

An Integrated Framework for Automatic Ki-67 Scoring in Pancreatic Neuroendocrine Tumor

Fuyong Xing^{1,2}, Hai Su^{1,2}, and Lin Yang^{1,2}

¹ Division of Biomedical Informatics, Department of Biostatistics

² Department of Computer Science, University of Kentucky, KY 40506, USA

Abstract. The Ki-67 labeling index is a valid and important biomarker to gauge neuroendocrine tumor cell progression. Automatic Ki-67 assessment is very challenging due to complex variations of cell characteristics. In this paper, we propose an integrated learning-based framework for accurate Ki-67 scoring in pancreatic neuroendocrine tumor. The main contributions of our method are: a novel and robust cell detection algorithm is designed to localize both tumor and non-tumor cells; a repulsive deformable model is applied to correct touching cell segmentation; a two stage learning-based scheme combining cellular features and regional structure information is proposed to differentiate tumor from non-tumor cells (such as lymphocytes); an integrated automatic framework is developed to accurately assess the Ki-67 labeling index. The proposed method has been extensively evaluated on 101 tissue microarray (TMA) whole discs, and the cell detection performance is comparable to manual annotations. The automatic Ki-67 score is very accurate compared with pathologists' estimation.

1 Introduction

Pancreatic neuroendocrine tumor (NET) cancer is one of the most common cancers worldwide. The Ki-67 labeling index, defined as the ratio between the numbers of immunopositive tumor cells and all tumor cells, has been considered as a valid biomarker to evaluate tumor cell progression and predict therapy responses. Manual Ki-67 assessment is subject to a low throughput processing rate and pathologist-dependent bias. Automatic Ki-67 assessment can provide more objective, high throughput, and reproducible results [6,10]. Automated cell detection can also provide access to computer-aided Ki-67 scoring. Parvin *et al.* [13] proposed an iterative radial voting algorithm based on oriented kernels to localize cell nuclei, in which the voting direction and areas are dynamically updated within each consecutive iteration. This algorithm has outstanding noise immunity and scale invariance. A computationally efficient single-pass voting for cell detection is reported in [15], which applies mean shift clustering instead of iterative voting to seed localization. Other methods [9,11,2,7,8] are also proposed for touching cell detection and segmentation. However, none of these methods addresses the automatic Ki-67 counting problem, which requires an accurate differentiation between tumor and non-tumor cells.

In this paper, we propose a novel integrated learning-based algorithm for automatic scoring of pancreatic NET with Ki-67 staining. In order to accurately detect cells in dense clusters, a robust and efficient region-based hierarchical voting algorithm is proposed to detect cell seeds (geometric centers). These seeds will be used to initialize a repulsive deformable model to extract touching cell boundaries with known object topology constraints. Next, a two stage learning-based scheme is employed to differentiate tumor from non-tumor cells. The algorithm combines both the cellular features and regional structure information. The Ki-67 labeling index is finally calculated using a color histogram to separate immunopositive (brown cells in Ki-67 staining) and immunonegative (blue cells in Ki-67 staining) tumor cells.

2 Methodology

2.1 Hierarchical Voting-Based Seed Detection

Robust cell detection is achieved by finding the geometric centers (seeds) of the cells. Let $T(x, y)$ denote the original image, and $\nabla T(x, y)$ be the gradient. For those pixels with relatively large magnitude $\|\nabla T(x, y)\|$, single-pass voting [15] defines a cone-shape voting area A with vertex at (x, y) and votes along the negative gradient direction: $\frac{-\nabla T(x, y)}{\|\nabla T(x, y)\|} = -(\cos(\theta(x, y)), \sin(\theta(x, y)))$, where θ represents the angle of the gradient direction with respect to x axis. A voting image $V(x, y)$ is initialized as zeros and then updated by weighting the gradient magnitude with a Gaussian kernel $g(m, n, \mu_x, \mu_y, \sigma)$:

$$V(x, y) = V(x, y) + \sum_{(m, n) \in A} \|\nabla T(x, y)\| g(m, n, \mu_x, \mu_y, \sigma), \quad (1)$$

where the voting area A is defined by the radial range (r_{min}, r_{max}) and angular range Δ . The isotropic Gaussian kernel is parametrically defined with $(\mu_x, \mu_y) = (x + \frac{(r_{max} - r_{min}) \cos \theta}{2}, y - \frac{(r_{max} - r_{min}) \sin \theta}{2})$ and scalar σ . After the voting map is generated, mean shift [4] is employed to calculate the final seed for each individual cell.

