

Detection of Pelger-Huet anomaly based on augmented fast marching method and speeded up robust features

Minglei Sun*, Shaobao Yang, Jinling Jiang and Qiwei Wang

School of Mechanical Engineering and Automation, Beihang University, XueYuan Road No. 37, HaiDian District, Beijing, China

Abstract. Pelger-Huet anomaly (PHA) and Pseudo Pelger-Huet anomaly (PPHA) are neutrophil with abnormal morphology. They have the bilobed or unilobed nucleus and excessive clumping chromatin. Currently, detection of this kind of cell mainly depends on the manual microscopic examination by a clinician, thus, the quality of detection is limited by the efficiency and a certain subjective consciousness of the clinician. In this paper, a detection method for PHA and PPHA is proposed based on karyomorphism and chromatin distribution features. Firstly, the skeleton of the nucleus is extracted using an augmented Fast Marching Method (AFMM) and width distribution is obtained through distance transform. Then, caryoplastin in the nucleus is extracted based on Speeded Up Robust Features (SURF) and a K-nearest-neighbor (KNN) classifier is constructed to analyze the features. Experiment shows that the sensitivity and specificity of this method achieved 87.5% and 83.33%, which means that the detection accuracy of PHA is acceptable. Meanwhile, the detection method should be helpful to the automatic morphological classification of blood cells.

Keywords: Pelger-Huet anomaly, fast marching method, SURF

1. Introduction

Pelger-Huet anomaly (PHA) is an autosomal dominant hereditary disease characterized by majority of the mature neutrophils showing hyposegmentation of their nucleus and excessive chromatin clumping. An acquired neutrophil dysplasia which is also known as pseudo Pelger-Huet anomaly (PPHA) has been described in hematological diseases and some clinical conditions. The nuclei of Pelger-Huet cells (PHC) are usually bilobed and appear in the form of a pince-nez, spectacle, or a dumbbell. In contrast, the nuclei of pseudo Pelger-Huet cells (PPHC) may be bilobed or unilobed. Pelger-Huet anomaly (PHA) is generally characterized by the presence of 55% to 95% PHC. In comparison, the percentage of PPHC in the so-called PPHA, which is associated with chemotherapy, myeloproliferative disorders, myelodysplasia, usually is 5% to 55%. In these conditions, the morphology of the nuclei usually returns to normal after the elimination of the primary disease, thus,

* Address for correspondence: Minglei Sun, School of Mechanical Engineering and Automation, Beihang University, XueYuan Road No. 37, Haidian District, Beijing, China. Tel.: 13683370429; Fax: 01082337098; E-mail: sunminglei@buaa.edu.cn.

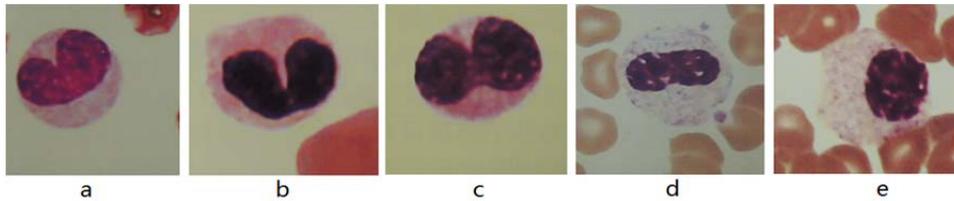


Fig. 1. a, b, c) Pelger-Huet cells (PHC), d, e) pseudo Pelger-Huet cells (PPHC).

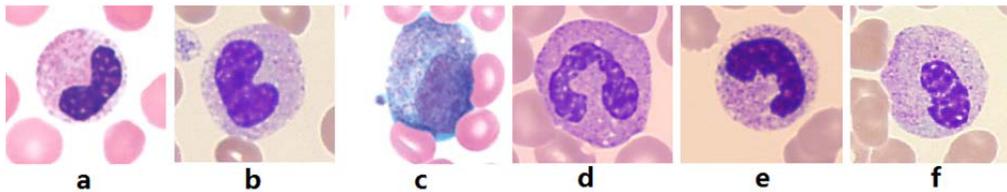


Fig. 2. Chromatin staining condition. a) Pelger-Huet cells (PHC), b, d, e, f) band neutrophil, c) myelocyte.

