Blood Cell Identification and Segmentation by Means of Statistical Models

CHUNYAN YAO[1,2], JIANWEI ZHANG[2], HOUXIANG ZHANG[2]
[1]The MOE Key Laboratory of Mechanical Manufacture and Automation,
Zhejiang University of Technology, 310032 Hangzhou, China
[2]TAMS, Department Informatics, University of Hamburg
22527 Hamburg, Germany

Abstract: - Automatic cell identification and segmentation are important steps for medical automation to greatly reduce human labors. The paper used a statistical model solving the practical issues in this problem, including construction of a training set, cell shape generalization, deformable model building, cell searching and segmentation. Practical experiments prove the validity of the proposed scheme.

Key-Words: - Cell; medical image; segmentation; statistical models; deformable shape

1 Introduction

Clinical pathological analysis of peripheral blood and bone marrow is one of the most widespread but time-consuming investigations [1]. It is tedious work that pathologists make diagnostic decision by observing the specimen cells and analyzing the geometric parameters of the cell. Along with the development of technology in medical image processing, the application of computer for accurate measurement of geometrical parameters of the cell is necessary. Automatic blood cell recognition can be realized with high accuracy in relatively short time, in which the key technique is image segmentation.

Accurate segmentation from complex scene images is not only required but also often crucial first step, which should be solved in consequential analysis of blood cell. The purpose of segmentation is to provide richer information than that exists in the original blood cell images alone. Although many researchers have set forth segmentation algorithm [2]-[15], it is a challenging problem due to the complex nature of the cells, low resolution and complex scenes in the microscopic images.

This paper is arranged as follows. In the next section we briefly summarize related works on blood cell image segmentation. In section 3, we introduce the statistical shape model with the algorithm. The experiment and experimental simulation results are given in Section 4. Finally, Section 5 gives a conclusion.

2 Related works

When the first research about automated blood slide examination was proposed [2], a number of methods for the segmentation of cell image have been researched.

Active Contour Models (or snakes) has been used successfully by Kass et al [3] to detect boundaries and edges. The basic idea is energy minimize curve which can evolve until its boundary segment. The potential problem with this approach is that the topology of the region to be segmented must be known in advance. Wang [4] have presented a new deformable contour method derived from a constrained contour energy minimization framework to identify cell borders, but the results of cell contours do not correspond to the exact border of the cells, as seen in Fig. 1. Hu [5] uses an improved active contour model to isolate each cell nucleus. A growing energy based on region similarity is added to the energy function to overcome the initialization problem of conventional snake. Marko [6] introduced a novel approach, by applying greedy for the computation of active contours, for automatic segmentation of unstained living cells in bright-field microscope images.

Ongun etc. [7] imposed an active contour implementation together with its initial positioning for blood cell contour detection. They applied a constraint of region, which can be any function characterizing features of the contour interior structures (including homogeneity, texture, and color, etc.), to minimize

boundary-based contour energy. The method can incorporate a very general class of modeling information from both boundary and interior region.

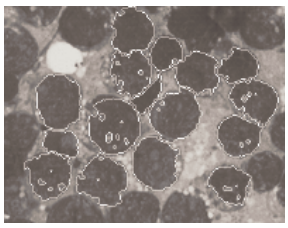

Fig 1. Segmentation result

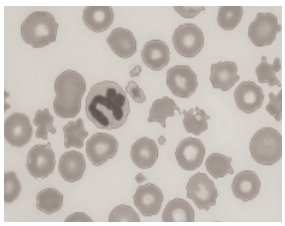

Fig 2. the final borders

Region growing algorithms [8][9] are also widely used to solve cell image segmentation problems. The algorithms begin with a set of pixels, called seeds, which mark the regions to be segmented and grow the regions around them based upon a certain homogeneity criteria. Because the color of cell nuclear is not homogeneous, it is difficult to correctly segment. It can bring to over-segmentation or less-segmentation.

Other approaches to various cell image segmentations have been proposed. Bamford [10] proposed an algorithm based on the concept of water immersion which ensures closed cell boundaries for cervical cell image segmentation. Tianzi [11] uses both the edge information and shape information of the cell images and then applies a parallel genetic algorithm to accurately segment human thyroid and small intestine cell images. Theera-Umpon [12] proposed a technique which is based on the fuzzy C-means clustering and mathematical morphology to segment nucleus of bone marrow white blood cells. Anoraganingrum [13] also applied both mathematical morphology operation and median filter to segment white blood cell.

