Published in final edited form as:

IEEE Trans Biomed Eng. 2010 October ; 57(10): 2613–2616. doi:10.1109/TBME.2010.2055058.

Computer-aided Detection of Centroblasts for Follicular Lymphoma Grading using Adaptive Likelihood based Cell Segmentation

Olcay Sertel[Student Member, IEEE],

Dept. of Biomedical Informatics, and the Dept. of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA

Gerard Lozanski,

Dept. of Pathology, The Ohio State University, Columbus, OH 43210 USA

Arwa Shana'ah, and

Dept. of Pathology, The Ohio State University, Columbus, OH 43210 USA

Metin N. Gurcan[Senior Member, IEEE]

Dept. of Biomedical Informatics, The Ohio State University, Columbus, OH 43210 USA

Gerard Lozanski: Gerard.Lozański@osumc.edu; Arwa Shana'ah: Arwa.Shanaah@osumc.edu; Metin N. Gurcan: Metin.Gurcan@osumc.edu

Abstract

Follicular lymphoma (FL) is one of the most common lymphoid malignancies in the western world. FL has a variable clinical course and important clinical treatment decisions for FL patients are based on histological grading, which is done by manual counting of large malignant cells called centroblasts (CB) in ten standard microscopic high power fields from H&E-stained tissue sections. This method is tedious and subjective; as a result suffers from considerable inter- and intra-reader variability even when used by expert pathologists. In this study, we present a computer-aided detection system for automated identification of CB cells from H&E-stained FL tissue samples. The proposed system uses a *unitone conversion* to obtain a single channel image that has the highest contrast. From the resulting image, which has a bi-modal distribution due to the H&E-stain, a cell-likelihood image is generated. Finally, a two-step CB detection procedure is applied. In the first step, we reduce evident non-CB cells based on size and shape. In the second step CB detection is further refined by learning and utilizing the texture distribution of non-CB cells. We evaluated the proposed approach on 100 region of interest images extracted from ten distinct tissue samples and obtained a promising 80.7% detection accuracy.

Index Terms

Biomedical image analysis; Cell segmentation; Follicular lymphoma; H&E stained tissue; Histology

I. INTRODUCTION

The last few years witnessed remarkable increase in research studies on digital pathology applications. This is mostly due to the recent advances in high-throughput whole-slide tissue scanning technology, which also allows the application of image analysis. Image analysis

can now be utilized to evaluate tissue samples as a second reader in order to supplement the decision making mechanism. However, there are challenging problems that need to be addressed before these systems can be reliably utilized in practice. These problems include: cellular architecture variations, image noise, artifacts and distortions due to tissue fixation, slide preparation and staining processes. One way to overcome these problems of variation is by developing adaptive approaches.

Follicular lymphoma (FL) is one of the most common lymphoid malignancies in the western world with a highly variable clinical course [1]. Currently, clinical decisions are guided by histological grading of the tumor. Recommended by the World Health Organization, histological grading of FL is based on the average number of cancerous cells, namely centroblasts (CBs), per standard 40 $\times$ microscopic high power field in representative malignant follicles. However, visual qualitative assessment is difficult, time consuming and subject to high inter- and intra-reader variability [2]. Fig. 1 shows sample region of interest image regions captured from different tissue samples and the inter-reader variability between five different pathologists.

Promising image analysis systems have been proposed in the last few years to classify tissue subtypes associated with various grades of cancer [3–4]. One of the problems is the segmentation of distinct cytological components. Due to the inherent discrimination provided by the staining, robust feature space analysis techniques (e.g., k-means, expectation maximization) in the color space is successfully applied in recent approaches [5–6]. Statistical texture operators based on co-occurrence and run-length matrices have been applied to achieve tissue classification [7]. There are also applications that target the detection of cells rather than global scene segmentation. In [8] a concavity based ellipse fitting method has been proposed, whereas in [9] a supervised classification approach is utilized that requires considerable amount of preprocessing to avoid staining variations.

