

Automatic calculation and visualization of nuclear density in whole slide images of hepatic histological sections

Akito Nagase^a, Masanobu Takahashi^{a,*} and Masayuki Nakano^b

^a*College of Systems Engineering and Science, Shibaura Institute of Technology, Saitama, Japan*

^b*Shonan Fujisawa Tokushukai Hospital, Kanagawa, Japan*

Abstract. An automatic method for estimating the density of cell nuclei in whole slide images (WSI's) of hepatic histological sections is presented. The method can be applied to hematoxylin-eosin (H&E) stained sections with slight histological atypism, such as early well-differentiated hepatocellular carcinoma (ewHCC). It is shown that measured nuclear density is affected by the nuclear size due to fragments of nuclei. This size-dependent problem has been solved by estimating the standard nuclear area for each image patch. The method extracts typical nuclei, which are used to automatically adjust the parameters, including the standard nuclear area. The method is robust for variations in contrast, color, and nuclear size. 40 image patches sampled from 20 WSI's of surgical sections were used for accuracy evaluation. The mean absolute percentage error of estimated nuclear densities was 8.2%. It was also confirmed that the distributions of nuclear density were successfully estimated and visualized for all 20 WSI's. The computation time for a WSI of typical surgical section (754 mm², about 1,280,000 nuclei) was about 57 minutes on a PC.

Keywords: Whole slide image, histopathology, liver, nuclear density, segmentation

1. Introduction

Because the differential diagnosis of borderline lesions is left to the pathologist's discretion, this form of diagnosis has become a serious problem. Quantitative information is necessary, because most pathological diagnosis is based on the subjective opinion of pathologists [1]. It is therefore important to analyze histological images that provide important quantitative information for diagnosis. Specifically, nuclei detection and segmentation are important for cancer diagnosis and grading [2]. Among features related to nuclei, the nuclear density, which is defined as the number of cell nuclei per unit area, is one of the key features to be observed. Figure 1 shows microscopic images of early well-differentiated hepatocellular carcinoma (ewHCC) and non-cancer. The histological atypism of the ewHCC is so slight that it is very difficult to differentiate from benign nodules [3]. The nuclear density is reported to be one of the important features to observe while differentially diagnosing ewHCC,

* Address for correspondence: Masanobu Takahashi, College of Systems Engineering and Science, Shibaura Institute of Technology, 307, Fukasaku, Minuma-ku, Saitama City, 3378570, Japan. Tel.: +81-48-687-5844; Fax: +81-48-687-5198; E-mail: mtaka@shibaura-it.ac.jp.

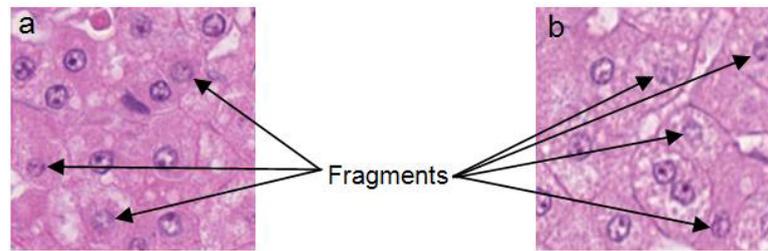


Fig. 1. Sample images of H&E stained hepatic histological section: (a) early well-differentiated hepatocellular carcinoma (ewHCC) and (b) non-cancer.

because the nuclear density of ewHCC is usually higher than that of non-cancerous areas [4]. Thus, a support system was developed for estimating nuclear density and diagnosing ewHCC [5]. The system can extract nuclear positions in hematoxylin-eosin (H&E) stained hepatic histological sections. Incorrect positions can be fixed by the users using a graphical user interface, and the corrected nuclear positions are used to calculate the nuclear density. The system is useful for obtaining the nuclear density of one ROI. However, the system is not automatic. Deciding the ROI is also problematic, because the nuclear density fluctuates within a single nodule. Although this problem might be relieved by increasing the number of ROI's, this requires more effort from the user; it takes about 2 minutes to extract and correct the nuclear positions in one ROI ($0.30 \text{ mm} \times 0.23 \text{ mm}$) [5]. It would be useful if the feature distribution were calculated fully automatically in whole slide images (WSI's) of large sections, such as surgical specimens. In this paper, a method to automatically calculate and visualize the distribution of nuclear density in whole slide images of ewHCC is proposed.

