

## PAPER

# Nuclei Detection Based on Secant Normal Voting with Skipping Ranges in Stained Histopathological Images

XueTing LIM<sup>†a)</sup>, Nonmember, Kenjiro SUGIMOTO<sup>†b)</sup>, and Sei-ichiro KAMATA<sup>†c)</sup>, Members

**SUMMARY** Seed detection or sometimes known as nuclei detection is a prerequisite step of nuclei segmentation which plays a critical role in quantitative cell analysis. The detection result is considered as accurate if each detected seed lies only in one nucleus and is close to the nucleus center. In previous works, voting methods are employed to detect nucleus center by extracting the nucleus saliency features. However, these methods still encounter the risk of false seeding, especially for the heterogeneous intensity images. To overcome the drawbacks of previous works, a novel detection method is proposed, which is called secant normal voting. Secant normal voting achieves good performance with the proposed skipping range. Skipping range avoids over-segmentation by preventing false seeding on the occlusion regions. Nucleus centers are obtained by mean-shift clustering from clouds of voting points. In the experiments, we show that our proposed method outperforms the comparison methods by achieving high detection accuracy without sacrificing the computational efficiency.

**Key words:** secant normal voting, skipping range, splitting point, nuclei detection, Hematoxylin and Eosin (H&E) staining

## 1. Introduction

Effective cell analysis is important as it could potentially help in the early detection of diseases. The traditional way to analyze cell images is done manually by pathologists or medical experts. However, the manual analysis could be labor intensive, this is because a biological specimen or pathological image usually involves a vast amount of cells.

In the past decades, researchers have been making an utmost effort in digital pathology analysis to help pathologists to reduce their workload by replacing to semi-automated or even fully automated methods. Computer Aided Diagnosis assists professionals and doctors in medical decision making by providing useful medical data interpretation without heavy human workload [1]–[4]. However, the major obstacle in constructing an effective automated nuclei segmentation technique is a nuclei occlusion (nuclei clump) problem. Besides this, detection results deteriorate when this problem comes with several specimen-induced artifacts such as imperfect staining or dust.

To segment the nuclei clump, Kong *et al.* [5] designed an integrated framework that comprises effective cell splitting algorithm with the support of supervised fore-

ground extraction. Veta *et al.* [6] reported a method of cell clump splitting algorithm by fast radial symmetry transform to mark the seeds based on regional minima extraction. Marker-controlled watershed algorithm undergoes intensive study in clumps segmentation problem. One seed per nucleus is required in this method to achieve accurate result. H-minima transform [12], high radial symmetry algorithm [8], and multi-scale Laplacian of Gaussian (LoG) [9] have been developed to detect the seeds. The LoG does not take ellipse orientation information into account, and hence it is hard to adapt all kinds of cell structures. To deal with this problem, generalized LoG (gLoG) kernel [10], [11] is proposed.

Despite the fact that the previous researches provide good detection performance, they still encounter some unresolved problems. The reason is that the above methods are sufficient only for solving homogeneous intensity cell images. Due to the cytological complexity or imperfect staining process, the intra-region intensity of nucleus is not always homogeneous. Texture information exists and multiple seeds are recognized in a single nucleus. The overlapping region of nuclei is often to be interpreted as a single nucleus by these methods. The detection result is relatively insufficient in high occlusion case.

In past decade, some researchers have found a new direction in nuclei detection in reference to nuclei saliency features motivated by symmetry, continuity [14] and closure information. These proposed methods are able to provide good detection performance for both homogeneous and heterogeneous intensity nuclei images by performing voting methods. Parvin *et al.* [13] proposed an automatic nuclei inferring method so-called Multi-pass voting (MPV). Nucleus center is determined by iterative changing and adjusting the voting direction based on optimization problems. Later, Qi *et al.* [15] reduced the computational complexity of the MPV by introducing Single-pass voting (SPV) method. Nucleus center is detected by cone shape voting area. From the detected seeds, nuclei clump are then segmented by minimizing an iterative model based on repulsive level set. Xu *et al.* [16] have performed some improvement and modification based on the SPV method. The computation time is successfully cut down and the accuracy is increased. However, these methods still encounter the problems of false seeding on overlapping and sharp corner regions.

In this paper, a novel nuclei detection method for Hematoxylin and Eosin (H&E) stained breast cancer images is proposed. The new method consists two major steps: se-

Manuscript received October 3, 2017.

Manuscript publicized November 14, 2017.

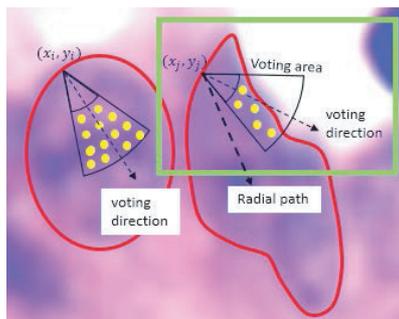
<sup>†</sup>The authors are with the Graduate School of Information, Production and Systems, Waseda University, Kitakyushu-shi, 808-0135 Japan.

a) E-mail: xtlm1123@fuji.waseda.jp

b) E-mail: ksugimoto@aoni.waseda.jp

c) E-mail: kam@waseda.jp

DOI: 10.1587/transinf.2017EDP7326



**Fig. 1** Illustration of voting method within a cone shape voting area oriented from negative image gradient defined voting direction [15]. Green box shows an inappropriate voting point falling on the sharp corner of nucleus, which tends to generate false seeds.

cant normal voting and determination of skipping ranges to overcome the drawbacks of previous work. We found that employing image gradient as voting direction may cause false seeding as shown in Fig. 1. Secant normal voting with skipping ranges manages to avoid the over-segmentation problem. The Skipping range is set to avoid voting on occlusion region. Furthermore, our proposed method is uninfluenced by the region sharp corners. In the experiments, we show that the proposed method can achieve good performance on both detection accuracy and processing time. The contributions of this paper are in the following:

1. We propose a novel nuclei detection framework that can accurately detect the nucleus center based on nuclei geometry information.
2. A skipping ranges scheme is introduced to avoid false seeding on the nuclei occlusion region.
3. In the experiments, we show the proposed method can achieve faster detection result for H&E stained breast cancer images.