PPHA has important significance in clinical diagnosis [1].

At present, the identification of PHA and PPHA is mainly based on the following factors: PHC and PPHC show bilobed or unlobed nucleus with rod like, dumbbell, peanuts, spectacle, round or oval shapes and excessive clumping chromatin commonly with deeper staining [2-6]. Figure 1 shows several kinds of shapes of PHC and PPHC described in a blood cell morphological diagnosis atlas.

More remarkable, the karyotype of PHC is similar to band neutrophil. Besides, unlobed PHC can be mistaken for myelocyte and metamyelocyte based on karyotype. Pelger-Huet cells (PHC) show significant differences in nucleus dimension, chromatin density and staining conditions. Generally, a PHC or PPHC has a smaller nucleus, crowd and darkly stained chromatin compared with those of leukocytes.

With the wide use of computers, many applications of different methods on automatic morphology classification of leukocyte have exhibited promising results. For example, leukocyte segmentation algorithm based on level set method, nucleus topology modeling method based on skeleton and graph, as well as detection and grading method of toxic granulation in cytoplasm. Meanwhile, classification depending on the condition of the nucleus composition has not received warranted concern. With the standardization of the blood smear preparation process and the improvement of staining quality, clinical significance of this feature is more and more highlighted.

Under the optical microscope, chromatin staining is clear with visible distribution including purple stain composition and light color composition. As we can see in Figure 2, these light color compositions (blob) present different shapes, mainly including circular and thin and curved shapes. Useful information to represent chromatin cluster condition in clinical cases is the size and distribution of these blobs. Due to the difference of blob and staining quality of different cells, accurate detection of features with different dimensions and robustness as well as rapidity are needed to extract blob features precisely.

This paper present a detection method of PHA and PPHA based on nucleus morphology and chromatin distribution. The skeleton of karyon was extracted with the augmented Fast Marching Method (AFMM) and width distribution along the direction of skeleton was obtained with distance transformation method. Besides, Speeded Up Robust Features (SURF) detecting method based on the scale invariant theory is used to detect the locations and size of the blobs in the nucleus. Several parameters were introduced including scale coefficient of variation, saturation contrast of blob and

kayon as well as uniformity of location and size. Then, K-nearest-neighbor (KNN) classifier is constructed to comprehensively describe the chromatin distribution based on these parameters. At last, experiment of 100 samples (including 40 PHCs and PPHCs) shows the great feasibility of this method.

2. Image acquisition and preprocessing

The peripheral blood samples were obtained from Peking University First Hospital, Beijing, China. These slides were stained with Wright Giemsa stain (supplied by BASO). Research equipment includes an Olympus BX51 microscope and a Sony DXC-390P 3CCD color video camera connected to a computer. A 100× oil immersion Plan Semi Apochromatic objective was used for the acquisition of the experimental images.

The method is based on analysis of the nucleus, thus, accurate segmentation of karyon can greatly reduce the computation time and seriously impact on the detection accuracy. Images captured by CCD camera are RGB mode which is mainly used for the physical device output modeling. This image space is not suitable for karyon segmentation as each of its three components is sensitive to the illumination. The HSV (hue, saturation, value) is chosen as the karyon segmentation space because of the reason that it realizes the decoupling of brightness and hue, of which the hue channel can describe the pure color. The nucleus displays a marked difference with other ingredients in the saturation channel, so it was chosen as the segmentation channel. The self-adapting Ostu method [7] obtained good results.

3. Karyon shape detection

3.1. Extraction of karyon skeleton

A majority of the cells with PHA and PPHA show bilobed or unilobed nucleus with rod like, dumbbell, round or oval shapes. The accurate recognition of the shape of karyon is a determining factor. The skeleton of karyon can preserve the geometric and topological properties of the initial shape. The conception of the skeleton was firstly given in a flame propagation model by H. Blum [8], and the implementation of the method is broadly divided into morphological thinning methods, geometric methods based on Voronoi diagram, skeletonization using distance transformation [9] and skeletons via boundary evolution, etc.