A review [2] reports automated methods for detection and segmentation of nuclei for H&E stained histological images. Many of them are, however, designed for use with breast cancer, and the literatures for liver images are limited [6, 7]. Intensity-based methods, such as [8], are common for segmenting nuclei, because nuclei usually have darker colors than their background. This approach is not appropriate for hepatic nuclei, because the intensity varies both on the boundaries of nuclei and within nuclei [6], as shown in Figure 1. Wienert proposed a cell nucleus segmentation method which uses minimal a priori knowledge and is applicable for several kinds of tissues, including normal liver tissue [7]. The method uses the staining intensity to classify the objects into cell nuclei. In the case of liver tissue, nuclear fragments sometimes have light colors, as shown in Figure 1. Since some of those nuclei would be ignored by this method, there are many unextracted nuclei observed in the resultant image for liver tissue [7]. Huang proposed a method for segmenting cell nuclei and grading hepatocellular carcinoma [6]. The method, however, achieves poor segmentation when the nuclei have light colors or vague contours. Although the method can be used for grading, it would be difficult to use for counting cell nuclei and estimating the nuclear density. Processing time would also be a serious problem when applying these methods for large WSI's, because the number of image patches is upwards of tens of thousands in the case of surgical sections.

Dealing with small nuclear fragments is another serious problem for estimating the nuclear density. The thickness of H&E stained sections is usually $3 \mu\text{m}$ to $4 \mu\text{m}$, which is less than the diameter of hepatocellular nuclei ($\sim 7 \mu\text{m}$). As a result, almost all of the cell nuclei in the sections are cut. The small fragments of nuclei cut near the end (Figure 1) make it difficult to estimate the nuclear density, because it is difficult to judge whether they are actually nuclei [9]. A constant lower limit on the area has sometimes been used to remove the small objects, including nuclear fragments [7]. Using a constant value for the lower limit is not appropriate, however, because the estimated nuclear density

depends on the nuclear size, which fluctuates within the section. For example, if the nuclear size is smaller, more nuclear fragments are ignored, and less nuclear density is observed. In the proposed method, this problem has been solved by estimating the standard nuclear size and setting the lower limit based on the estimated size in each image patch. The parameters used in the method were also automatically adjusted to improve the robustness. The method was successfully applied for the WSI's of ewHCC surgical sections. The method is automatic in the sense that no manual parameter tuning is necessary.

2. Methods

2.1. Outline of the method

The method used in the diagnosis support system [5] was modified and improved to create fully automated estimation and visualization of nuclear density. The proposed method consists of pre-processing, nuclear candidate extraction, contour extraction, estimation of the standard nuclear area, selection of nuclei, and visualization. First, a WSI is divided into image patches to estimate the nuclear density. After extracting the candidate positions for cell nuclei, contours are extracted for each candidate position. Extracted contours are used to select the objects that are more likely to be nuclei. The selected nuclei are used for both estimation of the standard nuclear area and adjustment of the parameters in the method. The nuclei are then selected again from all the candidate positions to estimate the nuclear density. Finally, estimated nuclear densities are visualized as color distributions.

2.2. Lower limit for nuclear area

Since the thickness of the section is less than the diameter of the cell nuclei, all the cell nuclei are cut. As a result, observed nuclei have a variety of areas. The distribution of nuclear areas was estimated by assuming an appropriate model for the nuclear size and the section thickness. The shapes of the nuclei were assumed to be spherical, because cellular atypism of ewHCC is slight. The variation of the nuclear diameter was assumed to follow a normal distribution. Figure 2 shows the distribution