Single-pass voting is computationally efficient, but it is not able to efficiently handle cell size and shape variations. Considering these challenges, in our algorithm, we introduce a region-based hierarchical voting in the distance transform map. A Gaussian pyramid is applied to both the voting procedure and the mean shift clustering step. The hierarchical voting is formulated as:

$$V_{RH}(x, y) = \sum_{l=0}^L \sum_{(m, n) \in S} I((x, y) \in A_l(m, n)) M_l(x, y) g(m, n, \mu_x, \mu_y, \sigma), \quad (2)$$

where $V_{RH}(x, y)$ is the confidence map, S represents the set of all voting pixels, $A_l(m, n)$ denotes the cone-shape voting area with vertex (m, n) at layer l , $M_l(x, y)$ represents the Euclidean distance transform map at layer l , and

$I(\mathbf{x}) = I((x, y) \in A_l(m, n))$ is the indicator function. The pixels with higher $M_l(x, y)$ values near the geometric center of a cell will enhance their contributions in Equation (2). For each pixel (x, y) , Equation (2) provides a weighted sum of all the voting values created by its neighboring pixels whose voting areas contain (x, y) , instead of only those created by its own. In comparison with single-pass voting, hierarchical voting is more robust with respect to the detection of cells with relatively large size variations.

2.2 Repulsive Deformable Model

The proposed cell detection algorithm indiscriminately detects both tumor and non-tumor cells, and therefore the results cannot be directly used to calculate the Ki-67 scoring index. Differentiation between tumor and non-tumor cells is the critical step for automatic Ki-67 scoring. In order to extract discriminative morphological features to separate tumor from non-tumor cells, accurate cellular segmentation is a prerequisite. This is challenging due to the complex color and intensity variations exhibited inside cells, especially within touching cell clumps. Based on the results obtained from Section 2.1, we propose to segment each individual cell using an improved deformable model. Motivated by Zimmer and Olivo-Marin's work [18], we introduce a contour-based repulsive term into the deformable model [3] to prevent evolving contours from crossing and merging with one another. In [3], the original pressure force is designed to deform contour $v(s)$ until the internal $F^{int}(v)$ and external $F^{ext}(v)$ forces achieve a balance:

$$F^{int}(v) = \alpha v''(s) - \beta v''''(s), \quad F^{ext}(v) = \gamma \mathbf{n}(s) - \lambda \frac{\nabla E_{ext}(v(s))}{\|E_{ext}(v(s))\|}, \quad (3)$$

where $\mathbf{n}(s)$ together with weight γ represents the pressure force, and $\nabla E_{ext}(v(s))$ denotes the image force where $E_{ext}(v(s)) = -\|\nabla T(x(s), y(s))\|^2$. α , β , and λ are weight parameters. Without a repulsive term, active contours guided by (3) will move independently and may cross with one another within touching cell clumps. In order to prevent contour overlapping, we introduce an external force:

$$F_R^{ext}(v_i) = \gamma \mathbf{n}_i(s) - \lambda \frac{\nabla E_{ext}(v_i(s))}{\|E_{ext}(v_i(s))\|} + \omega \sum_{j=1, j \neq i}^N \int_0^1 d_{ij}^{-2}(s, t) \mathbf{n}_j(t) dt, \quad (4)$$

