

# Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

## Automatic quantification of morphological features for hepatic trabeculae analysis in stained liver specimens

Masahiro Ishikawa  
Yuri Murakami  
Sercan Taha Ahi  
Masahiro Yamaguchi  
Naoki Kobayashi  
Tomoharu Kiyuna  
Yoshiko Yamashita  
Akira Saito  
Tokiya Abe  
Akinori Hashiguchi  
Michiie Sakamoto

Masahiro Ishikawa, Yuri Murakami, Sercan Taha Ahi, Masahiro Yamaguchi, Naoki Kobayashi, Tomoharu Kiyuna, Yoshiko Yamashita, Akira Saito, Tokiya Abe, Akinori Hashiguchi, Michiie Sakamoto, "Automatic quantification of morphological features for hepatic trabeculae analysis in stained liver specimens," *J. Med. Imag.* 3(2), 027502 (2016), doi: 10.1117/1.JMI.3.2.027502.

# Automatic quantification of morphological features for hepatic trabeculae analysis in stained liver specimens

Masahiro Ishikawa,<sup>a,b,\*</sup> Yuri Murakami,<sup>a</sup> Sercan Taha Ahi,<sup>a</sup> Masahiro Yamaguchi,<sup>a</sup> Naoki Kobayashi,<sup>b</sup> Tomoharu Kiyuna,<sup>c</sup> Yoshiko Yamashita,<sup>c</sup> Akira Saito,<sup>d,e</sup> Tokiya Abe,<sup>f</sup> Akinori Hashiguchi,<sup>f</sup> and Michiie Sakamoto<sup>f</sup>

<sup>a</sup>Tokyo Institute of Technology, Interdisciplinary Graduate School of Science and Engineering, 4259, Nagatsuta, Midori-ku 226-8503, Japan

<sup>b</sup>Saitama Medical University, Faculty of Health and Medical Care, 1397-1 Yamane, Hidaka-shi Saitama 350-1241, Japan

<sup>c</sup>NEC Corporation, Medical Solutions Division, 5-7-1 Shiba Minato-ku, Tokyo 108-8001, Japan

<sup>d</sup>Tokyo Medical University, Quantitative Pathology and Immunology Division, 6-1-1 Shinjuku, Shinjuku-ku Tokyo 160-8402, Japan

<sup>e</sup>Tokyo Medical University, Molecular Pathology Division, 6-1-1 Shinjuku, Shinjuku-ku, Tokyo 160-8402, Japan

<sup>f</sup>Keio University School of Medicine, Department of Pathology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan

**Abstract.** This paper proposes a digital image analysis method to support quantitative pathology by automatically segmenting the hepatocyte structure and quantifying its morphological features. To structurally analyze histopathological hepatic images, we isolate the trabeculae by extracting the sinusoids, fat droplets, and stromata. We then measure the morphological features of the extracted trabeculae, divide the image into cords, and calculate the feature values of the local cords. We propose a method of calculating the nuclear–cytoplasmic ratio, nuclear density, and number of layers using the local cords. Furthermore, we evaluate the effectiveness of the proposed method using surgical specimens. The proposed method was found to be an effective method for the quantification of the Edmondson grade. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License.

Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JMI.3.2.027502](https://doi.org/10.1117/1.JMI.3.2.027502)]

Keywords: quantitative pathology; segmentation; hepatic trabecula; hematoxylin and eosin-stained specimens.

Paper 15211PRR received Nov. 4, 2015; accepted for publication May 12, 2016; published online Jun. 3, 2016.

## 1 Introduction

Recently, digital image technologies have been introduced to pathology diagnosis, especially because of the great progress of whole slide imaging (WSI) systems.<sup>1</sup> WSI enables the observation of a microscopic image on a computer display while interactively changing the viewing field and magnification. Histopathology samples can then be handled as digital data instead of glass slides. As a result, image analysis-based applications are expected to support diagnosis in pathology. Although image analysis and pattern recognition techniques have been applied to histopathology before now, their clinical application was difficult because the diagnosis was done using the microscopic observation of glass slides, and digital image analysis required an additional digitizing process. However, because of the introduction of WSI technology, it is now possible to deploy image analysis tools in clinical use. This will allow quantitative evaluation rather than the qualitative or semi-quantitative analyses that have been done in conventional visual assessment. Digital image analysis will support a pathologists' decision with quantitative information obtained from the histopathology samples, enhancing the objectivity and reproducibility of the diagnosis.

In this paper, computer image analysis is applied to the quantitative assessment of hepatocellular carcinoma (HCC)—the most common histological type of primary liver cancer. The

histologic grade of HCC is usually assessed by the Edmondson–Steiner grading (Edmondson grading) system, which classifies cancers as groups 1–4 (G1, G2, G3, and G4). Categories G1 through G4 indicate the most and least differentiated neoplastic tissue, respectively. A higher grade indicates a higher malignancy of the cancer. As the morphology of low-grade cancer resembles noncancer tissue, its identification is a difficult task, and quantitative measurement would be helpful for objective and reproducible diagnosis. The modern histopathology practice mainly comprises the morphological assessment of cells and tissues and molecular analyses. Morphological assessment is primarily important in diagnostic pathology, and it is carried out using tissue sections stained with hematoxylin and eosin (HE). HE staining is a routinely used method for morphological assessment; it stains nuclei blue and cytoplasm, fiber, and other extracellular substances pink. Other special stains are also used for specific purposes. On the other hand, immunohistochemistry (IHC) staining is a technique especially valuable for molecular pathology. Digital image analysis has been applied to HE-stained, special-stained, and IHC-stained tissues.<sup>2</sup> In this paper, HE-stained tissue sample images are analyzed.

