Identifying Neutrophils in H&E Staining Histology Tissue Images*

Jiazhuo Wang[1], John D. MacKenzie[2],
Rageshree Ramachandran[3], and Danny Z. Chen[1]

[1] Department of Computer Science & Engineering, University of Notre Dame, USA
[2] Department of Radiology & Biomedical Imaging, UCSF, USA
[3] Department of Pathology & Laboratory Medicine, UCSF, USA

Abstract. Identifying neutrophils lays a crucial foundation for diagnosing acute inflammation diseases. But, such computerized methods on the commonly used H&E staining histology tissue images are lacking, due to various inherent difficulties of identifying cells in such image modality and the challenge that a considerable portion of neutrophils do not have a "textbook" appearance. In this paper, we propose a new method for identifying neutrophils in H&E staining histology tissue images. We first segment the cells by applying iterative edge labeling, and then identify neutrophils based on the segmentation results by considering the "context" of each candidate cell constructed by a new Voronoi diagram of clusters of other neutrophils. We obtain good performance compared with two baseline algorithms we constructed, on clinical images collected from patients suspected of having inflammatory bowl diseases.

1 Introduction

Identifying neutrophils, a major type of immune cells, is a crucial step towards diagnosing acute inflammation diseases; the locations and number of neutrophils with respect to different tissue layers determine whether there is acute inflammation [6]. H&E staining histology tissue image is a very common imaging modality in clinical use. In such imaging modality, neutrophils are characterized as having multiple lobes in nucleus per cell and often hardly visible cytoplasm (in Fig. 1(e)-(f), the lobes are indicated by green arrows).

There are known methods [3,4,10] for identifying neutrophils in blood smear (see [2,5] for complete reviews). But they utilize specific properties of blood smear, which are not valid in H&E staining histology tissue images. For example, in blood smear, immune cells have quite salient colors and are well separated; the lobes of a nucleus are usually distinctive with each other; the cytoplasm is also distinctive from either the nuclei or background, which can help group the lobes of a single cell together. Further, we are not aware of any work that marks

* This research was supported in part by NSF Grant CCF-1217906, a grant of the National Academies Keck Futures Initiative (NAKFI), and NIH grant K08-AR061412-02 Molecular Imaging for Detection and Treatment Monitoring of Arthritis.

P. Golland et al. (Eds.): MICCAI 2014, Part I, LNCS 8673, pp. 73–80, 2014.

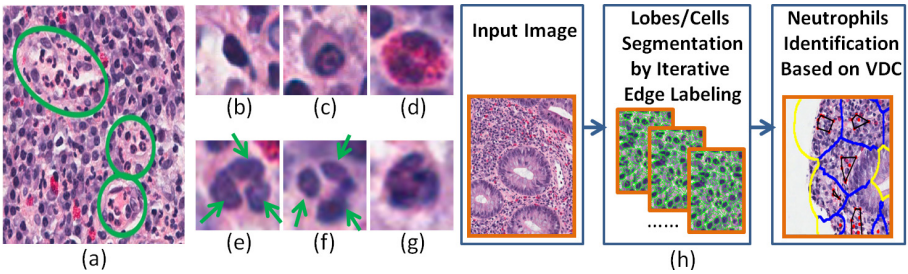

Fig. 1. (a)-(g) Image examples; (h) the framework of our method

neutrophils by a relevant antibody to analyze inflammation, and consequently, no immuno-staining method is known.

There are two main reasons for the lack of effective methods on H&E staining histology tissue images. (a) Identifying cells is inherently challenging [8] here, due to, e.g., non-uniform staining, complex and noisy tissue background, and crowded cells (Fig. 1(a)). (b) Identifying neutrophils is even harder. First, one has to group the lobes of each neutrophil together (otherwise, one may classify each lobe as a single-lobe cell). This is not easy due to the mixture of various types of cells, with either single lobe (e.g., lymphoctyes in Fig. 1(b) and plasma cells in Fig. 1(c)) or multiple lobes (e.g., eosinophils in Fig. 1(d)). Second, the lobes of a neutrophil may not appear perfectly well separated (i.e., lack of multiple-lobes characteristic), but crowded together instead (Fig. 1(g)), which makes grouping these lobes easier but classifying them as a neutrophil harder.

