divide the sample into two parts, one for training and the other for testing. 232 samples of test samples were collected, including 42 basophils, 47 eosinophils, 39 lymphocytes, 67 monocytes, and 37 neutrophils. The three-layer BP neural network includes 10 neurons in the input layer, 25 neurons in the hidden layer, and 5 neurons in the output layer. The extracted features are used as input to train the network. When the total error is less than 10-5, training is completed when the required error performance is achieved. The error function of the BP neural network classifier is shown in figure 4. White blood cell recognition results are shown in table 1.

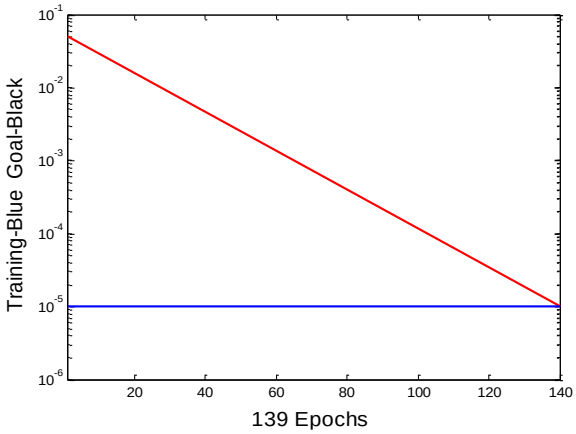

Figure 4 Neural network classifier error

The experimental results show that training the neural network and adjusting the weights can make the output closer to the target vector. The recognition rate of the five types of white blood cells in this algorithm is high. It can be seen from the results that the recognition rate of basophils and lymphocytes is relatively high, which is related to their cell characteristics.

CONCLUSION

Blood leukocytes have important clinical significance for the diagnosis of blood diseases, especially for the examination of leukocyte morphology, which has broad application prospects. The existing blood cell analyzer is not high enough in the accuracy of cell analysis, and has a large error from manual detection. This paper presents an automatic recognition algorithm of blood microscopic leukocyte images based on multi-feature fusion. This algorithm can achieve five classifications of leukocytes with high accuracy. With the continuous development of digitization of medical images, the research content of this paper has a positive application prospect, and the existing problems will be further solved in the future.

REFERENCES

Molina J R, Rojas R, Serrano J, et al. Invasive fungal diseases in children with hematological malignancies and stem cell transplant recipients[J]. Journal of Hematological Malignancies, 2012, 2(4).

Ruberto C D, Dempster A, Khan S, et al. Analysis of infected blood cell images using morphological operators[J]. Image & Vision Computing, 2002, 20(2):133-146.

Wick W D, Yang O O. War in the body: The evolutionary arms race between HIV and the human immune system and the implications for vaccines [J]. Slavic Review, 2014, 40(4):625-635.

Warren A L, Hecker K G, Nydam D V. Storage-associated changes in the bovine hemogram with the ADVIA 120 hematology analyzer [J]. Comparative Clinical Pathology, 2013, 22(6):1235-1240.