In the image analysis of HE-stained sections, the cell nuclei are extracted and their features such as area, perimeter, and ellipticity are analyzed.<sup>3,4</sup> The accuracy of nuclei extraction has been considerably improved<sup>5</sup> in recent years and it has been shown that the nuclear features automatically extracted by the image analysis are meaningful for grading and prognosis. In addition to the morphology of nuclei, tissue structure features are important in histopathology diagnosis. Structural features such as the

\*Address all correspondence to: Masahiro Ishikawa, E-mail: [ishikawa@saitama-med.ac.jp](mailto:ishikawa@saitama-med.ac.jp)

gland or lumen are used in the automated image analysis of prostate and breast cancers;<sup>6–8</sup> however, they are also important in the diagnosis of liver cancer.<sup>9,10</sup>

In HCC diagnosis,<sup>11,12</sup> the cord-like structure of the hepatocytes called the trabecula is of high importance. Normal trabeculae are one to two cells thick. Cancer cells mimic the normal hepatocytes and can arrange themselves into trabecular patterns that simulate normal liver trabeculae. While normal liver trabeculae radially extend from the central veins of the liver, the trabecular patterns of HCC become disordered and form a multi-layered structure as the trabeculae become thicker. However, as the color and texture of liver tissue widely varies among specimens, the cellular area containing the trabeculae is difficult to extract. Very few studies have been reported on the quantification of trabeculae. Kiyuna et al.<sup>13</sup> calculated six statistical distributions to serve as trabecular thickness features, which consisted of 10, 30, 50, 70, and 90 percentiles as well as the standard deviation. Although they quantified the thickness of trabeculae, they were not able to quantify the number of layers.

Here, we propose a technique that extracts the trabecular regions with high accuracy and quantifies the structural features such as the trabecular thickness using HE-stained liver specimens. We focus on the features useful for discriminating grades G1 and G2 cancers from noncancer because of the difficulty in identifying well-differentiated HCCs. After an overview of the methodology for the quantitative analysis of HCC tissue samples in Sec. 2, Sec. 3 describes the improved method for trabecular extraction that was originally proposed in Ref. 14. In Sec. 4, the method for quantifying trabecular regions is presented. The effectiveness of the proposed method is experimentally demonstrated in Sec. 5, and Sec. 6 presents the conclusions.

## 2 Overview of Hepatocellular Carcinoma Quantitative Analysis

An HE-stained image of a liver tissue specimen is shown in Fig. 1(a). The hepatocyte nuclei are stained violet-blue by hematoxylin, and the cytoplasm is stained pink. Lymphocytes are also stained blue by hematoxylin; however, they are distinguishable from the hepatocyte nuclei by their darker shade. Fibers are stained pink by eosin; however, the staining is a lighter shade than that of the cytoplasm. Sinusoids and fat droplets generally appear as white areas. A trabecula is a series of cells segmented by sinusoids and stromata. A manual extraction of the trabecula in Fig. 1(a) is shown in Fig. 1(c). In this figure,

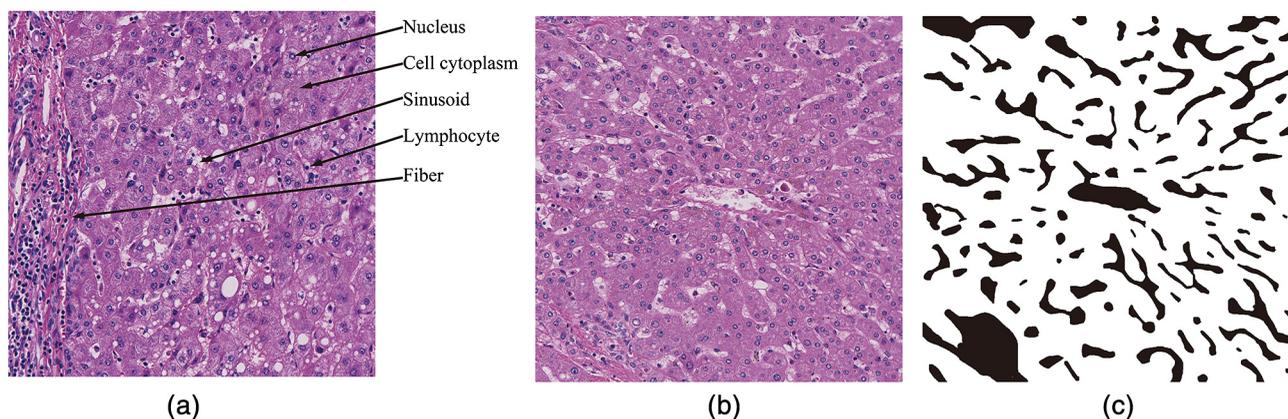
the white pixels indicate the trabecula regions and the black pixels indicate the other regions.

Figure 2 shows examples of trabecular differentiation in tissues of noncancer tissue and HCCs of different grades. In the tissues of higher grade HCC, the trabeculae thickness increases and the nuclei display various shapes. The trabecular thickness is known as a morphological feature that characterizes HCC. To quantitatively measure trabecular structure, it is necessary to segment the hepatocytes. However, as cytoplasm texture depends on the conditions of the cells (e.g., fatty metamorphosis or cancerous change) and specimens (staining time or specimen fixation), the extraction of trabeculae is a difficult task.