In fact, if based on only the appearance of each individual cell, pathologists are highly/moderately confident at marking roughly half neutrophils (easy cases). They must rely on the *context* of a cell to mark the rest neutrophils (hard cases): They compare a hard case with cells around that they can relatively easily identify, especially clusters of neutrophils (e.g., those inside green circles in Fig. 1(a)). Namely, the ambiguity and the subjectivity on the confidence scale of each annotation of neutrophil can be resolved after considering the context, making pathologists' annotations of all neutrophils still quite robust.

In this paper, we present a new method for identifying neutrophils in H&E staining histology tissue images. Our method combines a lobes/cells segmentation process based on iterative edge labeling and a neutrophils identification process based on Voronoi diagram of clusters (VDC). In segmentation stage (Fig. 1(h)), we seek to group the lobes of each cell (especially, each neutrophil) into one segment, which contains no lobes from other cells, since identification of neutrophils is dependent on the lobes. In identification stage (Fig. 1(h)), we capture quantitatively by VDC how pathologists resolve the ambiguity based on the context when identifying hard cases of neutrophils.

Our most significant contribution is an effective strategy for identifying neutrophils in H&E staining histology tissue images. Also, our iterative edge labeling approach is actually a generally applicable framework for segmenting multiple

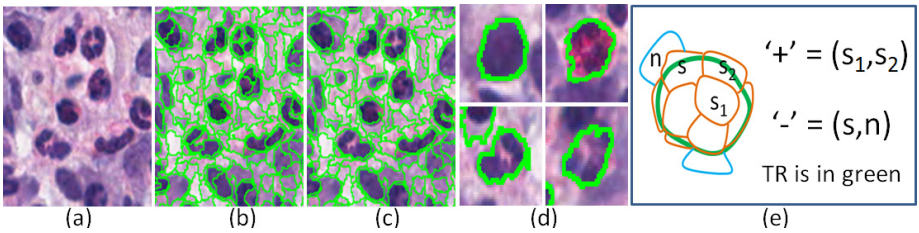

Fig. 2. Segmentation and construction of training examples

objects. Further, our VDC idea provides a new way to model quantitatively the local influence of clusters of objects for identifying objects around such clusters.

2 Method

2.1 Lobes/Cells Segmentation by Iterative Edge Labeling

Our goal of this stage is to obtain a segmentation of the input image, such that the lobes of each cell (especially, each neutrophil) are grouped into one segment, which contains no lobes of other cells. The resulted segments will be used in our later steps for identifying neutrophils.

Our main idea is to apply an iterative edge labeling process, which is essentially based on merging pairs of segments at different scales. At each iteration (i.e., scale), instead of hypothetically designing some criteria for segment merging, we decide to (a) learn such knowledge from training examples of edges in a segment adjacency graph (SAdjG, defined below), and (b) optimally refine it globally by applying MRF [9] to the edge dual graph (EDG, defined below) of SAdjG. By combining learning and refining characteristics exhibited by pairs of segments at different scales, we are able to alleviate or overcome difficulties caused by the high complexity nature of objects and background tissues in our imaging modality. Also, it is actually more natural to obtain segmentation of multiple objects by binary classification based on the *edge* view, rather than the *vertex* view, of SAdjG. This is simply because if two adjacent vertices of SAdjG are both labeled as foreground, one is not able to tell the corresponding two segments (sharing some common border pixels) are one or two objects.

Iterative Edge Labeling. We present steps of our algorithm in detail below.

1. Initialize the segmentation by applying some superpixel segmentation method (e.g., [1]). Tune the parameters of such a method so that the cells are broken into pieces to obtain an over-segmentation. Fig. 2(b) shows the initial segmentation of the image in Fig. 2(a). Let S denote the set of all segments in the current segmentation.