In recent years, Jiang [14] uses scale-space filtering and watershed clustering for white blood cell segmentation. Ritter [15] presents a fully automatic method for segmentation and border identification of all cells in an image taken from a peripheral blood smear slide. The algorithm combines automatic threshold selection with connected-components and a novel adaptation of Dijkstra's shortest path algorithm. The disadvantage of this algorithm is that the overlapped cell boundary can not be separated (Fig. 2).

In the clinical practice blood cell image, overlapped and cluttered blood cells are an inevitable. That is also a problem which should be resolved in the blood cell image automatic segmentation.

This paper proposes an alternative means of red blood cell segmentation algorithm based on statistical shape analysis. The aim is to realize blood cell automatic recognition by correct segmentation includes overlapped cells, and thus offers a useful and accurate method for improving the reliability in practical analysis and diagnosis

3 Statistical Shape Models

Statistical methods can be applied to analyze the shape differences and shape changes. One distinct statistical shape model was proposed by Cootes et. al. [16] with the most widely used techniques of point distribution models (PDM) [17]. A PDM is constructed in following three steps.

3.1 Construct training set

A statistical shape model is built from a training set of example shapes [18]. Each shape can be represented by a vector *xi* which is a set of landmark points along the shape's boundary.

$$x_i = [x_{i1}, y_{i1}, x_{i2}, y_{i2}, \cdots, x_{in}, y_{in}]^T \; 1 \le i \le n \quad (1)$$

where x_{ij} and y_{ij} are coordinate of the point in the i-th shape, *n* is the number of point.

3.2 Align shapes

In order to align all shapes, the distance, rotation angle, and scale between two cell shapes should be obtained. According to the shape definition [19], we have

$$\begin{bmatrix} \hat{x}_{ik} \\ \hat{y}_{ik} \end{bmatrix} = \begin{bmatrix} s\cos(\theta) & -s\sin(\theta) \\ s\sin(\theta) & s\cos(\theta) \end{bmatrix} \begin{bmatrix} x_{ik} \\ y_{ik} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (2)$$

where, θ is parameter of rotation angle, s is parameter of shape scale, tx and ty are parameter of translation. The Procrustes analysis is then performed to align each shape of the training set. More detail of alignment approaches can be found in [20].

3.3 Principal component analysis

After aligned all shape in training set, apply principal component analysis (PCA) to reduce vector dimensionality. The aligned shape is a set vector and form a cloud of points in the nd-D space. PCA computes the main axes of this cloud, allowing one to approximate any of the original points using a model with fewer than *n*d parameters [21]. The approach can be described as follows.

Calculate the mean shape x,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i \qquad (3)$$

Then compute the shape covariance matrix *C*,

$$C_{xa} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T \qquad (4)$$

Finally compute the eigenvectors pi and corresponding eigenvalues λi of *C* (sorted so that λi >λi+1). If P contains m eigenvectors corresponding to the largest eigenvalues, then we can approximate any one in the training set, x, using

$$x = \bar{x} + Pb \qquad (5)$$

where $P = (p_1, p_2 ... p_m)$ and b is a *m* dimensional vector.

For instance, Fig. 3 shows the principal axes of a 2D distribution of vectors. In this case any of the joint can be approximated by the nearest point on the principal axis through the mean vector. The number of modes to retain *m* can be chosen as to explain a given proportion *k*(eg 98%) of the variance exhibited in the training set.

$$k = \sum_{j=1}^{m}\lambda_j \Big/ \sum_{i=1}^{2n-1}\lambda_i \qquad (6)$$

where $\sum_{j=1}^{m}\lambda_j$ is the sum of the *m* largest eigenvalues, $\sum_{i=1}^{2n-1}\lambda_i$ is the sum of all the eigenvalues.

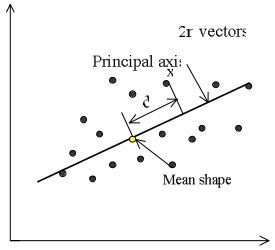

Fig 3. Principal Component Analysis

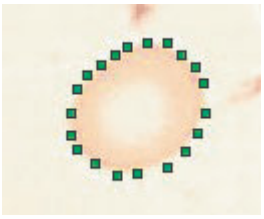

Fig. 4 The landmark points of the cell

3.4 Grey-level appearance model

Then, the new points on the target image should to be found to present the object, and the shape is transformed into a new and better location. The grey-level appearance [17][18] is also modeled to serve this purpose. Each landmark is moved along the direction approximately perpendicular to the contour to positions on either side, evaluating a total of positions. The landmark is put at the position with the lowest Mahalanobis distance. After moving all landmarks, the shape model is fitted to the displaced points, yielding an updated segmentation.