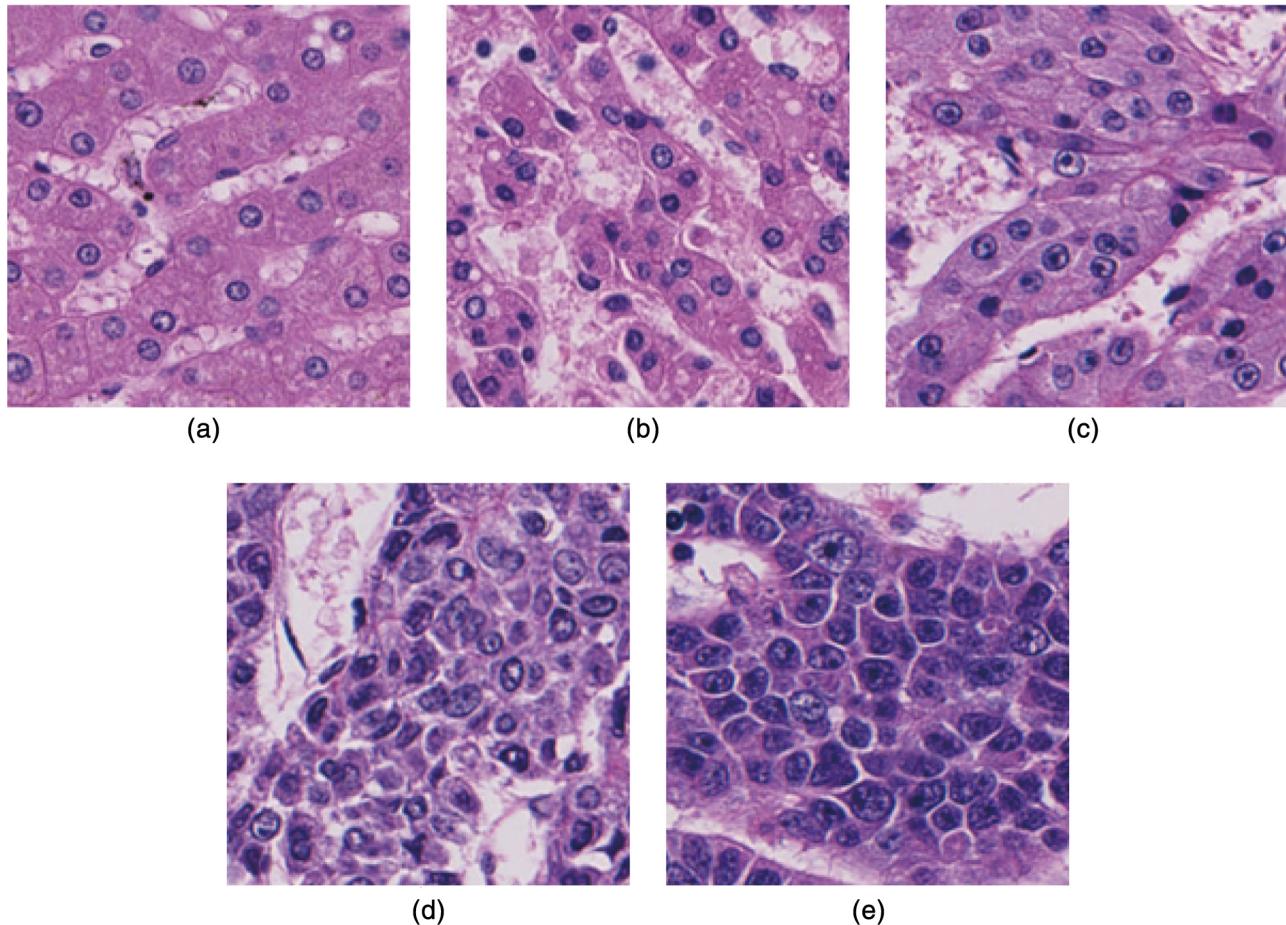
Figure 3 shows an overview of the proposed method for the extraction and quantification of trabecular structure. We previously reported that trabeculae can be extracted using the structures outside liver cells (sinusoids and stromata),<sup>14,15</sup> which can be extracted more easily than liver cells themselves. The method introduced in this paper also follows the same approach: the sinusoid and stroma regions are segmented first, and then we extract the trabeculae regions by excluding those regions. In addition, fat droplets are extracted and regarded as part of the liver cells<sup>16</sup> because fat droplets reside inside the cytoplasm. This process improves the trabeculae extraction results. Furthermore, the nuclei of liver cells are extracted because we use information about the nuclei when quantifying the trabeculae.

The proposed method addresses the problems that remained in the previous report:<sup>14</sup> (1) the erroneous extraction of fat droplets in the sinusoid extraction and (2) the erroneous extraction of lymphocytes in the nuclei extraction. The extraction accuracy of trabeculae is improved by solving problems (1) and (2). The extraction of fat droplets has been reported only in a few studies. Among these studies, Katoh et al.<sup>17</sup> extracted fat droplets to evaluate the risk HCC, but the average accuracy of their extraction was relatively low (77%). Turlin et al.<sup>18</sup> proposed an assessment of shepatic steatosis; however, this method is a semi-quantitative method. Lee et al.<sup>19</sup> proposed that HCC risk indexes could be established from fat droplets; however, they did not clarify the details. In addition, to solve the problem (1), we applied the fat droplet extraction method developed in one of our earlier studies.<sup>16</sup>

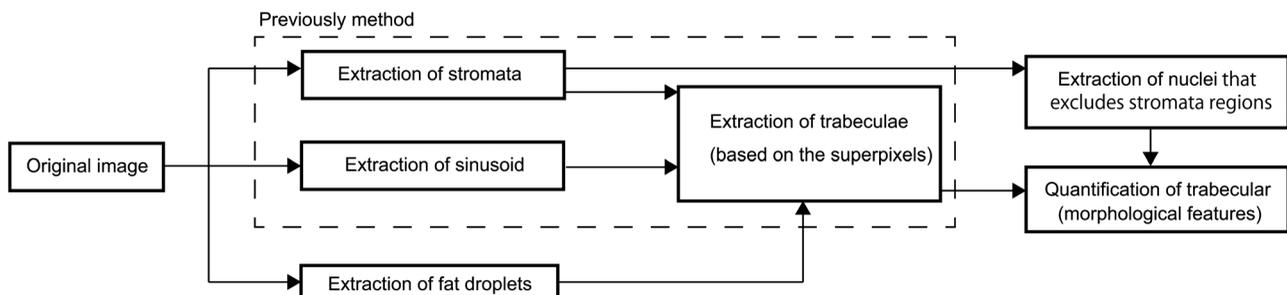
In addition, we also propose a quantification method for trabecular structure and verify its effectiveness. To quantify trabecular structure, it is necessary to consider the fact that the trabecular thickness varies locally. Hence, we propose a method



**Fig. 1** (a) HE-stained specimen of liver tissue (20× magnification), (b) normal tissue image, and (c) manually extracted trabeculae image.