2. Provide training data. For each cell considered as training data, select a group of points in that cell such that the union of the segments (in the **initial segmentation**) each containing at least one such point can roughly represent

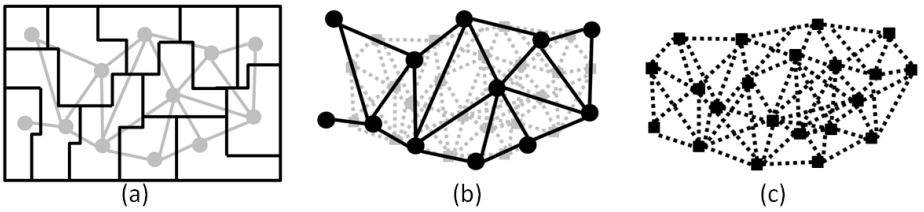

Fig. 3. Segmentation (a) and the corresponding graphs $SAdjG$ (b) and EDG (c)

well the region of that cell. We call the region of a cell thus constructed the training region (TR) of that cell (Fig. 2(d)). It is much easier to provide training data for the region of a cell in this way than delineating each boundary pixel of the cell. We will use the TR's to construct training examples of edges of the graph SAdjG, without making the user provide additional training data at each iteration. Note that the TR's, once constructed, will not change in the algorithm.

3. Construct the graph $SAdjG = (V_{SAdjG}, E_{SAdjG})$ based on S (Fig. 3(b)), where $V_{SAdjG} = \{s \mid s \in S\}$ and $E_{SAdjG} = \{(s_a, s_b) \mid s_a, s_b \in S$, and they share common border pixels$\}$.

4. Construct weighted training examples for the edges of the graph $SAdjG$. In iterations other than the initial one, segments obtained may cross the boundaries of the TR's, and thus we use the weights to compensate for the potentially noisy knowledge learned in such situations. For TR of each cell provided as training data, let $O = \{s \mid s \in S, s \cap TR \neq \emptyset\}$.

The set of '+' training examples (Fig. 2(f)) based on TR is defined as $PT = \{(s_1, s_2) \mid s_1, s_2 \in O, (s_1, s_2) \in E_{SAdjG}\}$. For each $(s_1, s_2) \in PT$, let

$$weight((s_1, s_2)) = \frac{\frac{(s_1 \cup s_2) \cap TR}{s_1 \cup s_2} + \frac{(s_1 \cup s_2) \cap TR}{TR}}{2}. \quad (1)$$

The set of '-' training examples (Fig. 2(f)) based on TR is defined as $NT = \{(s, n) \mid s \in O, n \notin O, (s, n) \in E_{SAdjG}\}$. For each $(s, n) \in NT$, let

$$weight((s, n)) = \frac{\frac{s \cap TR}{s} + \frac{s \cap TR}{TR}}{2}. \quad (2)$$

5. Build a binary classifier c for the edges of $SAdjG$, which represents our knowledge of what pairs of segments within or not within the same cell usually look like in the current iteration (or at the current scale). Note that we apply Random Forest here, but one can definitely use other classifiers.

6. Construct the edge dual graph $EDG = \{V_{EDG}, E_{EDG}\}$ (Fig. 3(c)), where $V_{EDG} = \{v \mid v = (s_1, s_2) \in E_{SAdjG}\}$ and $E_{EDG} = \{(v_a, v_b) \mid v_a = (s_1, s_2) \in E_{SAdjG}, v_b = (s_1, s_3) \in E_{SAdjG}\}$ (note that s_1 is shared by both v_a and v_b).