II. IDENTIFICATION OF CENTROBLASTS USING IMAGE ANALYSIS

Fig. 2 shows the flowchart of the proposed image analysis system, which mainly consists of three stages: segmentation of cellular components, indentifying individual cells and CB detection.

A. Segmentation of cellular components

H&E stain colors nuclear and cytoplasmic regions to hues of blue, and purple while protein rich collagen structures such as extra-cellular material is colored into hues of pink. Red blood cells (RBC) are intensely red and the background remains white in color. As can be seen from the images given in Fig. 1, due to the application of chemical dyes, these images have a considerably limited dynamic range in the color spectrum. We first convert the input images in the RGB color space onto a one-dimensional *unitone* image using the principal components analysis (PCA) [10]. The unitone image is computed by projecting the RGB image onto the first principle component associated with the highest variance; hence the resulting unitone image has the highest contrast. Due to the slide preparation process, there is a considerable variation between images acquired from different slides. By using PCA, we also avoid the global variations between different tissue slides, because the principle direction of variance is independent from global variations. The resulting unitone image is normalized to the range $I_u \rightarrow [0, 1]$.

The next step is the segmentation of individual cells. Modeling the distribution of both cellular and extra-cellular components with a Gaussian mixture model, we estimate the mixture parameters using the expectation maximization (EM) algorithm [10]. The unknown parameters are $\theta = \{\mu_H, \sigma_H \text{ and } \mu_E, \sigma_E\}$, where μ_H, σ_H and μ_E, σ_E are the mean and variance

of the distributions associated with cellular and extra-cellular structures, respectively. EM is an iterative method, which starts with a random initialization. It consists of two steps: expectation (Eq. 1), which computes the likelihood with respect to the current estimates, and maximization (Eq. 2), which maximizes the expected log likelihood:

$$Q(\Theta | \Theta^{(t)}) = E_{Z|x, \Theta^{(t)}} (\log L(\Theta; x, Z)), \quad (1)$$

$$\Theta^{(t+1)} = \arg \max_{\Theta} Q(\Theta | \Theta^{(t)}) \quad (2)$$

where $x = \{x_1, \dots, x_n\}$ are the observations (i.e., the pixel values) and $Z = \{z_1, z_2\}$ are the latent variables that determine the component from which the observation originates. Once the underlying distributions are estimated we compute the posterior probability for each pixel as follows:

$$p(\omega_i | x) = \frac{p(x | \omega_i) P(\omega_i)}{\sum_j p(x | \omega_j) P(\omega_j)}, \quad (3)$$

where $i = \{c, ec\}$ indicate cellular or extra-cellular components, and $p(x|\omega_i)$ is normally distributed as $p(x|\omega_i) \approx N(\mu_i, \sigma_i)$. Fig. 3 (a) and (b) show a sample unitone converted image and its histogram with estimated distributions of cellular and extra-cellular components are overlaid, respectively.

B. Identifying individual cells

Using the posterior probabilities and the estimated parameters of the unitone values, we construct a cellular likelihood image. We use a sigmoid function (see Fig. 3), which can be controlled with two parameters as follows:

$$f_{Cell_LK}(x) = 1 / (1 + e^{-\alpha(x-\beta)}), \quad (4)$$

where α controls the smoothness of the s-shaped likelihood curve and β indicates the offset where $f_{Cell_LK}(\beta) = 0.5$. These parameters are tuned adaptively for each image such that $\beta = \mu_H + 2\sigma_H$ and $\alpha = -50(\mu_E - \mu_H)$, where μ_H, σ_H are the estimated parameters of the distribution of the unitone values associated with cellular components, and $(\mu_E - \mu_H)$ is proportional to how well these distributions are separated from each other, so the exponential decay of the cell likelihood is adjusted accordingly. The cell likelihood image is smoothed using a 9×9 Gaussian window to enforce spatial connectivity.