**Fig. 2** Differentiation of trabeculae in Edmondson-graded tissues: (a) noncancerous and (b)–(e) Grade 1–4 tumors, respectively.



**Fig. 3** Overview of the proposed method.

to split a cell cord into small areas and quantify the structural features of the trabecular thickness to determine the local trabecular thickness. In the method in this paper, we target grades G1 and G2, since the diagnosis of G3 and G4 HCCs is relatively easy.

### 3 Automatic Segmentation of Trabecular Regions

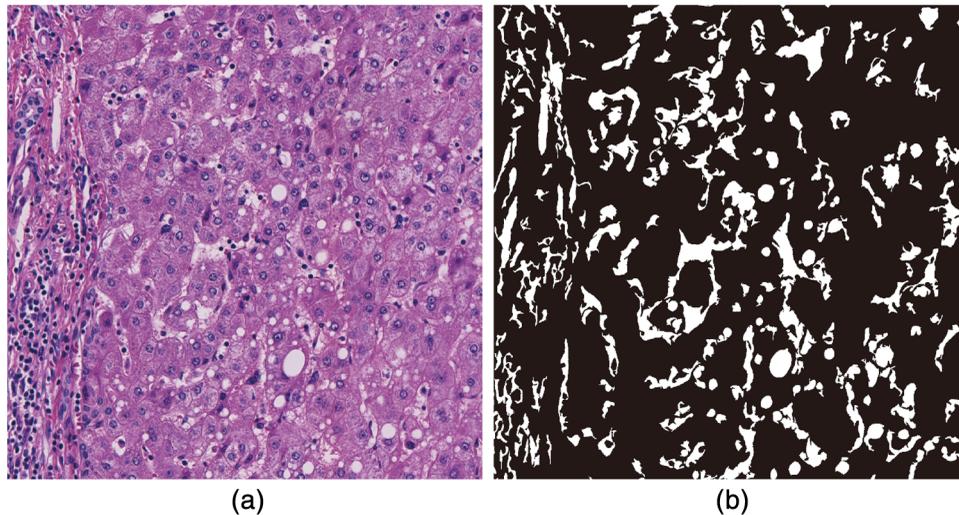
The proposed method illustrated in Fig. 3 is composed of two main steps: trabecular segmentation and trabecular feature measurement. This section presents the trabecular segmentation steps: stroma and sinusoid extraction<sup>14,15</sup> are presented in Sec. 3.1, the methods for extracting of fat droplets are described in Sec. 3.2, and trabecular extraction is explained in Sec. 3.3.

#### 3.1 Trabeculae Extraction

Trabecula extraction is thought to be difficult because liver cells take on various colors and appearances depending on tissue conditions. Therefore, we propose an approach to extract trabeculae from an HE-stained liver tissue slide by extracting the rest of the tissue, i.e., the sinusoids and stromal areas.

##### 3.1.1 Sinusoid extraction

We first apply an orientation-selective (OS) filter to improve the contrast of the sinusoid boundaries. We then apply the expectation-maximization (EM) algorithm to cluster bright pixels. Figure 4(a) shows the original image, and Fig. 4(b) shows the result of sinusoid extraction. The effectiveness of the OS



**Fig. 4** (a) Original HE-stained liver tissue image ( $1024 \times 1024$ ) and (b) result of sinusoid extraction. In (b), the white pixels indicate the sinusoid regions and the black pixels indicate the other regions.

filter is shown in Ishikawa et al.,<sup>15</sup> where details of this step can be found.

### 3.1.2 Stroma extraction

Stroma extraction consists of two steps. In the first step, we calculate the fiber probability of each pixel using color and texture features. The color and texture features are calculated over the surrounding  $64 \times 64$  region of each pixel. Color features consist of the (1) average, (2) variance, (3) skewness, and (4) kurtosis acquired from the histogram of each color channel, which corresponds to  $4 \text{ features} \times 3 \text{ color channels} = 12 \text{ features}$ . The texture features consist of the (1) angular second moment, (2) contrast, (3) correlation, and (4) homogeneity of Hararick's texture features, calculated from the gray-level co-occurrence matrix (GLCM). The pixel brightness values are normalized to the range of 0 to 1 and are quantized into eight levels.