7. Apply binary MRF to EDG: Each vertex of EDG (correspondingly, each edge of $SAdjG$) is assigned a value of '1' or '0', representing whether the corresponding pair of segments is within the same cell. By merging the corresponding

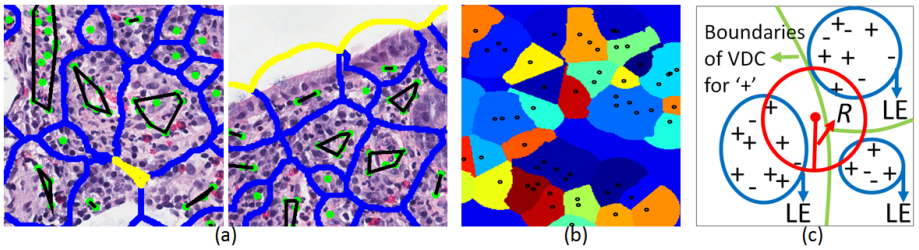

Fig. 4. (a)-(b) Examples of Voronoi diagrams of neutrophil clusters; (c) classifying the ambiguous cases using a VDC-based context

pair of segments for each edge of $SAdjG$ labeled as '1' into one segment, we obtain a new segmentation S (Fig. 2(c) shows the segmentation of another iteration). The unary cost of MRF comes directly form the probability output of the classifier c. We define the pairwise cost between the two labels '1' and '0' as

$$\begin{array}{ll} P(\text{'0'},\text{'0'}) = 0 & P(\text{'0'},\text{'1'}) = 0.5 \\ P(\text{'1'},\text{'0'}) = 0.5 & P(\text{'1'},\text{'1'}) = 1 \end{array} \tag{3}$$

Different from the typical usage of pairwise cost as smoothness constraint on labeling, we use it to penalize aggressive merging behavior of the segments (which may result in undesired under-segmentation).

8. Repeat steps 3 to 7, until the resulted segmentation is relatively stable. In practice, we find that 4 or 5 iterations are usually sufficient.

Feature Design. We construct features for the edges of $SAdjG$. They reflect either cues of each individual of the two corresponding segments (e.g., color, texture, compactness, etc), or cues exhibited between the two segments (e.g., color difference, texture difference, the number of common border pixels, etc).

2.2 Neutrophils Identification

Our goal of this stage is to identify neutrophils based on the segmentation results produced by the previous stage. Our main idea is to perform a two-phase process: First identify some "trustable" examples of neutrophils, and then construct a Voronoi diagram of clusters (VDC) based on the trustable neutrophils. We use VDC to extract a local context that can help resolve ambiguous cases of neutrophils around. Intuitively, we aim to capture quantitatively how pathologists identify neutrophils, especially the hard cases, using context information.

Trustable Examples. We first extract trustable examples for both neutrophils and non-neutrophils (i.e., other types of cells, or background tissue), by classifying the obtained segments and examining whether any of the two resulted probabilities, P(neutrophil) and P(non-neutrophil) (note that the sum of the two probabilities is one), is not below a certain threshold $pTrustableT$. The binary classifier (one can apply, e.g., Random Forest) is built based on all training annotations available for neutrophils and non-neutrophils.

VDC of Trustable Examples of Neutrophils. We observed that when pathologists are not sure whether a cell is a neutrophil based on only the appearance of this single cell, they try to compare this cell with some cells around on which they can relatively easily identify the types, to help make decision on the target cell. If neutrophils around form some clusters (i.e., in relatively high density), then such clusters usually attract more attention of the pathologists when they compare, since clusters of neutrophils often may provide appearance information on what a neutrophil generally looks like in the neighborhood.

Based on the observation above, first, we apply density-based clustering (DBC) [11] to the trustable examples of neutrophils, to capture the potential clustering behavior of neutrophils. In Fig. 4(a), the convex hull of each such cluster is colored in black, and trustable examples of neutrophils are in green.