where N is the number of cells, $d_{ij}(s, t) = \|v_i(s) - v_j(t)\|_2$ is the Euclidean distance between contour $v_i(s)$ and $v_j(t)$. The last term (with parameter ω) in Equation (4) models the repulsion schema: as the contours move closer ($d_{ij}^{-2}(s, t)$ becomes larger), they receive stronger repulsive forces from each other until stop evolving. Using the detected seed as initialization (Section 2.1), this repulsive deformable model can handle touching cells effectively. Compared with [18] that introduces an area-based penalty term to avoid contour overlapping, our model is computationally efficient and requires less memory. Meanwhile, as a parametric model, the proposed repulsive deformable model can preserve known object topology, and therefore each contour represents one cell without splitting or merging. This topology preserving property is significantly different from the widely used geometric repulsive deformable methods [17,12,5].

2.3 Two Stage Learning-Based Classification

In order to calculate the final Ki-67 labeling index, a two stage learning-based scheme combining cellular features and regional structure information is designed to differentiate tumor from non-tumor cells.

Stage I: Based on the results of cellular segmentation, a SVM classifier is trained to predict the probabilities of segmented cells (tumor or non-tumor cells) using the following cellular features: 1) Geometric descriptors: area, perimeter, circularity, axis ratio (length ratio between estimated major and minor axes), solidity; 2) Color intensity: mean, standard deviation (σ), smoothness ($1 - 1/(1 + \sigma^2)$), skewness, kurtosis, entropy, contrast, correlation, homogeneity; 3) Cell shapes with Fourier descriptor. In total we have extracted $5 + 9 \times 3 + 80 = 112$ features, where 3 represents R, G, and B color channels, and 80 denotes the first 20 harmonics (each corresponds to 4 coefficients) that are chosen in the Elliptical Fourier transformation. A sparse representation model is used to select a set of most discriminative features by solving:

$$\min_b \|Db - a\|_2^2, \text{ s.t. } \|b\|_1 \leq \eta, b \geq 0, \quad (5)$$

where $D \in \mathbb{R}^{(N^+ + N^-) \times W}$ represents the features extracted from N^+ tumor and N^- non-tumor cells, and $W = 112$ denotes the original dimension of the feature vector. The η is a parameter controlling the sparsity of b . The binary vector $a \in \mathbb{R}^{(N^+ + N^-) \times 1}$ represents the labels of cells used for training in D : $a_i = +1$ for tumor and $a_i = -1$ for non-tumor. Due to the l_1 norm constraint, the solution $b^* \in \mathbb{R}^{W \times 1}$ to (5) is sparse with nonzero elements corresponding to the selected discriminative features. Based on b^* with L nonzero elements, we can project all the features onto a low-dimensional, discriminative subspace. A SVM classifier is learned to predict the cell category in the transformed feature space. All the cells with low probabilities (usually representing typical non-tumor cells) will be removed before entering Stage II such that the second classifier would be able to focus on a reduced dataset containing more difficult cases.

Stage II: In Stage I, only cellular features are considered. The classification accuracy will be improved by introducing the local structure information. This image structural pattern can be described with texture, and modeled with texton [16] feature. A multiple scale Schmid filter bank [16] is used for image filtering:

$$F(r, \sigma, \tau) = F_0(\sigma, \tau) + \cos\left(\frac{\pi r \tau}{\sigma}\right) e^{-\frac{r^2}{2\sigma^2}}, \quad (6)$$

where τ is the number of cycles of the harmonic function within the Gaussian envelop of the filter and $r = \sqrt{x^2 + y^2}$. A texton library is constructed using the training pancreatic NET TMA specimens. For computational efficiency, the integral histogram [14] is utilized to calculate the multiscale windowed texton histogram. Finally, the logistic boosting is employed to calculate the probability map using the multiscale texton histograms as features.

Using the texture classification-based probability map, each individual cell will obtain a score to evaluate its probability belonging to tumor or non-tumor cells.