## 2. Voting Concept and Remaining Problems

In this section, background knowledge of voting method is briefly reviewed first before going to the remaining problems of previous works.

### 2.1 Voting Method and Voting Direction

With prior knowledge of radial information, a nucleus center can be automatically detected by a voting method. Figure 1 describes the voting method where a cone-shape voting area is oriented from a voting direction. The voting direction should be matched with the radial path to accurately detect a nucleus center. For previous works [13], [15], [16], image gradient is adopted as the voting direction. The voting direction should not be affected by noise in order to guarantee its effectiveness in detection. However, we found that employing image gradient as voting direction tends to generate false seeds. This is because image gradient is sensitive to the intensity variation inside the nuclei (intra-region).

### 2.2 Remaining Problems

Previous works often encounter sharp corner problem. Green box in the Fig.1 reveals a voting direction at a sharp corner, where it is far from the radial path (path towards a nucleus center). This causes the false seeding on the boundary nearby region. Moreover, for the methods [13], [15], [16], all boundary points along the tangential direction are used to compute the voting points. Employing all the boundary points deteriorates the detection performance. This is because the voting area may fall on occlusion regions, especially nucleus regions with high intensity variation. However, secant normal voting and skipping range in our proposed method are able to solve these problems.

## 3. Proposed Method

This section describes in detail how the proposed method can detect the nucleus center precisely. The proposed method consists of two major steps: secant normal voting and determination of skipping range, to overcome the drawbacks of previous works [13], [15], [16]. A list of parameters and notation used in this paper is shown in Table 1.

### 3.1 Brief Overview

Initially, some foreground regions are extracted from a histopathological image. These extracted regions are represented as a binary (black and white) mask and each region is categorized into clump or single nucleus regions based on some criteria. The extracted clumps are independently processed by the proposed voting method.

Let  $B = \{p_1, p_2, \dots, p_M\}$ ,  $p_m \in \mathbb{R}^2$  be a sequence of boundary points that encloses a clump region, where  $M$  is the total number of boundary points. The boundary is a cyclic curve traced in a clockwise direction. Our proposed method aims at generating a set of voting points  $C = \{c_1, c_2, \dots, c_N\}$ ,  $c_n \in \mathbb{R}^2$  from the sequence  $B$  based on a new voting method, where  $N$  is the total number of voting points. We name this approach as *secant normal voting* which will be described in Sect. 3.2. A voting point is expected to locate around a nucleus center.

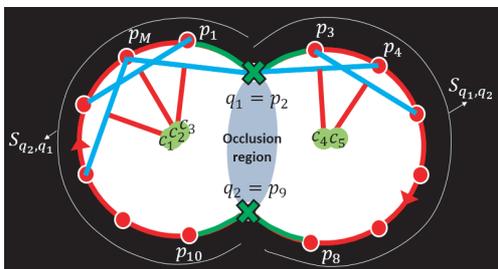
Figure 2 illustrates a simple region mask and some of the voting points allocated from the mask. From the figure, two nuclei are partially occluded to each other. As will be described in Sect. 3.3, we avoid voting on the occlusion region by our skipping range technique.

### 3.2 Secant Normal Voting

This subsection describes an alternative way to define voting direction by employing a normal line from a boundary secant. For information, a secant is a line passing through two points on an ellipse. The normal line of boundary secant mimics the function of image gradient in the previous works to infer a nucleus center without involving intensity

**Table 1** List of parameters and notations

Para.	Description	Value	Notation	Description
$r_{min}$	Estimated minimum radius of nucleus in the image	5	$B = \{p_1, p_2, \dots, p_M\}$	A sequence of boundary points
$r_{max}$	Estimated maximum radius of nucleus in the image	10	$C = \{c_1, c_2, \dots, c_N\}$	A set of voting points
$d$	Distance threshold	$[r_{min}, r_{max}]$	$Q = \{q_1, q_2, \dots, q_K\}$	A sequence of splitting points
$l$	Length threshold	$r_{min}$	$V = \{v_1, v_2, \dots, v_J\}$	A sequence of base vertices
$r$	Minimum distance of a pair of valid based vertices	6	$\mathcal{A}$	A sequence of alpha shape points
$\alpha$	Tightness control of alpha shape	10	$\mathcal{S}$	A set of subsequences
$\rho$	Threshold to decide single nucleus or clump	0.95	$S_{q_a, q_b}$	A subsequence begins with splitting point $q_a$ and ends with $q_b$
			$U_k$	A skipping range in reference to splitting point $q_k$
			$\beta_m$	A Boolean flag value attached to boundary element $p_m$



**Fig. 2** Secant normal voting points  $c_1, c_2, \dots, c_5$  are generated from sequence  $B$  in a foreground region. Voting on the occlusion region is avoided by splitting point (green cross). Notice that each red dot is actually adjacent to each other.

information. The proposed method allocates voting points around the nucleus center using the secant normal line generated from the elements of sequence  $B$ .