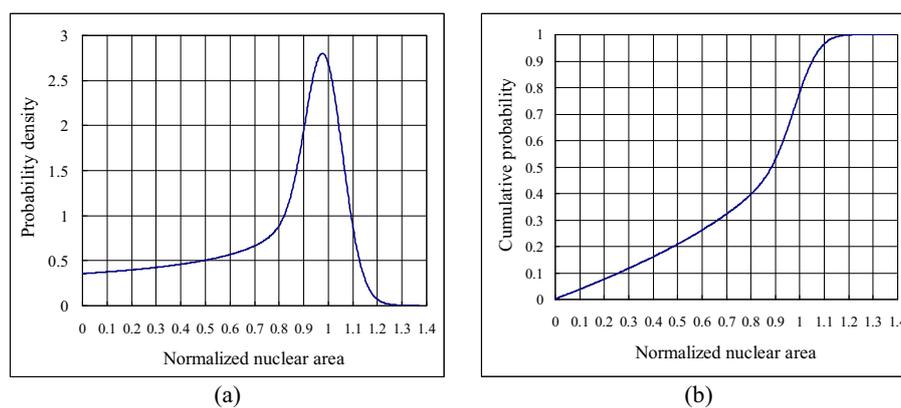


Fig. 2. Simulated distribution of nuclear areas. The nuclear area is normalized by the standard nuclear area given by $\pi \times (\text{average nuclear radius})^2$. (a) Distribution of normalized nuclear area. (b) Cumulative probability for normalized nuclear area.

of nuclear areas simulated using typical parameter values (average nuclear radius: 3.5 μm , standard deviation of nuclear radius: 0.13 μm , thickness of the section: 3.0 μm). Normalized nuclear area is the nuclear area normalized by the standard nuclear area which is given by $\pi \times (\text{average nuclear radius})^2$. As shown in Figure 2(b), about 20% of nuclear areas are smaller than half of the standard nuclear area. Therefore, defining the lower limit for the nuclear area proves to be a serious problem. If a constant value is used as the lower limit, measured nuclear density changes depending on the standard nuclear area. For a case where 75% of the standard nuclear area is used as the lower limit for a certain image patch P_1 , the nuclear density is calculated using about 65% of the nuclei, as the cumulative probability at the normalized nuclear area of 0.75 is about 35% in Figure 2(b). For another image patch P_2 whose standard nuclear area is 90% of P_1 , the lower limit used for P_1 corresponds to about 83% ($= 0.75/0.9$) of the standard nuclear area of P_2 . If the same constant lower limit is applied for P_2 , the percentage of the nuclei used for calculating the nuclear density is reduced to about 58% (cumulative probability of 42%). The nuclear density is underestimated by about 11% ($= |0.58 - 0.65| / 0.65$). In this case, 10% of change in the standard nuclear area causes about 11% of error in the nuclear density if a constant lower limit is used. The standard nuclear area was experimentally found to vary widely in a section. For example, the average and the standard deviation of the standard nuclear areas were estimated to be 40.7 μm^2 and 6.9 μm^2 , respectively, in the first WSI in section 3.3. The standard deviation corresponds to about 17% of the average. In the second WSI in section 3.3, the ratio of the standard deviation to the average was as high as 26%. Thus, a constant value is not appropriate as the lower limit. Therefore, the lower limit was decided based on the standard nuclear area estimated in each image patch. In this paper, 75% of the standard nuclear area was used as the lower limit. The method for estimating the standard nuclear area is described in section 2.6.

2.3. Pre-processing

First, a WSI is divided into a collection of image patches. Each image patch is then converted to an image with a resolution of 0.24 $\mu\text{m}/\text{pixel}$ for further processing. Next, the background color (the color of pixels without tissues) is automatically estimated using the WSI as the pixel value at the maximum peak in each histogram of R, G, and B. If the pixel value is larger than 91% of the background color in all three channels, the pixel is judged as the background. If the area of background pixels is more than 30% of the image area, the image patch is judged as background image patch and is excluded from the following processes.

2.4. Nuclear candidate extraction

The image patch is converted to a grayscale image as a weighted sum of color components. Then, the image is scaled down to 1/25 of its original size. After a Laplasian filter is applied, the image is binarized. Small areas are then removed to reduce noises, and small holes are filled up to remove the vacuoles in the nuclei. Opening and smoothing processes are applied to the resultant areas, and the positions where the pixel values have locally the peak value are extracted as the candidate positions.

2.5. Contours extraction

Contours are extracted for each extracted candidate position using the image without scaling down. Figure 3(a) shows the concept of the method.