Fast Marching Method (FMM) is similar to the regional growth process [10]. It presumes a velocity field F along the normal direction of the curve, and the time $T(x, y)$ when the curve reaching specified point (x, y) which can be obtained by solving the equation:

$$1 = |\nabla T|F$$

In case of two dimensions, it is equivalent to solve the equation:

$$\frac{1}{F} = \sqrt{\max(D_{x,y}^- T, 0)^2 + \max(D_{x,y}^+ T, 0)^2 + \max(D_{x,y}^- T, 0)^2 + \max(D_{x,y}^+ T, 0)^2}$$

Where D^- and D^+ respectively represent backward and forward difference operator.

In the augmented Fast Marching Method (AFMM) proposed by A. Telea and J. Wijk [11], the initial boundary is coded in order, and the coding propagates with the boundary. The code difference of Rendezvous point represents the importance of the point which is of high efficiency, strong robustness and parameters flexibility and adjustability.

The extraction process is illustrated as follows: firstly, determine the boundary and monotonic code of the boundary according to the principle of the region on the left of the curve, then, all the points in the area are divided into seed points, searching points and away points. After that, a circulatory process including marking the searching points who have the minimum $T = 1/F$ as seed points, marking the away points which are in the neighborhood of the seed points as searching points. This is followed by updating T of the away points until there is no searching point. In this process, code values propagate with the seed points and finally get to the skeleton, and the primary karyon skeleton is obtained by comparing the values. Figure 3 gives the skeleton extracting results of several blood cells. The method is of strong robustness and can extract the primary skeleton.

3.2. Width distribution along the skeleton

In this work, width distribution of points along the skeleton was obtained in order to describe the shape of the karyon. The width of an arbitrary point p :

$$d(p) = \min(\text{dist}(p, q))$$

The Euclidean distance $\text{dist}(p, q)$ is the distance from p to the boundary point q . The width distribution along the skeleton is shown in Figure 4.

The lowest value of the distribution curve generally indicates the splitting positions, as we can see in Figure 4, the wider nucleus width in the lowest value of the distribution curve and the overall smoothness showing in the curve are the expression of the hyposegmentation of the nuclei.

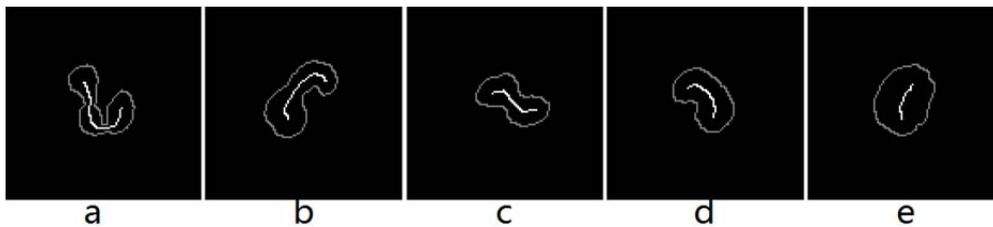


Fig. 3. a) segmented neutrophil, b) band neutrophil, c, d, e) Pelger-Huet cells (PHC).

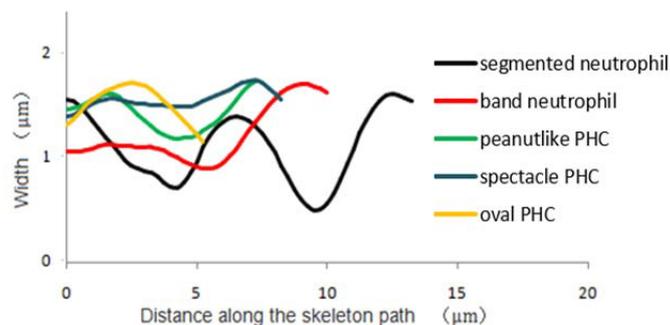


Fig. 4. Width distribution along the skeleton.