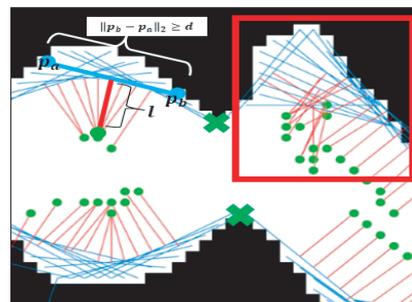
Figure 2 demonstrates a simple clump case where two nuclei are partially occluded to each other. The occluded nuclei can be imagined as two partially overlapped circles with two intersection points (the green crosses). Each intersection point locates at a boundary concave curve and touches to the lens-shaped intersection region. We call these intersection points as *splitting points*. This is because the lens-shaped intersection region is actually an occlusion region where two occluded nuclei should be split.

Let  $Q = \{q_1, q_2, \dots, q_K\} \subset B$  be a sequence of splitting points and  $K$  is the total number of points. The sequence of splitting points  $Q$  is useful to avoid voting on occlusion regions. They are skipped from constructing secant lines and allocating voting points. The detail of splitting points will be described in Sect. 3.3.

The sequence of splitting points  $Q$  partitions  $B$  into  $K$  number of subsequences  $\mathcal{S}$ . Let  $S_{q_a, q_b}$  be one of the subsequences of  $B$ , where  $q_a$  and  $q_b$  are the start point and end point respectively. Therefore, we get  $\mathcal{S} = \{S_{q_1, q_2}, S_{q_2, q_3}, \dots, S_{q_K, q_1}\}$ . For example, Fig. 2 shows that the sequence  $B$  is partitioned into two subsequences by two splitting points  $q_1$  and  $q_2$ .

For each subsequence  $S$ , two boundary points  $p_a \in S$  and  $p_b \in S$  are selected to compute the secant. We compute a voting point  $c$  by:

$$c = \frac{p_a + p_b}{2} + l \left( \mathcal{R}_{90^\circ} \frac{p_b - p_a}{\|p_b - p_a\|_2} \right), \quad (1)$$



**Fig. 3** Examples of secant normal voting with  $d = 10$  and  $l = 5$ : red box shows how  $d$  value selection can deal with the sharp corner on a boundary; red line: normal line from midpoint with length  $l$ ; blue line: secant linked by two boundary points; green cross: splitting point; green dot: voting point.

where

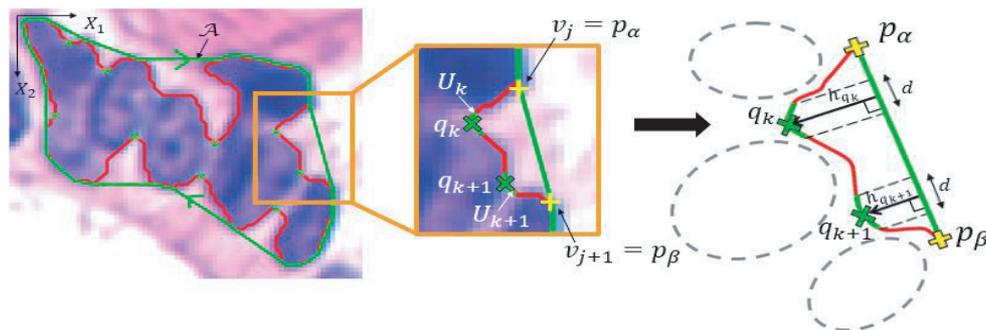
$$\mathcal{R}_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, \quad (2)$$

is a rotation matrix,  $\|\cdot\|_2$  indicates the  $\ell_2$ -norm of a vector, while length  $l$  is the shortest distance from a voting point  $c$  to a secant line. If  $p_a$  is fixed, then  $p_b$  is selected as the first succeeding boundary point that satisfies the relation  $\|p_b - p_a\|_2 \geq d$ . As  $d$  approaches to zero, a secant converges to a boundary tangent. The computation of secant normal is repeated by sequentially taking next two points:  $p_{a+1}$  and  $p_{b+1}$  until it reaches the final point in subsequence  $S$ . The procedure is then iterated for the following subsequences until  $S_{q_K, q_1}$  is selected.

The threshold  $d$  should be selected carefully. Large  $d$  can well tackle the problem of sharp corners in a region. This is because a voting point  $c$  is avoided to assign on the shape corner regions. In contrast, small  $d$  is sensitive to these sharp corners. In the experiments, we will show that the higher detection rate can be achieved when both  $d$  and  $l$  are selected between a range of minimum to maximum radius of nucleus. Figure 3 illustrates the process of secant normal voting and how the threshold  $d$  can handle the sharp corner of a region.

### 3.3 Determination of Skipping Range

This subsection discusses how to determine a sequence of



**Fig. 4** The slipping ranges (green line) and splitting points (green cross) are extracted from a pair of valid base vertices (yellow plus). Green cyclic curve is the  $\alpha$ -shape curve,  $\mathcal{A}$  with  $\alpha = 10$ , and red cyclic curve is the boundary curve.

splitting points  $Q$  and its corresponding skipping ranges. The motivation of splitting point and skipping range initiated by the area difference between a region enclosed by both boundary and alpha shape. Alpha shape ( $\alpha$ -shape) is a linear approximation of the original shape [20]. In Fig. 4, we draw an  $\alpha$ -shape as the green cyclic curve which encloses the foreground region. The tightness of  $\alpha$ -shape is controlled by a non-negative parameter- alpha radius,  $\alpha = [0, \infty]$ . If  $\alpha \rightarrow \infty$ , it converges to a convex hull. The reason for choosing  $\alpha$ -shape is that  $\alpha$ -shape allows users to control the tightness of the curve. It is more suitable than convex hull to “tie” a donuts-shaped clump region (as shown in Fig. 6) for determining the sequence  $Q$ .