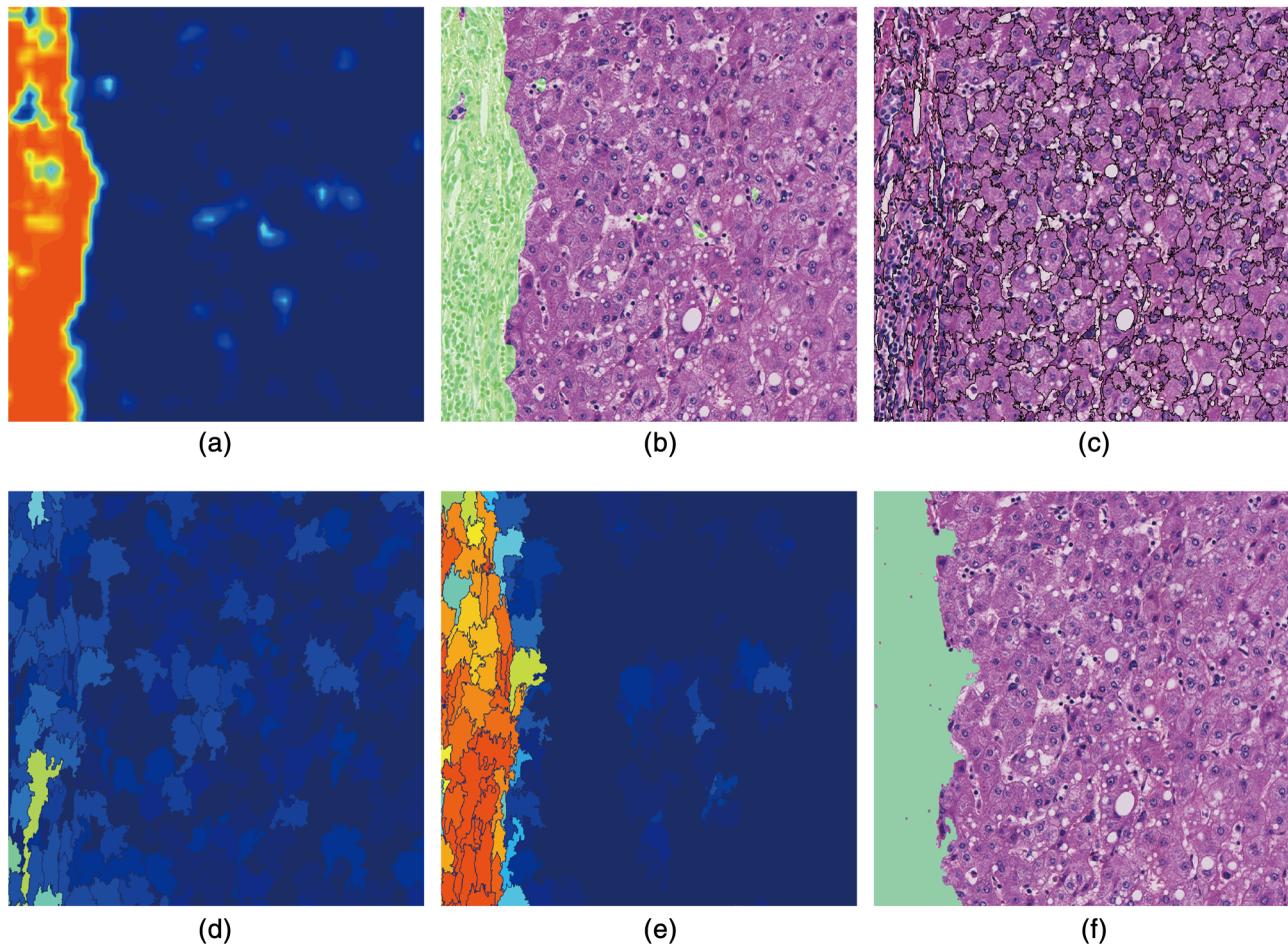
For the GLCM-based features, extraction was performed using a four-direction GLCM (0, 45, 90, and 135 deg) with a distance of 1. The four statistical quantities, calculated from the four co-occurrence matrices using different direction offsets, are summarized by two quantities, the mean and variance. Thus, the final texture feature is a 24-dimensional feature ( $\text{mean} \times \text{variance} \times 4 \text{ statistical features} \times 3 \text{ color channel} = 24 \text{ features}$ ). The fiber probability is calculated for each pixel using a linear support vector machine based on the 36-dimensional color and texture features. We constructed our learning data set using 4000 image patches of  $64 \times 64$  pixels, half of which were fiber regions. The remaining half included no fiber at all. An example of pixel-wise fiber probability indicated by pseudocolor is shown in Fig. 5(a). Figure 5(b) shows the stroma area in green, extracted by a threshold of 0.3. Basically, we have successfully extracted the stroma area. However, over-detection has occurred in the cytoplasm and under-detection has occurred in the stroma. A larger amount of stroma has been detected in the border between the stroma and the cytoplasm, and hence the border is not accurate. To solve this problem, we apply the next step.

In the second step, we extract stroma based on the "superpixel" image processing technique. We generated superpixels using simple linear iterative clustering (SLIC). SLIC is a fast and memory-efficient adaptation of the K-means algorithm to

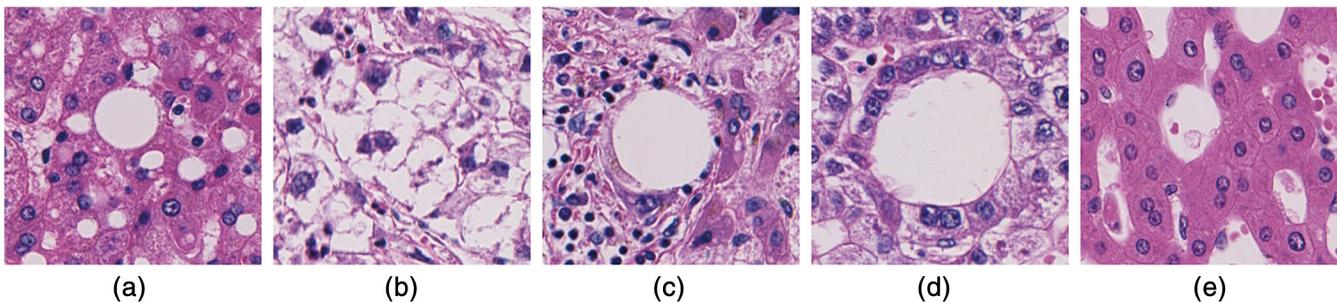
generate superpixels. Details can be found in Ref. 20. Figure 5(c) shows the SLIC result. Stroma identification is performed per superpixel using fiber probability and the number of lymphocytes. We empirically determined the thresholds for fiber probability and number of lymphocytes. Specifically, we determined the optimal threshold by applying various thresholds to a large number of images. The purpose of stroma extraction is to improve the trabeculae extraction. Over-detection in stroma extraction is a bigger problem than under-detection. Therefore, superpixels with a mean fiber probability of more than 0.3 are extracted as stroma. Additionally, superpixels containing more than seven lymphocytes are extracted as stroma. Lymphocytes can be extracted by template matching. In this paper, we prepared a circle template with a diameter of  $7.3 \mu\text{m}$ . Figure 5(d) shows a pseudocolor representation of the number of lymphocytes per superpixel, and Fig. 5(e) shows the pseudocolor representation of mean fiber probability per superpixel. Higher probability is indicated by warm colors, and lower probability is indicated by cool shades. Figure 5(f) shows the extracted stroma area in green. Compared with Fig. 5(b), the errors of Fig. 5(f) with respect to over-detection and the boundaries of the stroma and cytoplasm are reduced.