4. Chromatin distribution detection

4.1. Caryoplastin detection

Recognition of PHC and PPHC simply based on karyon morphological is not enough, for the reason that the nuclei of the rod like PHC and band neutrophil are similar. Besides, unilobed PHC can be easily mistaken for myelocyte or metamyelocyte. However, it can be differentiated according to the size of nucleus, coarse clustering degree and staining depth of the nuclear chromatin.

The size and shapes of the chromatin blobs are different. Clinically, detection only needs to distinguish the size and distribution of the blobs, thus, a fast robust feature extraction method is used to detect the roughness and compactness of the chromatin.

The scale of the image space expressed as a convolution of Gaussian function and the image represent the similarity of the image with Gaussian kernel. The multi-scale blob detection method confirms the location and optimum scale of the blobs via local maximum values in the normalized Gaussian scale space, and the characteristic scale of the blob represents the size of blob.

The strategy adopted in SURF algorithm [12] is an integral image, in which the image is iterated only once. Down-sampling method is increasing the size of the image kernel. The scale space of the image utilizes convolution of different scales of box-filter and image kernel, and obtain the extreme by Hessian matrix. The Hessian matrix expression of a point $p(x, y)$ in the integral image $I(x, y)$ at scale of σ :

$$L(p, \sigma) = G(\sigma) * I(x, y), G(\sigma) = \frac{\partial^2 g(\sigma)}{\partial p^2}$$

$$H(p, \sigma) = \begin{bmatrix} L_{xx}(p, \sigma) & L_{xy}(p, \sigma) \\ L_{xy}(p, \sigma) & L_{yy}(p, \sigma) \end{bmatrix}$$

Where σ is obtained by considering interesting points and interpolation tests; $g(\sigma)$ is a two-dimension Gaussian function; $L(p, \sigma)$ is the convolution of $G(\sigma)$ and an integral image, in this article, we use a box-filter to replace the Gaussian.

$$\det H(p, \sigma) = L_{xx}L_{yy} - L_{xy}^2$$

The location and optimum scale of caryoplastin can be verified through this algorithm, and the scale represents the size of these caryoplastins. Figure 5 shows that different blobs are marked with different feature scale vectors. Detections of blobs in different cells under the same scale vector are showed in Figure 6.

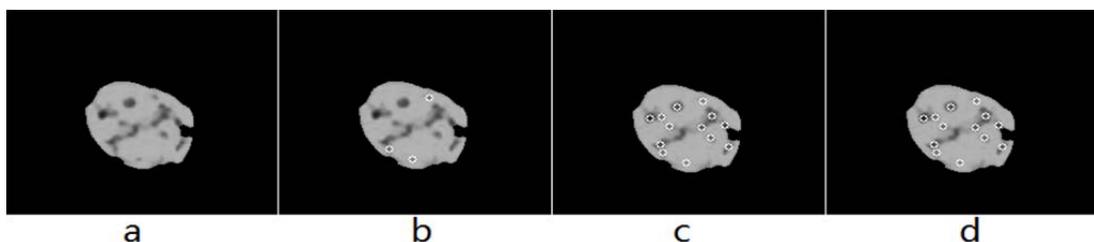


Fig. 5. Caryoplastin extraction of the same cell under different scale vector.

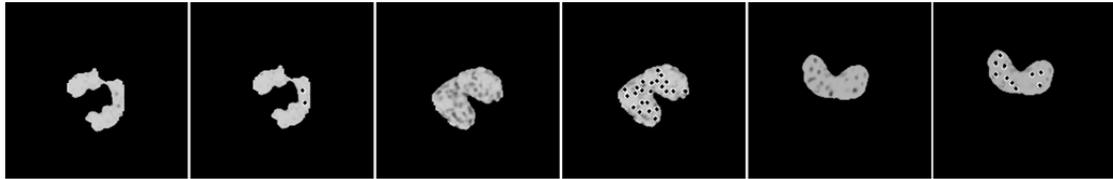


Fig. 6. Caryoplastin extraction of different cells under the same scale vector.