Based on our observation, a nearby region of a boundary concave curve has high possibility to be treated as an occlusion region. The boundary points of the boundary concave curve is a subsequence of  $B$  that cannot be touched by an  $\alpha$ -shape. Therefore, to find the boundary concave curve, we extract two points where the boundary and  $\alpha$ -shape start to branch away and merge together from each other (simply speaking, we want to find the starting and end points of some subsequences of  $B$  where these subsequences do not touch the  $\alpha$ -shape). We named these two points as base vertices. As shown in Fig. 4, the boundary concave curve can be drawn as a graph with some concave-up regions. The local minima of the graph are then taken as the splitting points and they should split the occluded nuclei into three.

Let  $\mathcal{A} \subseteq B$  be a sequence of the  $\alpha$ -shape points that encloses a clump region where  $\mathcal{A}$  is traced in a clockwise direction as a green cyclic curve as shown in Fig. 4. Before estimating the sequence  $Q$ , we attach a Boolean flag value to each boundary point in  $B$  by:

$$\beta_m = \begin{cases} 1, & \text{if } p_m \in \mathcal{A} \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

We use the element of  $B$  satisfying  $(\beta_m = 1) \cap (\beta_{m-1} \neq \beta_{m+1})$  as a base vertex. Hence, we can obtain a sequence of base vertices  $V = \{v_1, v_2, \dots, v_J\}$ ,  $V \subset B$ ,  $v_j \in \mathbb{R}^2$ , where  $J$  is the total number of base vertices.

Let us focus on one splitting point  $q_k$ . From  $V$ , if we set  $v_j = p_\alpha$ ,  $v_{j+1} = p_\beta$  and  $j$  is an odd number, then  $p_\gamma$  is any

boundary point in between  $p_\alpha$  and  $p_\beta$  that satisfies  $\alpha < \gamma < \beta$ . Two consecutive base vertices must satisfy the relation  $\|v_j, v_{j+1}\|_2 \geq r$ , where  $\|\cdot\|$  indicates the Euclidean distance from  $v_j$  to  $v_{j+1}$ . Threshold  $r$  has to be fixed large because  $q_k$  is unlikely to present within a small concave region that mostly leads to imperfect extraction.

Let  $p_\alpha = (x_\alpha, y_\alpha)$ ,  $p_\beta = (x_\beta, y_\beta)$ , and  $p_\gamma = (x_\gamma, y_\gamma)$ . We find the minimum distance of point  $p_\gamma$  to line  $p_\alpha p_\beta$  by the following equation:

$$h_{p_\gamma} = \frac{|(y_\beta - y_\alpha)x_\gamma - (x_\beta - x_\alpha)y_\gamma + x_\beta y_\alpha - y_\beta x_\alpha|}{\|p_\beta - p_\alpha\|_2}. \quad (4)$$

Hence, we obtain  $\mathcal{H} = \{h_{p_\alpha}, \dots, h_{p_\gamma}, \dots, h_{p_\beta}\}$ . Then, the point  $p_\gamma$  with local maxima distance value in  $\mathcal{H}$  is extracted as  $q_k$ . If multiple local minima are obtained, the next splitting points will be  $q_{k+1}$ ,  $q_{k+2}$ , and so on.

To make our detection method more robust to noise, we skip some parts of boundary points, that is *skipping range*. One skipping range is computed in reference to one splitting point and we get the skipping range  $U_k = \{p_{\gamma - \lceil \frac{d}{r} \rceil}, \dots, q_k, \dots, p_{\gamma + \lceil \frac{d}{r} \rceil}\} \subset B$ , where  $q_k = p_\gamma$  and  $\lceil \cdot \rceil$  is the ceil function. The computation of skipping range from splitting point is repeated until  $p_\beta = v_j$ . Single or multiple splitting points can be estimated between a pair of valid base vertices in Fig.4.

#### 4. Nuclei Detection

In this section, we describe the overall framework of nuclei detection. Initially, all regions of foreground mask are classified into single nucleus or nuclei clumps. We only implement the proposed method on clump regions. Then, seed point decision from clouds of voting points is explained. Figure 5 shows an overview of proposed nuclei detection framework that will be discussed shortly.

##### 4.1 Single Nucleus and Nuclei Clumps Classification

Nuclei detection of a single nucleus is an easy task by simply taking the region centroid. Implementing clump splitting method in single nucleus is redundant and unnecessary.

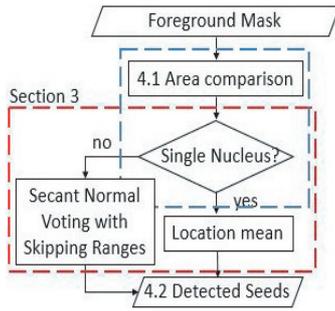


Fig. 5 Overview of proposed nuclei detection method.