### 3.2 Fat Droplet Extraction

The storage of excess glycogen in the liver cytoplasm is a pathological condition known as fatty liver. The fat droplets in an HE-stained liver specimen are shown in Fig. 6(a). However, liver tissue contains many structures that resemble fat droplets such as clear cells [Fig. 6(b)], hepatic arteries [Fig. 6(c)], pseudoglandular patterns [Fig. 6(d)], and sinusoids [Fig. 6(e)]. Generally, fat droplets reside inside the liver cytoplasm and should be regarded as cytoplasm when quantifying trabeculae. In this study, fat droplets were extracted and classified into the same region as the cytoplasm. Figure 4(a) shows a trabeculae image that includes fat droplets. The extracted result using the method of Sec. 3.1 are shown in Fig. 4(b). In this image, even fat droplets have been extracted as sinusoid areas, causing the trabeculae extraction to fail. This is because it is difficult to distinguish sinusoids, which are circular [as shown in Fig. 2(e)], from fat droplets. In addition, the sinusoid extraction method



**Fig. 5** (a) Fiber probability, (b) extraction result using fiber probability (threshold = 0.3), (c) SLIC result, (d) number of lymphocytes per superpixel, (e) mean fiber probability per superpixel, and (f) automatically extracted stroma region (indicated by green). For images (a), (d), and (e), higher values are represented by warm colors, and lower probabilities are represented by cool shades.



**Fig. 6** Typical appearance of fat droplets and similar structures in HE-stained liver specimens: (a) fat droplets, (b) clear cells, (c) an arterial vein, (d) a pseudoglandular structure, and (e) sinusoids.

discussed in Sec. 3.1.1 has a problem in that sinusoids and fat droplets are combined, as shown in Fig. 4(b), because this method classifies a red-green-blue (RGB) image into three classes and extracts the area of the brightest class. Therefore, we extracted fat droplets to reduce such detection errors. The fat droplet extraction was carried out by a method proposed for fatty metamorphosis quantification in Ref. 15. Fat droplet extraction consists of four steps: (1) smoothing of the fat droplet boundaries using an OS filter, (2) extraction of candidate fat droplet regions, (3) fat droplet discrimination, and (4) modification of the trabeculae extraction results using fat droplet regions.

### 3.2.1 Preprocessing

The lower-left boundary of the large fat droplet shown in the center of Fig. 7(a) has interrupted parts on its boundary. Such an interrupted boundary, which occurs on some droplets, affects the results of the circularity calculation. OS filtering is a sinusoid extraction preprocessing step because it connects the boundaries of a sinusoid. For the same reason, the OS filter is effective for fat droplet extraction. Therefore, we adopt an OS filter to smooth the fat droplet boundaries (step 1). The result of the OS filter is shown in Fig. 7(b). To clarify its effects,

Figs. 7(a) and 7(b) were binarized at a threshold of 175 and the results are shown in Figs. 7(c) and 7(d), respectively. It can be observed that the boundaries are clear when the OS filter is applied.

### 3.2.2 Extracting candidate fat droplet regions

We next extract the candidate fat droplet regions using an EM algorithm (step 2). The fat droplet basically looks whiter than the sinusoid because its fat and lipids liquidate during the process of creating a paraffin section for the slide, leaving a completely hollow space inside. Therefore, we classify the RGB image into 10 classes using the EM algorithm and extract images in the brightest class as fat droplet candidate regions.

### 3.2.3 Specific of fat droplets

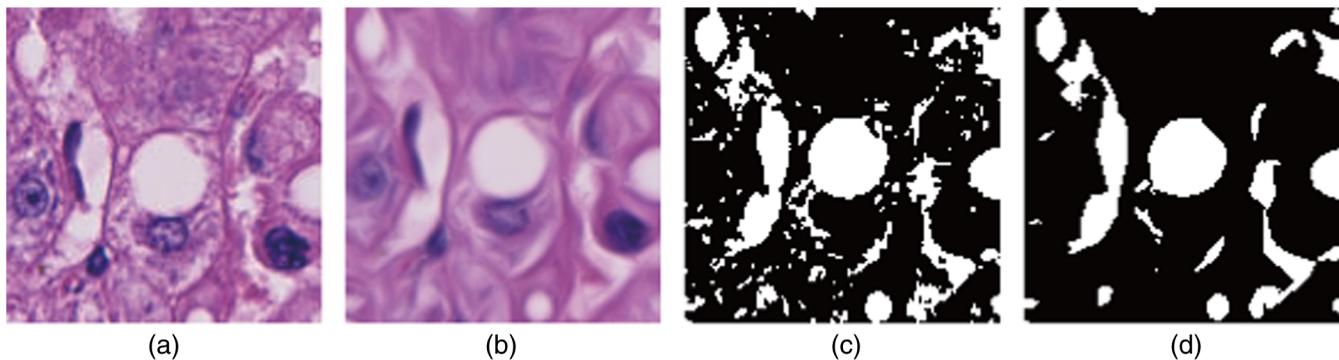
We next label the results of extraction and distinguish fat droplets from sinusoids and other structure in each labeled area. Fat candidate areas are classified by a random forest (step 3). As to feature values, circularity, eccentricity, area, complexity, clockwise turning ratio, outline concentration ratio, turning angle median value, derivative matching feature, fractal dimension, number of edge pixels, contrast, intensity mean, intensity variance, outside intensity mean, outside variance of intensity,

outside contrast, number of nuclei, area change ratio, and outside number of nuclei were used.