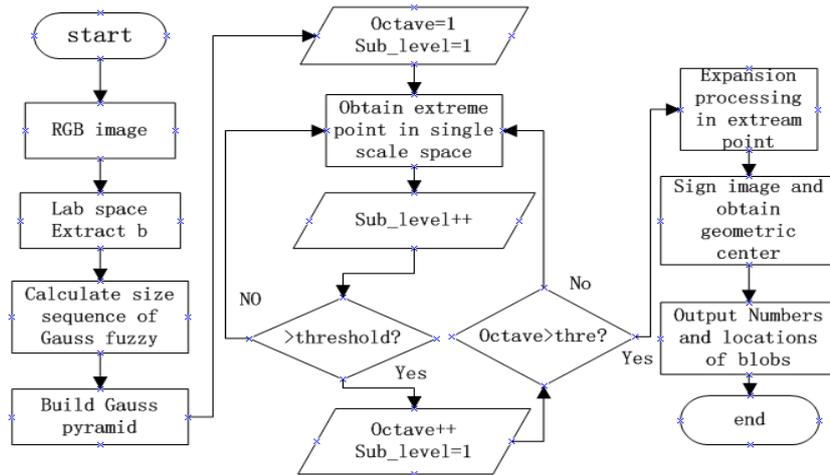


Fig. 7. The flowchart of blobs detection based on SURF.

4.2. Caryoplastin distribution description

Caryoplastin distribution contains much information of classification which is crucial. The numbers and locations of different scales of blobs are obtained with the SURF algorithm. Figure 7 shows the flowchart of SURF algorithm.

We introduce several parameters to describe the caryoplastin distribution:

- (1) Quantity of different scales of blobs N_i ;
- (2) Proportion of caryoplastin P :

$$P = \frac{N_i * i}{A_{karyon}}$$

A_{karyon} is the area of the karyon;

- (3) Average value of scale μ ;
- (4) Variable coefficient of scale CV :

$$CV = \frac{\mu}{\sigma}$$

σ is the standard deviation of scale;

- (5) Density of caryoplastin D :

$$D = \frac{\sum N_i}{A_{karyon}}$$

(6) Distribution uniformity of caryoplastin DU :

$$DU = \frac{\sum_{i=1}^m |D_i - D|}{mD}$$

m is the number of same-area regions along the skeleton. D_i is the caryoplastin density of i region.

4.3. K -nearest-neighbor (KNN) classifier

Principle component analysis (PCA) is used to reduce the correlation of the features extracted. A K -nearest-neighbor (KNN) classifier [13] is constructed based on the features and assessed in the test set. The fundamental of the KNN classifier is organizing the sample x according to the majority category of the k nearest training points. It partially views classification as a representation the test environment management and considers that a number of training samples that can provide a good approximation for the test sample should be selected as its “nearest neighbors”. The key point is the selection of initial seed points. This article designates the location of seed points in the multi-dimensional parameters, and finally achieves classification of PHA with other cells.

5. Experiment and analysis

To evaluate the proposed method, 100 microscopic images in $100\times$ magnification were selected to test the method, including 40 PHCs and PPHCs, band neutrophils, segmented neutrophil, myelocytes and metamyelocytes. As a result, the sensitivity was 87.5% and specificity was 83.33%.

Figure 8 indicates that the detection accuracy of PHC from myelocyte is superior because of the visible difference of the chromatin, and misdiagnosis as well as missed detection exist in classification of PHC (PPHC) from band neutrophil and metamyelocyte. Considering overall analysis of the film as an important reference, the experimental result is acceptable. In particular, this detection result may be relatively optimistic due to the limited sample numbers.

6. Conclusion

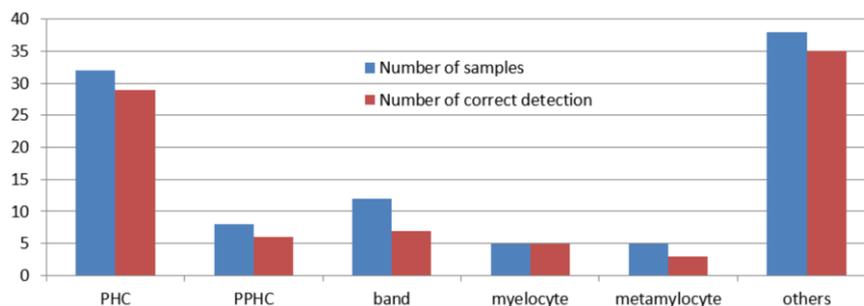


Fig. 8. Experimental results.