In addition, the ratio between the probability of one cell and the probability average for all its neighbouring cells provides a measurement of cell category distribution. Based on these observations, the mean/standard deviation of pixel probabilities in each cell, and the percentage of probability summation of one cell over the probability average for all cells in its local region are calculated. These statistical features are concatenated with the previously predicted cellular probabilities in Stage I to train a second SVM classifier. The output will produce the final labels to differentiate tumor from non-tumor cells. In order to calculate the Ki-67 labeling index, a color histogram is used to separate immunopositive from immunonegative tumor cells since immunopositive/immunonegative ones are usually stained as brown/blue.

3 Experimental Results

The proposed algorithm is tested with 101 whole slide scanned pancreatic NET tissue microarray (TMA) images, which are captured at $20\times$ magnification. Several representative image patches (3 or 4) are cropped from each whole disc scanned image slide (in total over 300 image patches). Each slide corresponds to thousands of mixed tumor and non-tumor cells. In total 20 slides are used for training and 81 slides are reserved for testing. The annotations are created by two pathologists and one pathology resident. The subjective nature is handled using a weighted majority voting considering experience. The cell detection algorithm is implemented with C++ for efficiency, while Matlab is used for cell segmentation and classification on a PC machine with $3.3GHz$ CPU and $16Gb$ memory. A rough estimate of cell diameter needs to be provided to the cell detection algorithm, and we set $\sigma = 0.3, l = 2$ in Equation (2), $\alpha = 0.05, \beta = 0, \gamma = 0.5, \lambda = 5, w = 0.7$ in Equation (4), and $\eta = 8$ in Equation (5). The parameters are selected based on cross-validation and fixed during the testing stage.

3.1 Cell Detection

Both qualitative and quantitative analyses are conducted for the proposed cell detection algorithm. In Figure 1, thousands of cells are correctly detected and segmented in one TMA disc with several zoomed-in patches for better illustration. The computational time for a digitized 2310×2150 pixels TMA disc is 96.5 seconds, compared with more than thirty-minute labor intensive work even for manual counting of 3-4 representative patches from the whole disc. We present the detection results using the proposed method on several small randomly selected patches extracted from the whole dataset in Figure 2 compared with three recent state-of-the-art methods: Laplacian-of-Gaussian filters (LoG) [1], iterative radial voting (IRV) [13], and single-pass voting (SPV) [15]. The proposed algorithm is more robust with respect to the variations of cell scale, shape, and intensity. This can be attributed to the region-based hierarchical voting in the distance map.

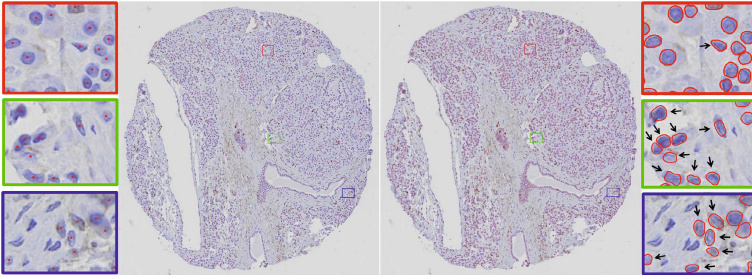


Fig. 1. Cell detection (left panel) and segmentation (right panel) using the proposed method. Several patches are zoomed in for better illustration. The cells pointed out by black arrows in each patch of the right panel are non-tumor cells, and the other segmented cells are tumor cells.

To quantitatively evaluate seed detection accuracy, we present the mean value (MV) and standard deviation (STD) of the Euclidean distances between manually located seeds and those created by automatic algorithms in Table 1, where the missing rate (MR), over-detection rate (OR), and effective rate (ER) are also presented. The missing or over-detection means no seeds or more than one seed are detected for one ground-truth cell, respectively. These two cases are excluded when we compute the MV for fair comparison. The effective rate is to calculate a ratio between the number of detected seeds and ground truth seeds, which measures the methods' robustness to background clutter. $ER = 1$ indicates the strongest robustness. LoG is sensitive to image background, IRV and PSV miss or over-detect some seeds, while the proposed method produces the best performance.