Second, we compute a Voronoi diagram (VD) of these clusters (the VDC boundaries are colored in blue or yellow in Fig. 4(a)). Different from the common VD [7] whose sites are each a single point only, every site of our VDC can be a set of points (i.e., the center locations of the neutrophils in a cluster). The distance between a point p and a cluster $C = \{p_1, p_2, \ldots, p_m\}$ is defined as

$$D(p, C) = \sum_{i=1}^{m} F(p, p_i), \qquad F(p, p_i) = \begin{cases} \frac{w(p_i)}{d(p,p_i)^2}, & \text{if } d(p, p_i) \leq dist_T \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $F(p, p_i)$ is a function measuring the "influence" of a point $p_i \in C$ on p, which depends on both the weight $w(p_i)$ (measuring the importance of p_i, and set as 1 here) and the Euclidean distance $d(p, p_i)$. We assume no influence is from p_i if $d(p, p_i)$ is larger than a threshold $dist_T$ (set as 160 pixels). Essentially, $D(p, C)$ measures the amount of collective influence of a cluster C on any point p. Thus, suppose there are k clusters, then a point p belongs to the (geometric) VD cell C_{VD} (i.e., not a biological cell, see Fig. 4(b)) associated with cluster C, if $D(p, C) = max\{D(p, C_i) \mid i = 1, 2, ..., k\}$ (i.e., with the biggest influence). Note that since $F(p, p_i)$ is truncated by a Euclidean distance threshold $dist_T$, there may exist points far away from all clusters that do not lie in any VD cell (e.g., in Fig. 4(a), points in regions bounded by only yellow curves).

Third, as shown in Fig. 4(c), for each VD cell C_{VD}, we build a binary classifier (such as Random Forest) such that its '+' and '-' training examples are trustable examples for respectively neutrophils and non-neutrophils whose center locations lie in C_{VD}. We call this classifier a *Local Expert* (LE) for C_{VD}, which represents our knowledge of what neutrophils and non-neutrophils look like around the corresponding cluster of neutrophils of C_{VD}.

Finally, for each ambiguous segment s, we define its *context* as the set of all VD cells overlapped with the circle with a radius R (set as 60 pixels; see Fig. 4(c)) centered at the center of s (note that it is possible for an ambiguous segment to have an empty context, and we take the segment as non-neutrophil in such cases). The probability outputs of the segment s is the average of the probability outputs by LE of each VD cell in its *context*, weighted by normalized distance $D(p, C)$ between the center point p of the segment s and the corresponding

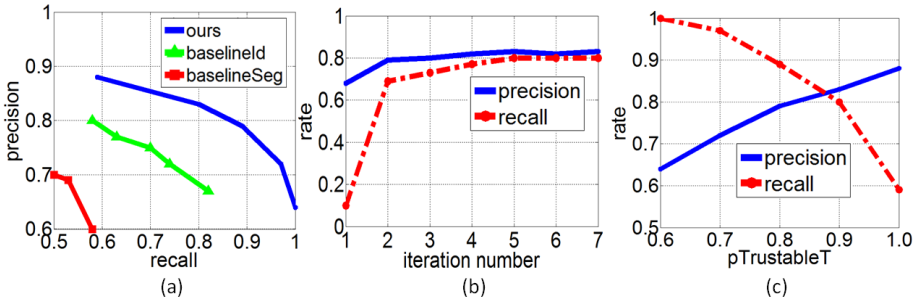

Fig. 5. (a) Precision and recall curves of identifying neutrophils by different methods; (b)-(c) analysis of influence of each algorithm stage

cluster C of each such VD cell. We take s as a neutrophil if such probability output for neutrophil is above a certain threshold p_T (set as 0.5).

3 Experiments and Evaluation

Clinical Data Sets. We collected images scanned at 40x magnification from patients suspected of having inflammatory bowl diseases. Pathologists annotated 339 neutrophils (not a small set, considering neutrophils do not infiltrate as prevalently as some other immune cells), and based just on appearance of each individual cell, they were highly/moderately confident at marking only about half of them (nevertheless, our method is blind to such confidence scales because such ambiguity to pathologists is finally resolved by considering the context).