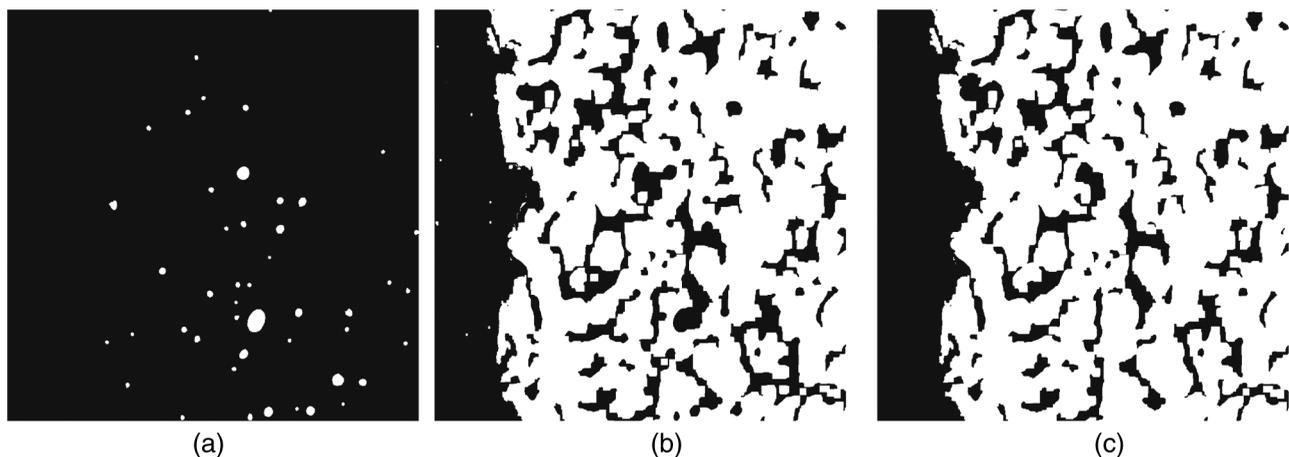
Figure 8(a) shows the final detected fat droplet regions. Finally, we modify trabeculae extraction results using the fat extraction results (step 4). We correct the cytoplasm results by excluding the fat droplet areas extracted from the sinusoidal areas. To be more specific, the fat areas of Fig. 8(a) are used to exclude the detected fat droplets in the sinusoidal area of Fig. 8(b). The corrected results using this proposed method is shown in Fig. 8(c).

### 3.3 Trabecular Extraction

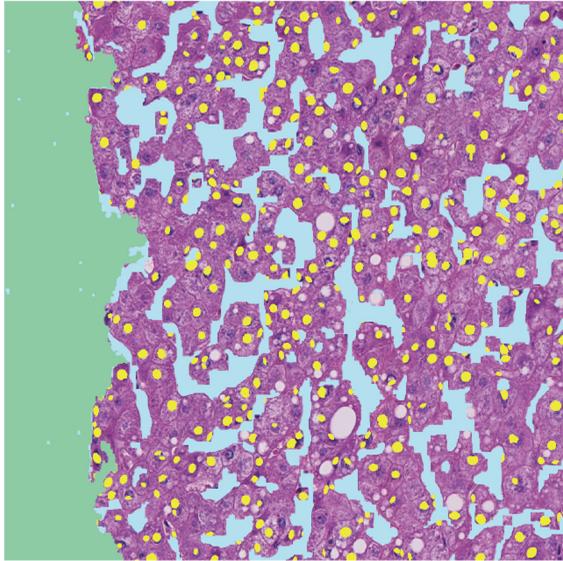
The trabecular area is extracted as the regions that remain after the sinusoids, stroma, and glass regions have been excluded, and the regions of fat droplets are included. Figure 9 shows the result of trabecular extraction. Figure 9 shows the extracted stroma area in green and the extracted sinusoidal area in blue. In the proposed system shown in Fig. 3, nuclear extraction is required. The proposed method adopts the method proposed by Kiyuna et al.<sup>13</sup> Figure 9 shows the extracted nucleus areas in yellow. The proposed method effectively reduces false nucleus detection by excluding the extracted fiber and sinusoidal areas. As a result, it becomes possible to quantify trabeculae.



**Fig. 7** Results of OS filter: (a) fat droplets, (b) result of OS filter, (c) binary image of (a), and (d) binary image of (b).



**Fig. 8** Sinusoid extraction in tissue that includes fat droplets: (a) result of fat droplet extraction, (b) result of trabecular extraction after extracting the fibers and sinusoids, and (c) result of trabecular extraction after extracting the fibers, sinusoids, and fat droplets. In (a)–(c), the white pixels indicate the target regions (fat droplets and trabeculae) and the black pixels indicate the other regions.



**Fig. 9** Automatically extracted trabeculae. Green regions indicate the fiber regions, blue regions indicate the sinusoid areas, and yellow regions indicate the nucleus areas.

#### 4 Automatic Quantification of Trabecular Features

This section describes the method for trabecula quantification; the morphological features of the trabecular structure are measured using the hepatic trabeculae segmented by the method of the previous section. These features include the number of layers (number of cells in thickness) containing the trabeculae of a local shape and the nuclear–cytoplasmic (N/C) ratio, which is a histological feature of the tissue. Measurements are taken along the centerline of the segmented trabeculae, as explained in the following sections.

##### 4.1 Extraction and Division of Trabeculae Centerlines

To compute the features related to the trabecular structure, the centerline of the trabeculae is extracted by a morphological