Table 1. Seed detection accuracy compared with ground truth

	$MV \pm STD$	MR	OR	ER
LoG [1]	3.23 ± 1.84	0.03	0.09	1.28
IRV [13]	2.62 ± 2.11	0.12	0.16	1.2
SPV [15]	2.59 ± 2.04	0.10	0.02	0.91
Proposed	2.39 ± 2.0	0.06	0.01	0.99

3.2 Ki-67 Scoring

In the experiments at Stage I, circularity, axis ratio, color mean, standard deviation, kurtosis, contrast, correlation, and homogeneity are selected by the sparse representation model as the most discriminative features to separate tumor from non-tumor cells. This is reasonable since in our dataset tumor cells usually appear more circular than non-tumor cells with a more inhomogeneous texture and relatively lighter staining. The first SVM classifier applies a Gaussian kernel (the parameter $\sigma = 0.3$ and the penalty $c = 1$) with these selected discriminative features. Combined with the texton histogram-based probabilities, the second SVM classifier with the same set of parameters is trained to produce the final classification of tumor and non-tumor cells. Compared with manual annotations by pathologists, we have achieved 87.68% classification accuracy with 87.12%

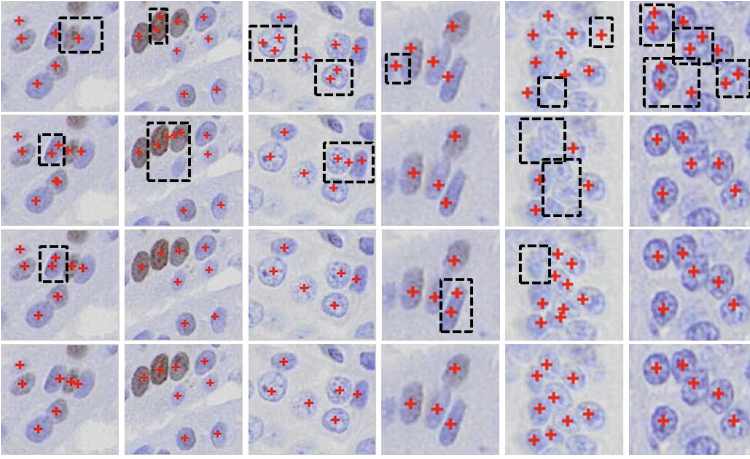


Fig. 2. The geometric centers of cells (seeds) detection on several randomly selected image patches. Rows 1, 2, 3, and 4 denote results created by LoG [1], IRV [13], SPV [15], and the proposed algorithm, respectively. The missing or false seeds are surrounded with black dashed rectangles.

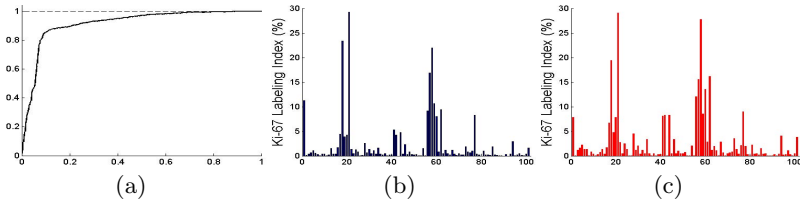


Fig. 3. ROC curve (a) of the classification of tumor cells and lymphocytes, automatic Ki-67 score (b), and manual Ki-67 score (c) for 101 patients (x -axis)

specificity and 88.01% sensitivity. The ROC curve is displayed in Figure 3(a). The Ki-67 indexing numbers for 101 pancreatic NET patients obtained by our method are shown in Figure 3(b), and they are very close to human manual evaluation of Ki-67 scoring (Figure 3(c)). The absolute mean error rate between automatic and manual Ki-67 scores is 0.88%.

4 Conclusion

In this paper, we have introduced an automatic algorithm for Ki-67 scoring on pancreatic NET TMA images. The novel cell detection and Ki-67 scoring system can efficiently and accurately detect thousands of cells in the whole TMA discs and provide accurate Ki-67 score.