After constructing the cellular likelihood, we apply a locally adaptive thresholding step to obtain the binary representation of cellular structures such that the threshold value is computed differently for each pixel based on the distribution of likelihood values within its neighborhood as follows:

$$\tau(p, q) = \frac{1}{N^2} \sum_{i=-N_w/2}^{N_w/2} \sum_{j=-N_w/2}^{N_w/2} I_{CellLK}(i, j), \quad (5)$$

where p, q are the row and column indices, i, j are the offset indices within the local neighborhood, and $N_W=15$ defines the neighborhood window size and I_{CellLK} is the cell likelihood image (see Fig. 3(c) and (d)). The thresholding step is followed by morphological post-processing operations consisting of filling small holes within cells, opening and removing small sized components.

One of the generic problems in microscopic image analysis applications is the separation of spatially clustered cells that are touching or even overlapping each other. The watershed transform based on shape topology is a common approach to deal with this problem [11]. However, it is not suitable in cases where more than a few cells are clustered. Using the fast radial symmetry transform [12], we propose a spatial voting based approach. Accordingly, for a radius value of r , each pixel p contributes a spatial voting matrix proportional to its gradient magnitude $\|g(p)\|$ at the spatial location $s_r(p)$. $s_r(p)$ is calculated based on the gradient direction:

$$s_r(p) = s_r(p) + \text{round} \left(\frac{g(p)}{\|g(p)\|} r \right) \quad (6)$$

where $g(p)$ is the gradient of pixel p and s_r denotes the spatial voting matrix computed for the radius value r . We compute the radial symmetry votes for a range of radii values $r=\{7,9, \dots, 21\}$ that covers the typical range of cell sizes. The corresponding voting spaces for different radii values are merged based on maximum vote. Finally, individual locations of cells are computed using non-maxima suppression. Fig. 4 shows a sample image region, its cell likelihood image and the binary segmentation of cells after locally adaptive thresholding. As can be seen from Fig. 4(c), a group of touching cells is identified as a single component. Fig. 5 shows the cell separation using the radial symmetry based voting approach.

C. CB Detection

In FL tissue, the malignant follicles are composed of varying proportions of centrocytes, small often cleaved cells with coarse chromatin and scant cytoplasm and centroblasts (CBs), large cells with open vesicular chromatin and one to multiple nucleoli that are frequently associated with nuclear membrane. In low grade FL, centrocytes predominate while in high grade FL, CBs increase in number. Apart from the differences in size and shape; due to the complex spatial organization of sub-cellular components, cell texture provides an important perceptual clue that allows us to differentiate CBs from centrocytes.

After the segmentation of individual cells, we use a two-step procedure to identify CBs. The first step is based on eliminating evident non-CB cells using the size and eccentricity criterion as follows:

$$cell_i = \begin{cases} CB, & a_i > \mu_a + \sigma_a \text{ and } e_i < 0.85 \\ notCB, & otherwise \end{cases}, \quad (7)$$

where a_i is the area and e_i is the eccentricity of i^{th} cell defined as the ratio of the major to minor axis' length of the best fitted ellipse; μ_a and σ_a are mean and standard deviation of cell area. As we designed this initial CB detection step to be very sensitive, it provides a relatively high number of false positives. However, we can still identify evident non-CB cells and utilize their texture distributions in order to further refine our detection in the subsequent step.

We used the *gray-level run length matrix* (GLRLM) method to obtain a statistical measure of the spatial organization of intensity variations within each cell. GLRLM is a two dimensional matrix from which higher order statistics can be derived to represent image texture [13]. Each entry $c(g, l | \theta)$ in GLRLM represents the number of occurrences of run-length l associated with pixels having the gray level of g along the direction θ . From the GLRLM representation with a 16 gray-level quantization, we compute 10 features averaged over four different directions $\theta = \{0, \pi/4, \pi/2, 3\pi/4\}$. These features cover a wide spectrum of patterns of different scales and texture information.