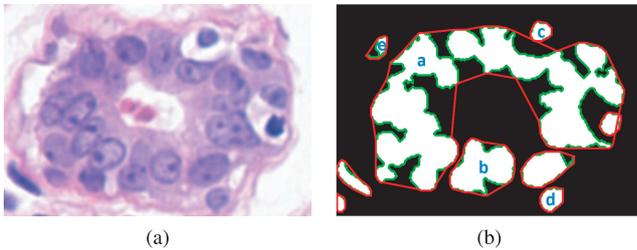


Fig. 6 Nuclei boundary and convex hull delineation. (a) Original histopathological image. (b) Foreground mask where region boundary shown by green line, convex hull shown by red line.

Thus, the single nucleus and nuclei clump must be treated separately. Let  $\mathcal{R} = \{R_1, R_2, \dots, R_g, \dots, R_G\}$  be a set of connected regions in a foreground mask. All regions are then classified into single nucleus, SN or nuclei clump, NC.

To determine which class a region  $R_g$  belongs to, we draw a boundary and its corresponding convex hull for each connected region. A convex hull is the smallest convex that encloses a set of points. Single nucleus often tends to be nearly perfect ellipse shape, hence its boundary is fitted well with the convex hull. Figure 6(b) shows several connected regions in a foreground mask which are enclosed by boundary curve and convex hull.

Let  $|P_B|$  and  $|P_C|$  be total number of pixels within a region which is enclosed by the boundary and convex hull respectively. We get the region area ratio  $\Phi$  by

$$\Phi = \frac{|P_B|}{|P_C|}. \quad (5)$$

Table 2 shows  $|P_B|$ ,  $|P_C|$  and  $\Phi$  of regions marked **a**, **b**, **c**, **d**, and **e** in Fig.6 (b). From the table, we found that single nucleus often has a higher  $\Phi$  than a clump. However, a region with low  $\Phi$  ratio but small area should be classified as SN as well, for example region **e**. These small connected regions with imperfect elliptical shape are the outcomes of imperfect extraction results. They are usually occluded by background region with highly similar intensity information.

Thus, an extracted region can be classified as SN if it satisfies the condition  $(\Phi \geq \rho) \cap (|P_B| \leq A_{\mathcal{R}})$ . The remaining regions are belonging to class NC. In our work, we assign  $\rho$  to 0.95.  $A_{\mathcal{R}}$  is the mean area of all connected regions  $\mathcal{R}$  in an image. For each SN region, final seed location is selected as the centroid location  $(\mu_{x_1}, \mu_{x_2})$ , where  $\mu_{x_1}$  and  $\mu_{x_2}$  are the

Table 2  $P_B$ ,  $P_C$  and ratio  $\Phi$  of five connected regions in Fig.6(b).

Region	a	b	c	d	e
$ P_B $	5869	884	118	158	70
$ P_C $	8799	1026	123	165	116
$\Phi$	0.67	0.86	0.96	0.96	0.60

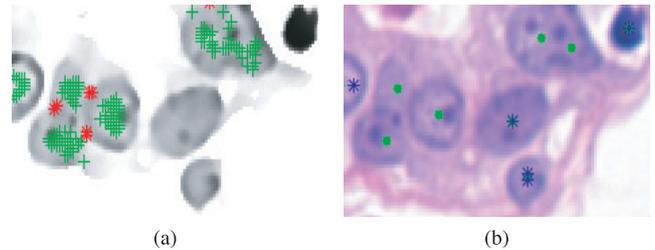


Fig. 7 Seed decision from voting points by mean-shift clustering. (a) voting points (green) and splitting points (red); (b) detected seeds: clump (green) and single nucleus (blue).

mean of  $x_1$  and  $x_2$  coordinates of the region respectively. For all the NC regions, our proposed voting method is applied to get  $C$ .

## 4.2 Seed Decision

A set of seed points is then selected from voting points set  $C$  by mean-shift clustering. The complete match between radial path and gradient direction addressed by [13] can be ignored by using mean-shift clustering. The voting points do not require to be placed at the exact location of nucleus center. As long as all the voting points are assembled near the center, a finalized seed point can be correctly determined.

We assign the cluster bandwidth to nucleus diameter in this case. Left image in Fig.7 shows the detected splitting points (red asterisk) as well as the voting points (green cross). Right image in Fig.7 illustrates some final detected seeds after mean-shift clustering. The detection of seeds in clumps and single nucleus are carried out separately to reduce the computational time. The whole procedure of voting based nuclei detection is expressed in Algorithm 1.

---

### Algorithm 1: Proposed nuclei detection algorithm

---

**Input:**  $N$  nuclei clump regions of a foreground mask

**Initialize parameters:**

Set  $d = [r_{min}, r_{max}]$ ,  $l = r_{min}$ ,  $r = 6$ ,  $\alpha = 10$ .

**for**  $i = 1, \dots, N$ -th clump **do**

1, Obtains boundary points sequence  $B$

2, Gets  $\alpha$ -shape points sequence  $\mathcal{A}$  from  $B$  and  $\alpha$

3, Extracts splitting points sequence  $Q$  from  $B$  and  $\mathcal{A}$  by Eq. (4).

4, Partitions  $B$  into  $K$  subsequences  $S$  by  $Q$

**for**  $k = 1, \dots, K$ -th subsequence **do**

Constructs secant line from two boundary points with minimum  $d$  apart and gets the voting point  $c$  by Eq. (1).

5, Gets final decision seed points from set of voting points  $C$

by mean-shift clustering.

---

## 5. Experimental Results and Discussions

The UCSB Bio-Segmentation Benchmark [19] dataset is employed for evaluation purpose in this study. The dataset is a set of 58 H&E stained breast cancer images with two different sizes: 896x768 and 768x512. For each image, the dataset provides a small region of ground truth mask with approx. 250x250 size (in range of 30 to 100 nuclei). The proposed nuclei detection algorithm is implemented by MATLAB on a Intel Core i5-6200U CPU 2.30 GHz environment.