(a) Setting the central point

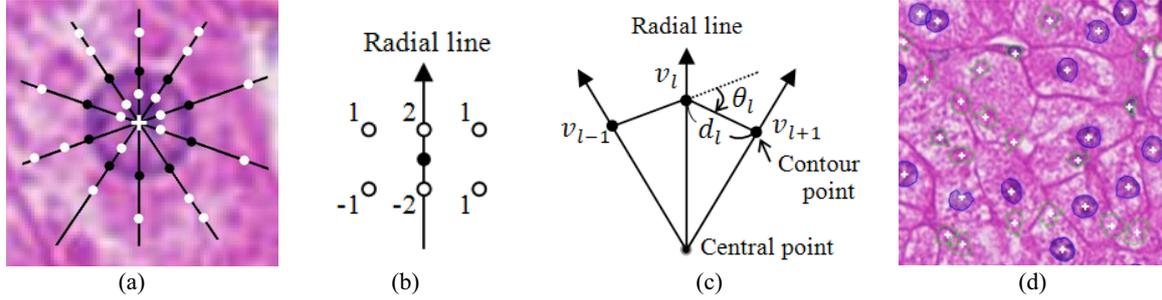


Fig. 3. Nuclear contours extraction. (a) Concept of the method (cross: central point, white dots: possible points, black dots: contour points). (b) A filter to calculate the gradient of the pixel value. (c) Distance d_l between adjacent contour points and the angle θ_l formed by adjacent contour lines. (d) Extracted contours and central point. Blue contours denote true nuclei to be detected, and green contours denote other objects or small nuclear fragments.

One point, which we call central point, is initially set at the candidate position.

(b) Setting radial lines

Radial lines are set from the central point with an even angular interval.

(c) Setting possible points

Possible points, which are the candidates for contour points, are selected for each radial line. Several positions, where gradient of image intensity in the radial direction are larger than those of other positions on the same radial line, are selected as the possible points. Since the selection is based on relative amplitude, weak edges of fragment nuclei having light colors can be detected.

(d) Selection of contour points

An initial contour point is randomly selected from the possible points for each radial line. The selection of contour points is improved so as to increase the energy function E_{total} defined in Eq. (1).

$$E_{total} = \sum_{l=1}^L (\alpha E_{img} + \beta E_{dist} + \gamma E_{ang} + \eta E_{col}) \quad (1)$$

$$E_{img} = \nabla_l I \quad (2)$$

$$E_{dist} = -\max(d_l - d_{max}, d_{min} - d_l, 0) \quad (3)$$

$$E_{ang} = \cos(\theta_l - 2\pi/L) \quad (4)$$

$$E_{col} = - \sum_{i=R,G,B} |I[i][l] - AVE_I[i]| \quad (5)$$

L is the number of contour points (radial lines), and α , β , γ and η are weight parameters. E_{img} is the energy of the image, and it is calculated as the gradient of the pixel value I in the radial direction. A Sobel-like filter shown in Figure 3(b) was used to calculate the gradient. By using this filter, unwanted effects caused by both noise and partial lack of contours can be reduced. E_{img} works so that the possible points having larger gradient value are apt to be selected as contour points. E_{dist} works so that the distance between adjacent contour points d_l falls inside or near a range $[d_{min}, d_{max}]$ and prevents the selection of wrong possible points, which sometimes have long or short contour lines. θ_l in Eq. (4) is the angle formed by two adjacent contour lines, as shown in Figure 3(c). E_{ang} works so that the contour approaches a circle, because E_{ang} becomes the maximum when $(\theta_l - 2\pi/L) = 0$. Since the cosine function makes little difference around a value of zero, this restriction on the shape is not highly strict.

The contours can be adapted to various shapes similar to a circle, as shown in Figure 3(d). In Eq. (5), $I[i][l]$ is the pixel value of i -channel ($i=R, G, B$) at the l -th contour point, and $AVE_I[i]$ is the average pixel value of all contour points for i -channel. E_{col} becomes the maximum when all the contour points have the same color. Since the contours of the same nucleus have the similar color, E_{col} prevent the selection of wrong possible points having different colors. The improvement of the contour points is repeated until convergence or for a predetermined number of iterations.

(e) Improvement of the central point

The position of the central point is updated as the center of bounding rectangle of the contour points, and steps (b) to (d) are repeated using the improved central point. This step is repeated for a predetermined number of iterations, and the best result having the largest energy is selected. The resultant contour points are interpolated to make the final contours. Figure 3(d) shows an example result.