Using GLRLM features, each cell is represented as a vector in the resulting feature space. Then, we model the distribution of non-CB cells identified in the initial detection step as a Gaussian distribution (i.e., $\approx N(\mu_{non-CB}, \Sigma_{non-CB})$) and parameters are estimated using maximum likelihood as follows:

$$\mu_{notCB} = \frac{1}{n} \sum_{i=1}^n f_i, \Sigma_{notCB} = \frac{1}{n} \sum_{i=1}^n (f_i - \mu_{notCB})^2, \quad (8)$$

where n is the number of non-CB cells initially detected based on size and shape constraints, and f_i is the corresponding feature vector of the i^{th} cell. Fig. 6 shows the scatter plot of GLRLM texture features in a two-dimensional feature space obtained by applying PCA and keeping the first two dimensions associated with the largest variance, as well as the estimated distribution of non-CB cells plotted in elliptical contours associated with 0.5σ , 1σ , 2σ and 3σ , respectively. As can be seen in Fig. 6, initially detected CB cells have unique texture features clustered densely, while the features of non-CB cells clearly diverge from this dense cluster. Based on this observation, we identify CB cells as follows:

$$f_i = \begin{cases} CB, & P(f_i | \mu_{notCB}, \Sigma_{notCB}) > P(\mu_{notCB} + 3U\Lambda^{-1/2}) \\ notCB, & otherwise \end{cases} \quad (9)$$

where $P(\mu_{notCB} + 3U\Lambda^{-1/2})$ corresponds to the likelihood value computed as the value at the iso-contour of $N(\mu_{notCB}, \Sigma_{notCB})$, where covariance matrix can be decomposed into equidensity contours, i.e., iso-contours, using the eigenvalue decomposition $\Sigma = U\Lambda U^T$ and $\Lambda^{-1/2}$ corresponds to semi-major axis lengths of the associated iso-contours.

III. EXPERIMENTAL EVALUATION

We evaluate the proposed approach over a dataset consisting of 100 ROI images from ten different whole-slide FL tissue samples digitized at 40 $\times$ magnification using an Aperio Scope XT scanner (Aperio, San Diego, CA). Each ROI image has a digital resolution of 1353 $\times$ 2168 pixels and includes approximately 2500 cells. The ground-truth information for the CB cell locations is generated by five expert board-certified hematopathologists. Since there is a considerable amount of variation between different readers, we construct the ground-truth from CBs that are marked by at least two pathologists. Based on this ground-truth, the average CB detection accuracy of an expert hematopathologist is 65% with 5.1 false positive (FP) CBs per ROI image on average.

Table I summarizes the evaluation of the proposed computerized CB detection system. For each tissue slide, we report the average CB detection count per ground-truth CB count (CB), average CB detection accuracy (Acc), and the FP detection count, at the initial and the final CB detection stages, respectively. As can be seen from these results, the initial CB detection step provides 90% accuracy with an average FP count of 200. In the final refinement stage,

the FP count is reduced by 85% down to 30 FPs per ROI image on average, compromising roughly 10% accuracy in CB detection accuracy. The results show that the accuracy of the computerized system is higher than the accuracy of human readers; however, it also generates a relatively higher number of false positives (~30 FPs/ROI).

IV. CONCLUSIONS

We are developing a fully automated computerized system for the detection of CB cells from digitized H&E-stained FL tissue samples. The proposed system demonstrates the feasibility of robust segmentation of individual cells in the tissue image by a novel adaptive likelihood approach based on both global and local image characteristics. CB detection is carried out in a two step procedure where we first identify evident non-CB cells based on the size and shape constraints, and then model their distribution in the feature space to refine the final CB detection step. Experimental results provide promising performance of this system, which can supplement the decision making mechanism in order to improve current agreement among different human readers. In our future work, we will develop complementary approaches to improve the current performance of the proposed approach.

Acknowledgments

This work was supported in part by award number R01CA134451 from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute, or the National Institutes of Health.