Detection of PHA and PPHA based on karyomorphism and chromatin distribution was reported in this paper.

Extraction of karyon skeleton using AFMM and width distribution along the skeleton including distance transformation.

Detection of caryoplastin based on SURF, analysis of distribution parameters and construction of K-nearest-neighbor classifier.

Follow-up work of this method is the selection of seed points of the classifier. According to the difference of staining condition, a self-adapting selection method of seed points should be applied: before the analysis of cells like PHA, band neutrophil, myelocyte and metamyelocyte, a certain number of obvious characteristic cells (generally segmented neutrophil) are selected firstly to estimate the staining condition, and location of seed points is adjusted correspondingly. In addition, recognition of the nucleolar region should be added to improve the detection.

References

- [1] Q. Wang, M.G. Zhang and Y.R. Zhang, Atlas of Hemocytology Diagnostics, Science Press, Beijing, 2008.
- [2] L.M.S.A. Dusse, A.M.B. Moreira, L.M. Vieira, D.R.A. Rios, R.M. Morais e Silva and M.D.G. Carvalho, Acquired Pelger-Huët: What does it really mean? *Clinica Chimica Acta* **411** (2010), 1587–1590.
- [3] R. Colella and S.C. Hollensead, Understanding and recognizing the Pelger-Huët anomaly, *American Journal of Clinical Pathology* **137** (2012), 358–366.
- [4] E. Wang, E. Boswell, I. Siddiqi, C.M. Lu, S. Sebastian, C. Rehder and Q. Huang, Pseudo-Pelger-Huët anomaly induced by medications a clinicopathologic study in comparison with myelodysplastic syndrome-related Pseudo-Pelger-Huët anomaly, *American Journal of Clinical Pathology* **135** (2011), 291–303.
- [5] L.M.S. Dusse, R.M. Morais e Silva, V.M. Freitas, G.M. Medeiros de Paula, L.M. Vieira and Maria das Gracas Carvalho, Pseudo-Pelger-Huët in kidney-transplanted patients, *Acta Haematologica* **116** (2006), 272–274.
- [6] E. Shanbrom, Z. Collins and S. Miller, "Acquired" Pelger-Huet cells in blood dyscrasias, *The American Journal of the Medical Sciences* **240** (1960), 100–106.
- [7] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Transactions on Systems, Man and Cybernetics* **SMC-9** (1979), 62–66.
- [8] H.A. Blum, A transformation for extracting new descriptors of shape, in: *Models for the Perception of Speech and Visual Form*, Weiant Wathen-Dunn, ed., Air Force Cambridge Research Laboratories (U.S.), Data Sciences Laboratory, M.I.T. Press, Cambridge, 1967, 362–380.
- [9] Y.A. Chahadeh, Skeletonization algorithm using chamfer distance transformation adapted to rectangular grids, the 13th International Conference on Pattern Recognition, Vienna, 1996, 131–131.
- [10] J.A. Sethian, A fast marching level set method for monotonically advancing fronts, *The National Academy of Sciences of the United States of America* **93** (1996), 1591–1595.
- [11] A. Telea and J.J. Van Wijk, An augmented fast marching method for computing skeletons and centerlines, in: *Proceedings of the Symposium on Data Visualisation 2002*, Eurographics Association, Barcelona, Spain, 2002, pp. 251–ff.
- [12] H. Bay, T. Tuytelaars and L. Van Gool, SURF: Speeded up robust features, *Proceedings of Computer Vision* **3951** (2006), 404–417.
- [13] P. Viswanath and T.H. Sarma, An improvement to k-nearest neighbor classifier, *Recent Advances in Intelligent Computational Systems (RAICS)*, Trivandrum, Kerala, India, 2011, pp. 227–231.