Acknowledgement. The project described is supported by the National Center for Research Resources, UL1RR033173, and the National Center for Advancing Translational Sciences, UL1TR000117.

References

1. Al-Kofahi, Y., Lassoued, W., Lee, W., Roysam, B.: Improved automatic detection and segmentation of cell nuclei in histopathology images. *TBME* 57(4), 841–852 (2010)

2. Arteta, C., Lempitsky, V., Noble, J.A., Zisserman, A.: Learning to detect cells using non-overlapping extremal regions. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 348–356. Springer, Heidelberg (2012)
3. Cohen, L.D.: On active contour models and balloons. *CVGIP: Image Understanding* 53(2), 211–218 (1991)
4. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. *PAMI* 24(5), 603–619 (2002)
5. Dzyubachyk, O., van Cappellen, W.A., Essers, J., Niessen, W.J., Meijering, E.: Advanced level-set-based cell tracking in time-lapse fluorescence microscopy. *TMI* 29(3), 852–867 (2010)
6. Funel, N., Denaro, M., Faviana, P., Pollina, L.E., Perrone, V.G., Lio, N.D., Boggi, U., Basolo, F., Campani, D.: The new fully automated system for ki67 evaluation in pancreatic neuroendocrine tumors (pnets). Would it be possible to obtain a standard to grade evaluation? *J. of Pancreas* 13(5S), 562 (2012)
7. Kong, H., Gurcan, M., Belkacem-Boussaid, K.: Partitioning histopathological images: An integrated framework for supervised color-texture segmentation and cell splitting. *TMI* 30(9), 1661–1677 (2011)
8. Liu, X., Harvey, C.W., Wang, H., Alber, M.S., Chen, D.Z.: Detecting and tracking motion of myxococcus xanthus bacteria in swarms. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 373–380. Springer, Heidelberg (2012)
9. Lou, X., Koethe, U., Wittbrodt, J., Hamprecht, F.: Learning to segment dense cell nuclei with shape prior. In: *CVPR*, pp. 1012–1018 (2012)
10. Mohammed, Z.M., Elsberger, B., McMillan, D.C., Going, J.J., Orange, C., Mallon, E., Edwards, J., Doughty, J.C.: Comparison of visual and automated assessment of Ki-67 proliferative activity and their impact on outcome in primary operable invasive ductal breast cancer. *Br. J. Cancer* 106(2), 383–388 (2012)
11. Monaco, J., Hipp, J., Lucas, D., Smith, S., Balis, U., Madabhushi, A.: Image segmentation with implicit color standardization using spatially constrained expectation maximization: detection of nuclei. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 365–372. Springer, Heidelberg (2012)
12. Mosaliganti, K., Gelas, A., Gouaillard, A., Noche, R., Obholzer, N., Megason, S.: Detection of spatially correlated objects in 3D images using appearance models and coupled active contours. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 641–648. Springer, Heidelberg (2009)
13. Parvin, B., Yang, Q., Han, J., Chang, H., Rydberg, B., Barcellos-Hoff, M.H.: Iterative voting for inference of structural saliency and characterization of subcellular events. *TIP* 16, 615–623 (2007)
14. Porikli, F.M.: Integral histogram: A fast way to extract histograms in cartesian spaces. In: *CVPR*, pp. 829–836 (2005)
15. Qi, X., Xing, F., Foran, D.J., Yang, L.: Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set. *TBME* 59(3), 754–765 (2012)
16. Schmid, C.: Constructing models for content-based image retrieval. In: *CVPR*, pp. 39–45 (2001)
17. Yan, P., Zhou, X., Shah, M., Wong, S.T.C.: Automatic segmentation of high-throughput RNAi fluorescent cellular images. *TITB* 12(1), 109–117 (2008)
18. Zimmer, C., Olivo-Marin, J.C.: Coupled parametric active contours. *PAMI* 27(11), 1838–1842 (2005)