### 5.1 Parameters Selection

All parameters involved in the proposed algorithm are listed in Table 1. These parameters are tuned carefully and repeatedly in the experiment to achieve an optimal result for UCSB Bio-Segmentation Benchmark dataset.

In this experiment, the minimum and maximum of nucleus radius are chosen as 5 and 10 respectively. Both estimated radius values should be adjusted based on the input image. Threshold  $d$  is selected to be the average value of  $r_{min}$  and  $r_{max}$  while length  $l$  is assigned to be  $r_{min}$  as shown in Table 1. Value  $\alpha$  can be set as infinity (convex hull). However, we found that  $\alpha = 10$  can increase detection accuracy as we have explained in Sect. 3.3. Parameter  $r$  and  $\rho$  can keep constant throughout the experiment. Parameter  $r$  is used for eliminating unlikely concave region. Setting  $r$  to very small value does not influence the detection result but extends execution time. Since the processor speed is fast, the time lagging due by small  $r$  is negligible.

### 5.2 Secant Normal Voting Based Nuclei Detection Performance

The performance of proposed nuclei detection framework is evaluated by comparing with some existing works: gLoG [11], MPV [13] and improved SPV [16]. The detection results are then evaluated by calculating the precision  $P$  and recall  $R$  percentages [21] as follows:

$$P = \frac{N_{TP}}{N_{DS}} \times 100\%, \quad R = \frac{N_{TP}}{N_{GT}} \times 100\%, \quad (6)$$

where  $N_{DS}$  is the number of detected seeds,  $N_{GT}$  is the number of manually annotated seeds from ground truth image, while  $N_{TP}$  is the number of true positive seeds. A detected seed is considered as true positive if and only if it is within 5 pixels distance from a ground truth annotated seed.

To evaluate the detection efficiency of all methods, we modified the dataset images. We remove all remaining stains and dusts on the background region based on the provided ground truth masks. Hence, no pre-processing is required and none of the methods will be influenced by the background noises or unwanted particles. Table 3 illustrates the detection performances of four methods. From the table, the proposed method shows its effectiveness in achieving more accurate and faster detection results by recording

highest recall and precision percentages. From Table 3, the improved SPV requires at least 5 times more the execution time than the other methods. This is because clustering vast of voting points from cone-shaped voting area is time consuming.

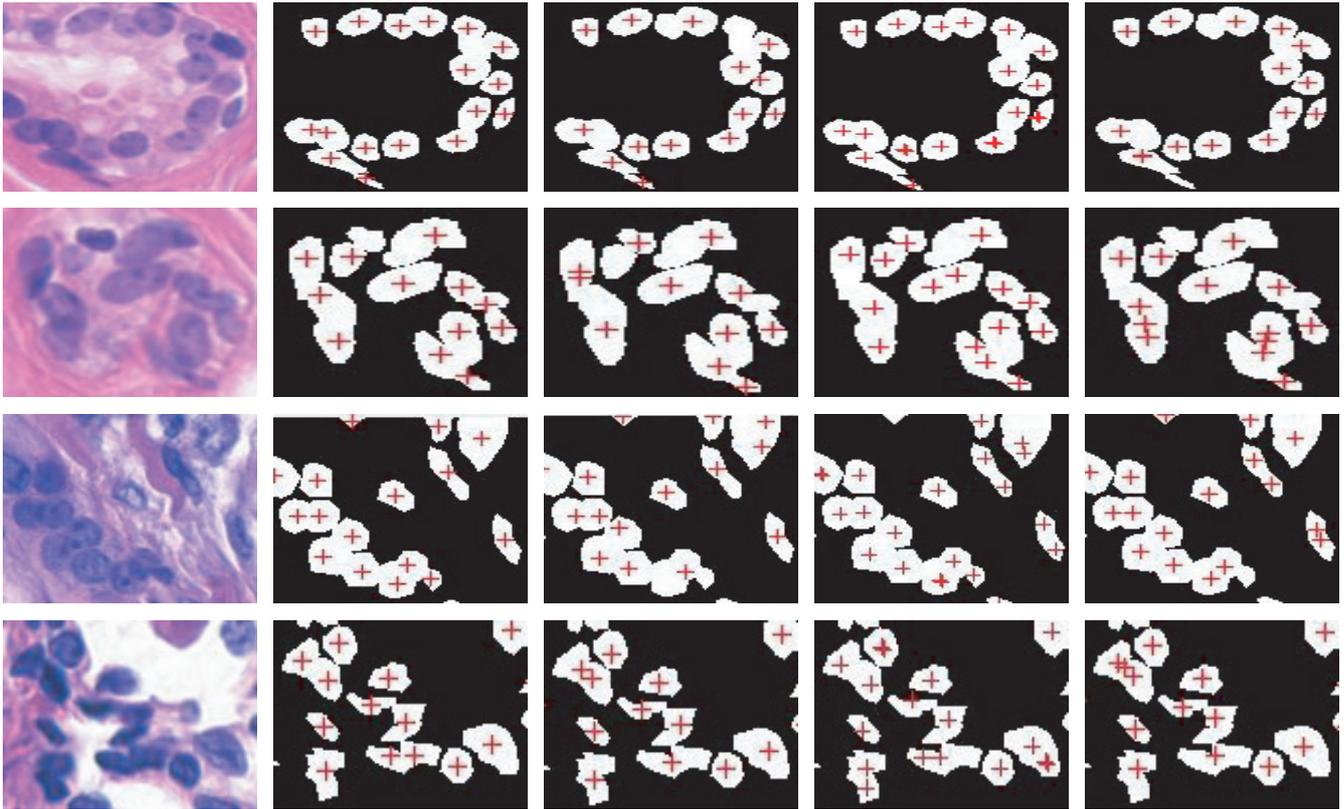
Figure 8 visually shows the nuclei detection and segmentation results of four methods on a set of breast cancer tissue images. Generalized LoG tends to generate false seeds for non-homogeneous images. For high heterogeneous images, seeds are inaccurately assigned on nucleoli (the small dark dot that appears within a nucleus) or dark overlapping regions. In the fourth column of Fig. 8, at first glance, MPV presents a better detection results. However, we notice that multiple seeds are detected in single nucleus even though they are adjacent to each other. For more accurate detection, an additional clustering step is required. MPV's detection results are strongly affected by the parameters selection. Furthermore, MPV fails to detect most the seeds in a high occluded image.