4 Experiment and Results

4.1 Morphological pre-processing

Mathematical morphology offers a powerful tool for image processing, for example, enhancement, segmentation, restoration, edge detection, texture analysis, etc.[13][22]. In this paper, we are also interested in morphological techniques for blood cell image pre-processing. The blood cell image in our experiment is peripheral smear of a patient with moderately severe liver disease, getting from [23]. Some red blood cells (RBC) are hypochromic, probably due to iron deficiency. We neglect the appearance of cells and get the center position of cell in the image by performing a series of morphological openings.

4.2 Point distributions

In the experiments, we selected 30 single cell construct shape training set. Each cell has been labeled with 20 landmarks by hand. Figure 3 illustrates these landmarks of one cell.

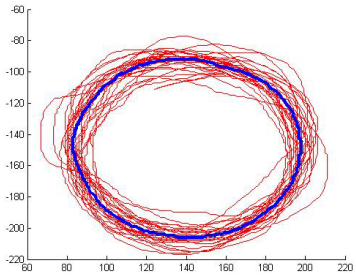

Fig. 5. the mean shape and training set of the cell

We used the first shape of the training set as the first mean shape. By iterative processing, the true first mean shape is picked up from the training set. After a process of five iterations, the mean shape became stable. Here the first shape of the training set influence greatly on mean shape. Therefore the common shape is selected as the first mean shape. Figure 4 illuminates the blood cell

training set drawn by red line and mean shape drawn with bold blue line.

In our experiment, let the proportion k = 0.97, and we perform the principal component analysis to decrease dimensionality of the aligned point set. Figure 5 illustrates the results of PCA. As can be seen from Fig. 6, the first 12 main Eigenvalue contribute to over 97% of the variation in the component set. Therefore, we truncate the vector after 12th component, neglecting the rest components.

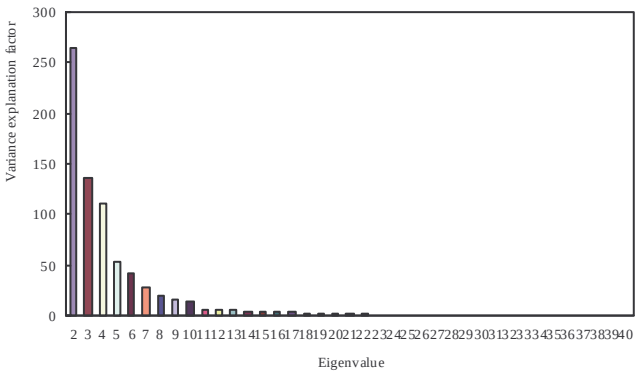

Fig.6 The result of PCA

4.3 Cell segmentation

In the experiment, we set forth 5-points sampling rule to orientate the direction of landmark point. As shown in Fig. 7, a cell object is a closed contours, each landmark point can be seen as the first point or the end point the direction perpendicular to a landmark (x_i, y_i) is calculated by its four neighbor points (x_{i-1}, y_{i-1}), (x_{i+1}, y_{i+1}), (x_{i-2}, y_{i-2}) and (x_{i+2}, y_{i+2}). At first, we compute two normal lines of four points, where one is for the two most near point and the other is the rest two points. Then, we calculate the mean of two normal lines. Let the mean value as the direction of this landmark point. Each landmark is moved along this direction to positions on both inner and out side. On each side, 10 pixels are sampled using a fixed step size. Repeat this process until convergence and finally we obtain the identified cells and their segmentation boundaries. The experimental results are illustrated in Fig. 8.

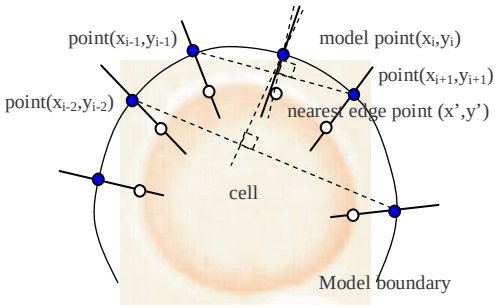

Figure 7 cell shape searching

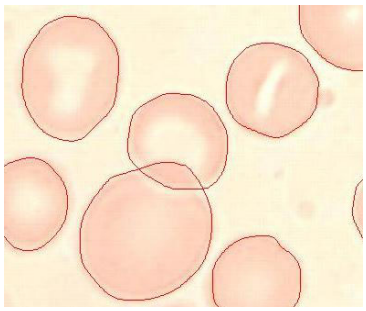

Figure 8 cell shape identification and segmentation

5 Conclusion

Automatic cell identification and segmentation are important steps for medical automation to greatly reduce human labors. The paper used a statistical model solving the practical issues in this problem, including construction of a training set, cell shape generalization, deformable model building, cell searching and segmentation. Practical experiments prove the validity of the proposed scheme.