REFERENCES

1. Jaffe, ES., et al. Pathology and genetics: Tumours of haematopoietic and lymphoid tissues. IRAC Press; 2001.
2. Martinez AE, et al. Grading of follicular lymphoma: Comparison of routine histology. Arch of Pathol Lab Med. 2007; vol. 131:1084–1088. [PubMed: 17616995]
3. Tabesh A, et al. Multifeature prostate cancer diagnosis and gleason grading of histological images. IEEE Trans. on Med. Im. 2007; vol. 26:1366–1377.
4. Kong J, et al. Computer-aided evaluation of neuroblastoma on whole-slide histology images. Pattern Recog. 2009; vol. 42:1080–1092.
5. Sertel O, et al. Histopathological image analysis using model-based intermediate representations and color texture: follicular lymphoma grading. J. Sig. Proc. Systems. 2009; vol. 55(1):169–183.
6. Fatakdawala, H., et al. EM driven geodesic active contour with overlap resolution: application to lymphocyte segmentation on breast cancer histology; IEEE Int. Conf. on BIBE; 2010.
7. Gurcan MN, et al. Histopathological Image Analysis: A Review. IEEE Reviews in Biomedical Engineering. 2009; vol.2:147–171. [PubMed: 20671804]
8. Kothari S, Chaudry Q, Wang MD. Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques. Proc. IEEE ISBI. 2009:795–798.
9. Cosatto, E.; Miller, M.; Graf, HP.; Meyer, JS. Grading Nuclear Pleomorphism on Histological Micrographs; IEEE ICPR; 2008.
10. Duda, RO.; Hart, PE.; Stork, DG. Pattern classification. Wiley-Interscience; 2001.
11. Meyer F. Topographic distance and watershed lines. Signal Processing. 1994; vol. 38:113–125.
12. Loy G, Zelinsky A. Fast radial symmetry for point of interest detection. IEEE Trans. on PAMI. 2003; vol. 25(8):959–973.
13. Galloway MM. Textural analysis using gray level run lengths. Comput. Graphics Image Process. 1975; vol. 4:172–179.

Fig. 1.
Sample region of interest images (512×512 pixels) from H&E stained FL tissue samples viewed at 40× magnification. CB cells are identified by five different pathologists indicated by circles in different colors.

Fig. 2.
Block diagram of the proposed CB detection system.

Fig. 3.

(a) A sample ROI image of 300×300 pixels converted to unitone, and (b–e) the subsequent intermediate results showing the progress at different steps of CB detection process. (f) Final CB detection result is given together with the ground-truth CB markings by different pathologists indicated circles with different colors.

Fig. 4.
A sample image region demonstrating cell segmentation and the resulting cluster of cells.

Fig. 5.

(a)–(d) show the resulting voting spaces for a range of radii, $r = \{19, 15, 11, 7\}$; (e)–(h) show the corresponding locations (green circles) of segmented individual cells with varying radii.

Fig. 6.

Scatter plot of cell texture features. CB and non-CB cells are defined based on the initial detection using the size and shape information, and further refined based on estimated distribution $P(f|\omega_{\text{notCB}})$ of non-CB cells.

TABLE I

Evaluation of the proposed CB detection system

	Initial CB Detection			Final CB Detection		
	CB	Acc (%)	FP	CB	Acc (%)	FP
Slide 1	10/11	92.3	218	9/11	81.3	41
Slide 2	8/9	85.9	208	7/9	73.7	21
Slide 3	10/11	90.8	210	9/11	82.7	36
Slide 4	2/2	94.1	172	2/2	88.2	34
Slide 5	6/6	96.7	220	5/6	85.3	42
Slide 6	10/12	81.8	187	9/12	80.3	30
Slide 7	5/6	89.7	192	4/6	75.5	16
Slide 8	3/3	90.1	200	2/3	73.7	18
Slide 9	7/7	98.1	211	6/7	88.4	27
Slide 10	9/11	82.8	180	8/11	77.5	43
Average:		90.2	200		80.7	30