2.6. Estimation of standard nuclear area

The area of each object (nuclear candidate) is calculated as the number of pixels on and inside the contours. Figure 4 shows the combination of the energy E_{total} and the area for all the objects in one image patch. Blue circles are true nuclei to be counted, and green triangles correspond to either other objects or nuclei whose areas are less than the threshold (75% of the standard nuclear area). Although automatically-extracted contours include wrong contours, most of the objects having large energies are true nuclei. It would be expected that most contours of those nuclei are correctly extracted. Thus, the objects having the energy of the top $R_{energy}\%$ are selected as true nuclei and are used to estimate the standard nuclear area.

Two methods were examined to estimate the standard nuclear area using the selected objects:

(a) Mode of area histogram

Figure 2(a) shows the mode of the histogram is almost the same as the standard nuclear area. Therefore, the mode of the area histogram of the selected objects was used as the standard nuclear area.

(b) Average area of the objects

Since the number of the selected objects is limited, the mode of the area histogram might fluctuate due to statistical error. The objects having the size between the top $R_{size1}\%$ and the $R_{size2}\%$ were selected, and the average of those areas was used as the standard nuclear area. This method would be more stable than the method that uses the mode.

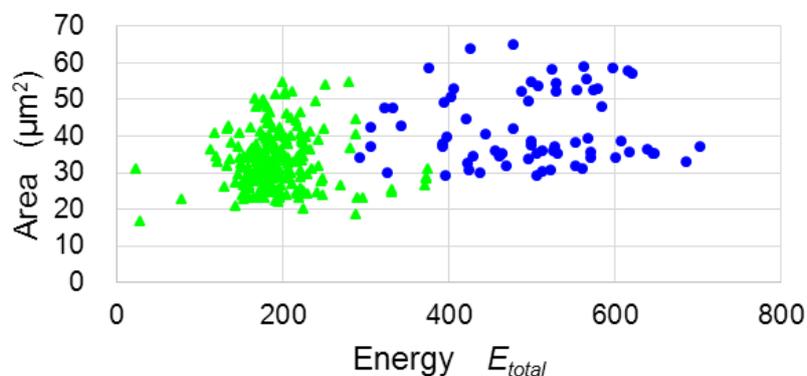


Fig. 4. Energy E_{total} and area of all candidate positions. Blue circles: true nuclei to be detected, green triangles: other objects or small nuclear fragments.



Fig. 5. Hepatocellular nuclei sampled from 20 images.

2.7. Selection of nuclei

After eliminating the objects whose area is less than the threshold (75% of the standard nuclear area), the objects whose energy is less than a threshold are also eliminated. The number of resultant objects is then divided by the area of the image patch to calculate the nuclear density. An energy E_{select} used for this selection process is the same as the energy E_{total} defined by Eq. (1), except for the normalization of the weight parameter, α . Since the energy of the image E_{img} is proportional to the edge intensity, the portion of E_{img} in the energy E_{total} is affected by the edge intensity at the nuclear contours, which may vary widely (Figure 5). In order to avoid this undesirable dependency, the weight parameter, α , was normalized for each image patch to be inversely proportional to the edge intensity at the nuclear contours. The nuclei selected for the estimation of standard nuclear area (2.6) was used to estimate the edge intensity for each image patch.

2.8. Visualization

The distribution of the nuclear density is visualized by assigning an appropriate color for each image patch depending on the nuclear density. Histogram equalization was employed for color assignment, because the variance of nuclear densities among sections is very large. The number of levels is the only parameter that can be changed by the user if desired. The same color is automatically assigned to the same number of image patches, except for background patches.

3. Experimental results

3.1. Materials

H&E stained surgical specimens of ewHCC were scanned using the digital slide scanner NanoZoomer (Hamamatsu Photonics K.K.) with a 20 \times objective lens. 20 specimens obtained from 20 patients were used in the experiment. The image size ranged from 24.2 mm \times 24.5 mm (53,248 \times 54,016 pixels) to 45.6 mm \times 24.5 mm (100,352 \times 54,016 pixels), and the image resolution was 0.455 μ m \times 0.455 μ m. Each WSI was divided into about 7000 to 13,400 image patches (635 \times 635 pixels), and each image patch was converted to 1200 \times 1200 pixels (0.24 μ m/pixel).