For improved SPV, the deficiency of voting direction selection results in false detection of nuclei. The voting direction deviates from pointing towards the nucleus center if the boundary does not have significant concavity. In contrast to all the methods above, the proposed method increases the detection accuracy by employing secants and skipping ranges. Voting on the overlapping region is avoided. Moreover, selecting suitable alpha radius value prevents over-segmentation on red bean-shaped nucleus as shown in the right-hand side of third-row images. The proposed method also successfully reduces the computation time as it only involves simple point-based voting scheme.

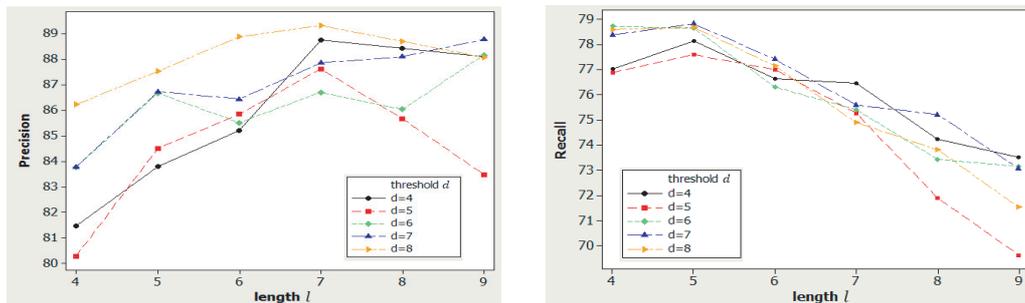
Figure 9 shows the precision and recall results of proposed method with various threshold  $d$  values (4 to 8) and length  $l$  (4 to 9) on a set of randomly selected breast cancer images. We found that the recall percentage starts to decrease when length  $l$  increases from  $r_{min}$ . Setting  $l \rightarrow r_{max}$  can only detect large radius nuclei in the image. However, selecting  $l \rightarrow r_{min}$  is more suitable to detect both small and large radius nuclei. This is because small  $l$  value still allows voting points to be accumulated near to the nucleus center. In contrast, large  $l$  may accidentally assign voting points beyond the nucleus boundary. The recall result does not deviate much for each  $l$  value by altering the threshold value  $d$ . It shows the robustness of our proposed method towards threshold  $d$ .

**Table 3** Nuclei detection performance of proposed method and the comparison methods [11], [13], [16] in the UCSB Bio-Segmentation Benchmark dataset.

Method	Precision, $P$	Recall, $R$	Average time, $t(s)$
Proposed Method	93.46%	90.92%	1.13
gLoG [11]	86.57%	83.60%	1.69
MPV [13]	77.13%	73.92%	<1
Improved SPV [16]	89.71%	88.00%	6.47



**Fig. 8** H&E breast cancer nuclei detection result comparisons. First column is a set of color images. From second column to last column are the proposed method, gLoG [11], MPV [13], and improved SPV [16] respectively.



**Fig. 9** Precision and recall for  $d = 4, 5, 6, 7, 8$  and  $l = 4, 5, 6, 7, 8, 9$

### 6. Conclusions and Future Work

In this paper, we propose a novel nuclei detection method. The problem of voting on a sharp corner area is solved by employing secant in the voting scheme. We named the new voting method as secant normal voting. The detection accuracy is strengthened by the predetermined skipping range. Skipping range reduces the chance of over-segmentation by avoiding the false seeding on occlusion regions. In the experiments, we show that the proposed method be able to provide more accurate detection result. At the same time, computation time is reduced.

In future work, we will mainly focus on the multi-scale

problem of cell nuclei. Besides this, we hope to construct a more efficient segmentation algorithm based on the detected seeds and splitting points.

### Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Number 15K00248 and Waseda University Grant for Special Research Projects (Project number: 2016B-201).