Performance Evaluation. Since no direct effective method was known before (we have actually tried out related methods for blood smear on our images, but the results are quite poor to be used as a fair comparison), to show the necessity and effectiveness of both our segmentation and identification stage, we compare our method with two baselines: *baselineSeg* and *baselineId*, constructed by us in a structured manner. *BaselineSeg* applies superpixel segmentation [1] followed by the same VDC identification process as in our method. *BaselineId* applies the same segmentation process as in our method followed by directly thresholding the classification results for segments, without using VDC.

To measure the performance, we compute $precision = \frac{TP}{TP+FP}$ and $recall = \frac{TP}{TP+FN}$ for neutrophils, using 10-fold cross validation on ground truth data. Fig. 5(a) presents the precision and recall curves, obtained by varying key parameters (specifically, number of superpixels for *baselineSeg*, thresholding value for *baselineId*, and $pTrustableT$ for our method) respectively for each method. One can see that both our segmentation and identification processes help improve the overall performance significantly compared with the baselines.

Analysis of Influence of Each Algorithm Stage. We examine the influence of each stage of our method on the final performance, by varying key parameters

to which final results are sensitive. Fig. 5(b) shows the performance for different numbers of iterations. One can see that at early iterations, bad segmentation leads to bad identification; at later iterations, identification becomes better and stable as segmentation becomes better and stable. Fig. 5(c) shows the performance by varying the $pTrustableT$, which results in changes of the clustering settings of neutrophils in VDC. One can see that the more trustable the examples for neutrophils are, the higher the precision is; but the recall decreases, since trustable examples above the threshold become fewer.

4 Conclusions

We presented a new and effective method for identifying neutrophils in H&E staining histology tissue images, based on iterative edge labeling and VDC. This lays a crucial foundation for successful acute inflammation analysis.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. TPAMI 34(11), 2274–2282 (2012)
2. Gurcan, M., Boucheron, L., Can, A., Madabhushi, A., Rajpoot, N., Yener, B.: Hitopathological image analysis: A review. IEEE Rev. Biomed. Eng. 2, 147–171 (2009)
3. Hiremath, P., Bannigidad, P., Geeta, S.: Automated identification and classification of white blood cells (leukocytes) in digital microscopic images. IJCA, Special Issue on RTIPPR (2), 59–63 (2010)
4. Huang, D.C., Hung, K.D., Chan, Y.K.: A computer assisted method for leukocyte nucleus segmentation and recognition in blood smear images. J. Syst. Software 85, 2104–2118 (2012)
5. Jatti, A., Urs, V.M.: Review paper of conventional analysis of cell smear under a microscope. IJIRD 3(2) (2013)
6. Naini, B.V., Cortina, G.: A histopathologic scoring system as a tool for standardized reporting of chronic (ileo) colitis and independent risk assessment for inflammatory bowel disease. Human Pathology 43, 2187–2196 (2012)
7. Okabe, A., Boots, B., Sugihara, K., Chiu, S.N.: Spatial Tessellations: Concepts and Applications of Voronoi Diagrams. Wiley (2000)
8. Saraswat, M., Arya, K.: Leukocyte classification in skin tissue images. In: Bansal, J.C., Singh, P.K., Deep, K., Pant, M., Nagar, A.K. (eds.) BIC-TA 2013. AISC, vol. 201, pp. 65–73. Springer, Heidelberg (2013)
9. Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., Rother, C.: A comparative study of energy minimization methods for Markov random fields. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 16–29. Springer, Heidelberg (2006)
10. Theera-Umpon, N., Dhompongsa, S.: Morphological granulometric features of nucleus in automatic bone marrow white blood cell classification. IEEE Trans. Inf. Technol. Biomed. 11, 353–359 (2007)
11. Xu, B., Chen, D.Z.: Density-based data clustering algorithms for lower dimensions using space-filling curves. In: Zhou, Z.-H., Li, H., Yang, Q. (eds.) PAKDD 2007. LNCS (LNAI), vol. 4426, pp. 997–1005. Springer, Heidelberg (2007)