To evaluate the accuracy, two image patches, one from ewHCC and one from non-cancer, were randomly selected from the hepatocyte region for each of 20 WSI's. 5094 hepatocyte nuclei from 40 images were labeled by the authors under the supervision and check of the pathologist (M. Nakano), who is a specialist in liver pathology. Figure 5 shows typical nuclei of hepatocytes sampled from 20 images. These nuclei were successfully extracted in the experiments. The experimental data set includes nuclei having a variety of color, contrast, nuclear size, shape, and textures.

3.2. Evaluation

The accuracy of the method was evaluated using the 40 images by 4-fold cross validation. The mean

Table 1
Mean absolute percentage error of nuclear density

Method	Mean absolute percentage error
Using “Average area of objects”	8.2 %
Using “Mode of area histogram”	13.3 %
Without normalization of weight parameter	13.8 %
Constant standard nuclear area	12.2 %

absolute percentage error (the average of $|\text{measured density} - \text{correct density}| / \text{correct density}$) was evaluated for test images using the parameters optimized for training images. The result is shown in Table 1. As the method to estimate the standard nuclear area (2.6), “Average area of objects” showed better result of 8.2% than “Mode of area histogram”. Thus, “Average area of objects” was used in other experiments. Concerning the improvements employed in the energy E_{select} , the error was degraded to 13.8% if the normalization of the weight parameter α was excluded, showing the effectiveness of the parameter normalization. The accuracy was also evaluated using a constant value for the standard nuclear area. Although the constant value was optimized in order to minimize the mean absolute percentage error, the error was degraded to 12.2%. Therefore, estimating and using the standard nuclear area was shown to be effective.

3.3. Visualization of nuclear density

The proposed method was applied to all the image patches, and it was confirmed that the distribution of the nuclear density in each WSI was successfully visualized. Figures 6(b) and 6(d) show two samples of visualized density distributions where the number of levels is set to the default value of 7. In the case of Figures 6(a) and (b), the physical dimension of the WSI is 35.3 mm \times 24.5

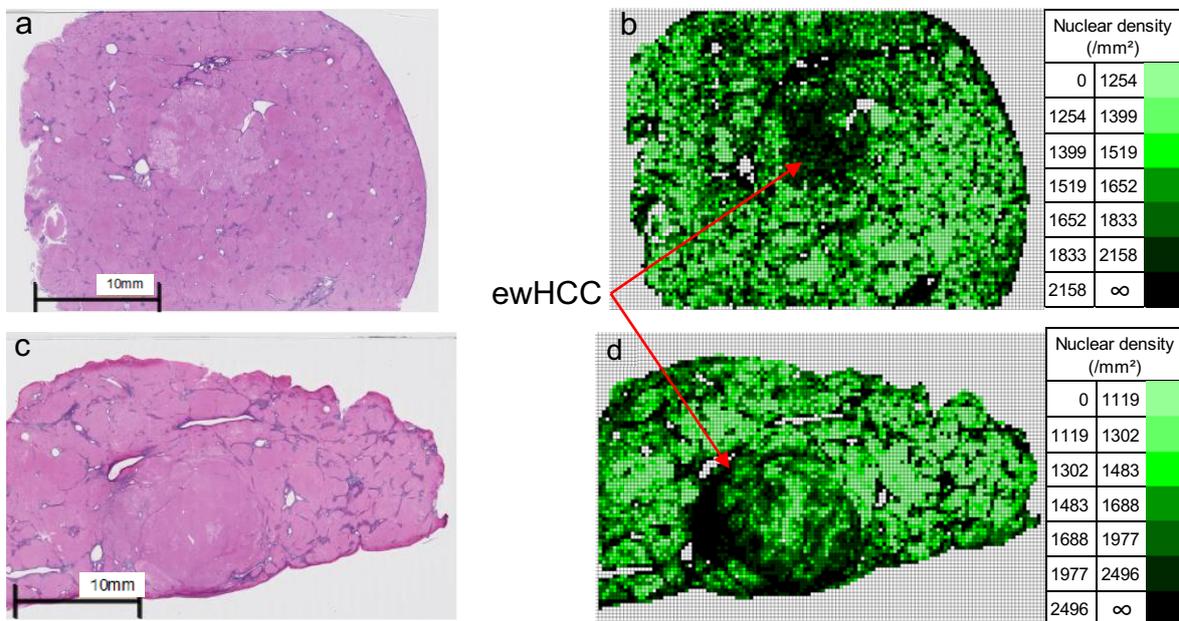


Fig. 6. (a) (c) WSI's and (b) (d) visualized distribution of their nuclear densities, respectively.