### References

[1] J. Gil, H. Wu, and B.Y. Wang, "Image analysis and morphometry in the diagnosis of breast cancer," *Microsc. Res. Tech.*, vol.59, no.2,

- pp.109–118, 2002.
- [2] W.N. Street, W.H. Wolberg, and O.L. Mangasarian, “Nuclear feature extraction for breast tumor diagnosis,” *Proc. Int. Symp. Electron. Imag.: Sci. Technol.* San Jose, CA, USA, vol.1905, pp.861–870, 1-4 Feb. 1993.
  - [3] C. Wahalby, I.M. Sintorn, F.Erlandsson, G. Borgefors, and E. Bengtsson, “Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections,” *J. Microsc.*, vol.215, no.1, pp.67–76, 2004.
  - [4] K. Nguyen, A.K. Jain, and B. Sabata, “Prostate cancer detection: fusion of cytological and textural features,” *J. Pathol. Informat.*, vol.2, no.2, pp.7–27, Dec. 2011.
  - [5] H. Kong, M. Gurcan, and K. Belkacem-Boussaid, “Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting,” *IEEE Trans. Med. Imag.*, vol.30, no.9, pp.1661–1677, Sept. 2011.
  - [6] M. Veta, P.J. van Diest, R. Kornegoor, A. Huisman, M.A. Viergever, and J.P.W. Pluim, “Automatic nuclei segmentation in H&E stained breast cancer histopathology images,” *PLoS ONE*, vol.8, no.7, p.e70221, July 2013.
  - [7] A.C. Ruifork and D.A. Johnston, “Quantification of histochemical staining by color deconvolution,” *Anal. Quant. Cytol. Histol.*, vol.23, no.4, pp.291–299, 2001.
  - [8] G. Loy and A. Zelinsky, “Fast radial symmetry for detecting points of interest,” *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol.25, no.8, pp.959–973, 2003.
  - [9] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, “Improved automatic detection and segmentation of cell nuclei in histopathology images,” *IEEE Trans. Biomed. Eng.*, vol.57, no.4, pp.841–852, April 2010.
  - [10] H. Kong, H.C. Akakin, and S.E. Sarma, “A generalized laplacian of gaussian filter for blob detection and its applications,” *IEEE Trans. Cyber.*, vol.43, no.6, pp.1719–1733, 2013.
  - [11] H. Xu, C. Lu, R. Berendt, N. Jha, and M. Mandal, “Automatic nuclei detection based on generalized laplacian of gaussian filter,” *IEEE J. Biomed. Health Inform.*, vol.21, no.3, pp.826–837, 2017.
  - [12] C. Jung and C. Kim, “Segmenting clustered nuclei using h-minima transform based marker extraction and contour parameterization,” *IEEE Trans. Biomed. Eng.*, vol.57, no.10, pp.2600–2604, Oct. 2010.
  - [13] B. Parvin, Q. Yang, J. Han, H. Chang, B. Rydberg, and M.H. Barcellos-Hoff, “Iterative voting for inference of structural saliency and characterization of subcellular events,” *IEEE Trans. Image Process.*, vol.16, no.3, pp.615–623, March 2007.
  - [14] A. Shashua and S. Ullman, “Structural saliency: The detection of global salient structures using a locally connected network,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp.321–327, 1998.
  - [15] X. Qi, F. Xing, D.J. Foran, and L. Yang, “Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set,” *IEEE Trans. Biomed. Eng.*, vol.59, no.3, pp.754–765, March 2012.
  - [16] H. Xu, C. Lu, and M. Mandal, “An efficient technique for nuclei segmentation based on ellipse descriptor analysis and improved seed detection algorithm,” *IEEE J. Biomed. Health Inform.*, vol.18, no.5, Sept. 2014.
  - [17] K. Sugimoto and S.I. Kamata, “Compressive bilateral filter,” *IEEE Trans. Image Process.*, vol.24, no.11, pp.3357–3369, 2015.
  - [18] C.H. Li and C.K. Lee, “Minimum cross-entropy thresholding,” *Pattern Recognit.*, vol.26, no.4, pp.617–625, 1993.
  - [19] E.D. Gelasca, J. Yun, B. Obara, and B.S. Manjunath, “Evaluation and benchmark for biological image segmentation,” *IEEE Int. Conf. Image. Process.*, pp.1816–1819, Oct. 2008.
  - [20] H. Edelsbrunner, D. Kirkpatrick, and R. Seldel, “On the shape of a set of points in the plane,” *IEEE Trans. Inf. Theory*, vol.29, no.4, pp.551–559, July 1983.
  - [21] R. Baeza-Yates and B. Ribeiro-Neto, “Modern information retrieval,” 1st ed. Reading, MA: Addison-Wesley, 2005.



**XueTing Lim** received the B.E. degree from Universiti Teknologi Malaysia in 2015. She is now in Graduate School of Information, Production and Systems, Waseda University. Her research interests includes medical image processing and pattern recognition.



**Kenjiro Sugimoto** received the B.E. and M.E. degrees from the Kurume National College of Technology, Japan and Waseda University, Japan in 2007 and 2009, respectively. He had been a Research Fellow of the Japan Society for the Promotion of Science, from 2010 to 2012, and is currently a Junior Researcher with the Graduate School of Information, Production and Systems, Waseda University. His current research interests are in pattern recognition, image processing, and data compression. He is a member

of the IEEE.



**Sei-ichiro Kamata** received the M.S. degree in computer science from Kyushu University, Fukuoka, Japan, in 1985, and the Ph.D. degree in computer science from Kyushu Institute of Technology, Kitakyushu, Japan, in 1995. From 1985 to 1988, he was with NEC, Ltd., Kawasaki, Japan. In 1988, he joined the Kyushu Institute of Technology, as a Faculty Member. From 1996 to 2001, he was an Associate Professor with the Department of Intelligent System, Graduate School of Information Science and Electrical Engineering, Kyushu University. Since 2003, he has been a Professor with the Graduate School of Information, Production and Systems, Waseda University. In 1990 and 1994, he was a Visiting Professor with the University of Maine, Orono, US. His research interests include image processing, pattern recognition, image compression, and space-filling curve application. He is a member of the IEEE, ITEJ, and IPSJ.