References:

[1] Morphometric Investigations of Erythrocytes, http://www.videotest.ru/

[2] Bentley, S. & Lewis, S. (1975), "The use of an image analysing computer for the quantification of red cell morphological characteristics", *British Journal of Haematology* 29, 81/88.

[3] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. 1st International Conference on Computer Vision, pages 259{268, London, June 1987.

[4] Wang, X., He, L. Wee, W., `Deformable contour method: A constrained optimization approach', International Journal of Computer Vision, 59(1), 2004, 87- 108.

[5]Min Hu; Xijian Ping; Yihong Ding Automated cell nucleus segmentation using improved snake Image, Processing, ICIP 2004 , 24-27 Oct. 2004 Page(s): 2737 - 2740 Vol. 4.

[6] Marko Tscherepanow, Frank Zollner, and Franz Kummert, Automatic Segmentation of Unstained Living Cells in Bright-Field Microscope Images, Workshop on Mass-Data Analysis of Images and Signals, MDA 2006, pages 86-95, July 2006.

[7] Ongun, G., Halici, U., etc., Automated contour detection in blood cell images by an efficient snake algorithm, Nonlinear Analysis-Theory Methods & Applications 47(9), 2001, 5839-5847.

[8] Mehnert Andrew, Jackway Pau1, An improved seeded region growing algorithm. Pattern Recognition Letters, 1997, 18(6)：1065- 1071.

[9] S. A. Hijjatoleslami, and J. Kittler, “Region growing: A new approach,” *IEEE Trans. Image Processing*, Vol. 7, 1998, pp. 1079–1084.

[10] Bamford, P. and Lovell, B. C., A Water Immersion Algorithm for Cytological Image Segmentation. APRS Image Segmentation Workshop, , Sydney, Australia, 1996, 75-79.

[11]Tianzi Jiang Faguo Yang , An evolutionary tabu search for cell image segmentation; IEEE Transactions Systems, Man and Cybernetics, Part B, Oct 2002, Vol. 32, No. 5 pp: 675- 678.

[12] Theera-Umpon, "Patch-Based White Blood Cell Nucleus Segmentation Using Fuzzy Clustering," ECTI Transactions on Electrical Engineering, Electronics, and Communications, Vol.3, No.1, pp.15-19, February 2005.

[13] D. Anoraganingrum, “Cell Segmentation with Median Filter and Mathematical Morphology Operation,” Intl Conf on Image Anal and Proc, pp. 1043-1046, Sept 1999.

[14] K. Jiang, Q. Liao, and S. Dai, “A Novel White Blood Cell Segmentation Scheme Using Scale-Space Filtering and Watershed Clustering,” 2nd Intl Conf on Machine Learning and Cybern, pp.2820-2825, Nov 2003.

[15]Nicola Ritter ,James Cooper, Segmentation and Border Identification of Cells in Images of Peripheral Blood Smear Slides, Thirtieth Australasian Computer Science Conference, Ballarat Australia , 2007, 161-169

[16]T. F. Cootes, D. H. Cooper, C. J. Taylor, and J. Graham. A trainable method of parametric shape description. 2nd British Machine Vison Conference. Springer-Verlag, Sept. 1991, pp. 54–61.

[17] T. F. Cootes, C. J.Taylor, D. Cooper, and J. Graham. Training models of shape from sets of examples. In D. Hogg and R. Boyle, editors, 3rd British Machine Vision Conference. Springer-Verlag, Sept. 1992, pp. 9–18.

[18] Cootes, T., A. Hill, C. Taylor, and J. Haslam, The use of Active shape models for locating structures in medical images. Image and Vision Computing, 1994. 12: p.355-366.

[19] I. L. Dryden and K. V. Mardia. Statistical Shape Analysis. John Wiley & Sons, 1998.

[20] M. B. Stegmann and D. D. Gomez. A brief introduction to statistical shape analysis, March 2002. www2.imm.dtu.dk

[21] http://www.isbe.man.ac.uk/~bim/Models/

[22] Di Rubeto, C.; Dempster, A.; Khan, S.; Jarra, B. Segmentation of blood images using morphological operators, 15th International Conference on Pattern Recognition, Volume 3, 2000, pp.397 - 400

[23]Stanley Schrier, Red Blood Cell Diseases (RBC), ASH Image Bank, 2002.