mm (77,824×54,016 pixels). The WSI was divided into 10,578 (=123×86) image patches. 1483 image patches were judged as the background, and the nuclear densities of the resultant 9095 image patches (754 mm²) were estimated. The number of nuclei counted was about 1,280,000. The calculation time for this WSI was about 57 minutes on a PC (Intel Core i7-3930K, 6 cores, 12GB memory).

In both sample results of Figures 6(b) and (d), ewHCC regions have darker colors compared to their surroundings. Some other dark regions, mainly at outer regions, might be present due to artifacts. Visualized distribution may help the user easily notice the areas having nuclear densities different from others. In Figure 6(d), the central region of the ewHCC nodule has light colors similar to the background non-cancerous area. If the ROI's are selected from this central region, the difference of the nuclear density would be very small. Therefore, visualizing the nuclear distribution is important to relieve this sampling problem.

4. Conclusion

A method for estimation and visualization of nuclear density in whole slide images of H&E stained hepatic histological sections is presented. The method is automatic and computationally efficient. A WSI with the tissue area of 754 mm² was processed in about 57 minutes. Experimental results using 40 image patches sampled from ewHCC and non-cancerous regions showed good accuracy. Future work will include extraction and visualization of other features such as morphological features. Applying the method for more advanced cancer would be promising. Reduction of the computation time would also be possible by optimizing the program.

Acknowledgment

This work was supported in part by KAKENHI 24500514.

References

- [1] M.N. Gurcan, L.E. Boucheron, A. Can, A. Madabhushi, N.M. Rajpoot and B. Yener, Histopathological image analysis: A review, *IEEE Reviews in Biomedical Engineering* **2** (2009), 147-171.
- [2] H. Irshad, A. Veillard, L. Roux and D. Racoceanu, Methods for nuclei detection, segmentation, and classification in digital histopathology: A review—current status and future potential, *IEEE Reviews in Biomedical Engineering* **7** (2014), 97-114.
- [3] M. Nakano, A. Saito, M. Doi and K. Takasaki, Stromal and blood vessel wall invasion in well-differentiated hepatocellular carcinoma, *Liver* **7** (1997), 41-46.
- [4] S. Nakajo, M. Yamamoto and E. Tawara, Measurement cytology, base and clinical, advancement of image cytometry, form measurement academic analysis in hepatocellular carcinoma, about the utility of nuclear density and nuclear density ratio, *Japan clinical* **50** (1992), 2525-2529.
- [5] M. Takahashi, M. Nakano, M. Tomita and K. Matsunaga, Support system for histopathologic diagnosis of hepatocellular carcinoma using nuclear positions, *Proceeding of the 2nd IASTED International Conference on Biomedical Engineering (BioMED2004)* **1** (2004), 417-123.
- [6] P.W. Huang and Y.H. Lai, Effective segmentation and classification for HCC biopsy images, *Pattern Recognition* **43** (2010), 1550-1563.
- [7] S. Wienert, D. Heim, K. Saeger, A. Stenzinger, M. Beil, P. Hufnagl, M. Dietel, C. Denkert and F. Klauschen, Detection and segmentation of cell nuclei in virtual microscopy images: A minimum-model approach, *Scientific Reports* **2** (2012), 1-7.
- [8] S. Liu, C. Yin, Z. Kiu, Z. Zhang, J. Gao, M. Zhu, J. Gu and K. Xu, Microscopic image analysis and recognition on

- pathological cells, Proceedings of 20th Canadian Conference on Electrical and Computer Engineering **1022** (2003), 1022-1025.
- [9] M. Abercrombie, Estimation of nuclear population from microtome sections, The Anatomical Record **94** (1946), 239-247.