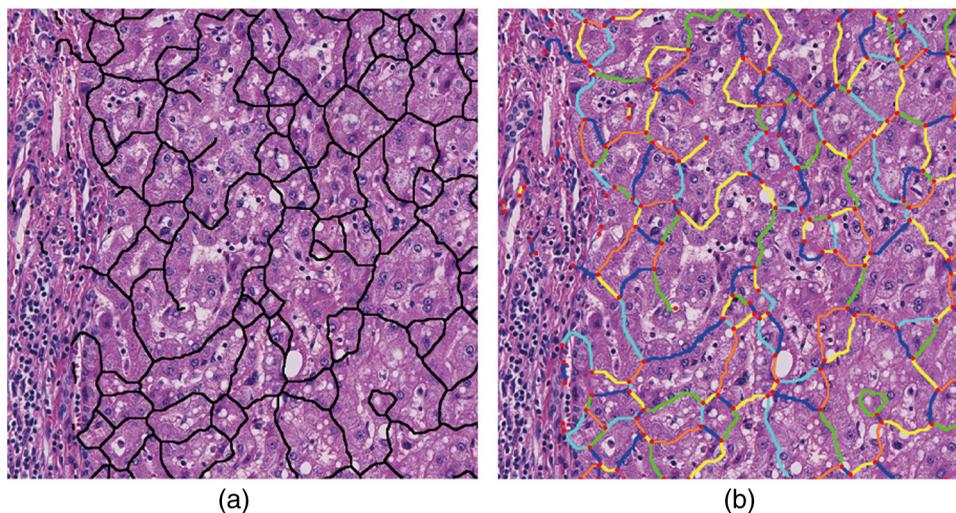
thinning operation.<sup>21</sup> To minimize false branches during the thinning process, branches shorter than a certain threshold (in the following experiment, 20 pixels from their tips to the branching point) are deleted. Furthermore, to reduce the erroneous extraction of trabeculae containing endothelial cells and fibers within their sinusoids, the boundaries of the trabeculae areas are smoothed by a preprocessing morphological operation. The centerline is therefore segmented at the branches or at the end-points. A trabecula label is assigned to each segment of the centerline, and the features of the trabecula are measured and associated with each cord label.

An example of centerline extraction is shown in Fig. 10(a). Figure 10(b) shows a pattern diagram for each cord label. Branches are marked in red and the centerlines are depicted in colors that are randomly selected for each cord label. In this paper, we propose to calculate the features for each cord label. This is because the number of layers is different from place to place even in one image.

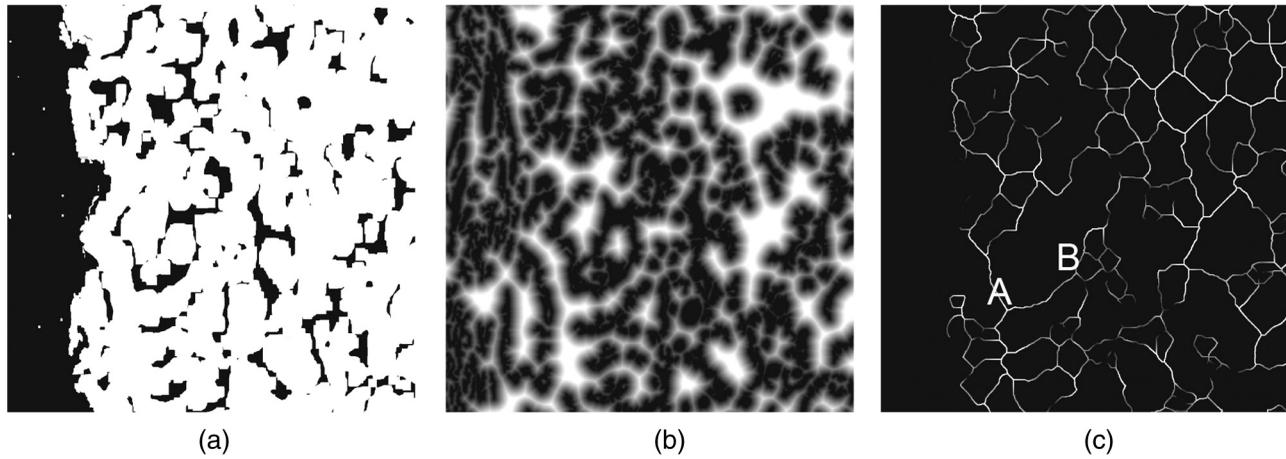
##### 4.2 Measurement of Trabecular Thickness

Increased trabecular thickness is one of the characteristic features of HCC. Trabecular thickness is measured in the direction orthogonal to the centerline of the trabeculae. The distance from an individual centerline to the boundary of the trabecula is derived by overlapping the distance-transformed image obtained from the thinning results. Doubling this center–boundary distance, we obtain the trabecular thickness at each point on the centerline.

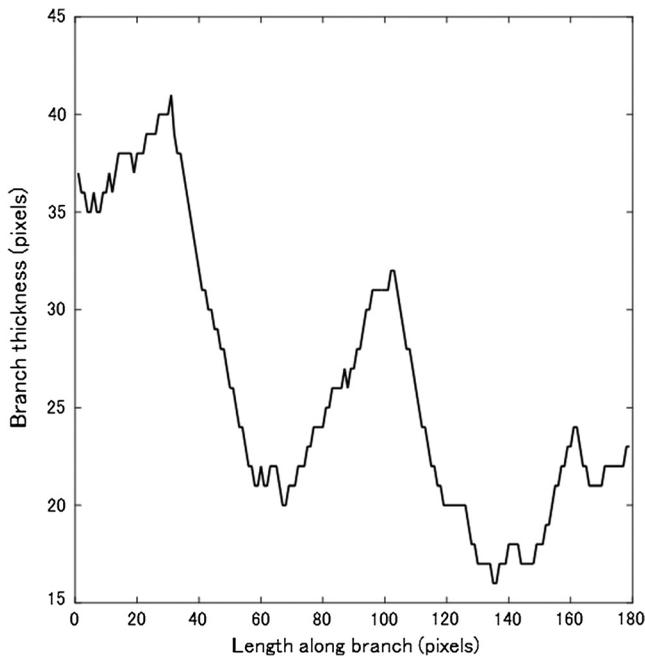
Figure 11(a) shows a trabecular image. Figure 11(b) shows the Euclidean-distance transformed image of Fig. 11(a), where the intensity increases with distance from the black pixels of the centerlines, shown in Fig. 11(c). The thickness along the centerline between branches A and B in Fig. 11(b) is depicted in Fig. 12. The thickness, computed by the above algorithm, increases in the vicinity of the branching points A and B in Fig. 11(b). Therefore, the region near point A has a large value because it is far from outside. This is because the trabecula is in contact with other hepatocytes, and hence this computed thickness does not reflect the real thickness. To reduce such effects, the pixels belonging to the centerline close to both



**Fig. 10** Results of trabecular centerline extraction: (a) centerlines of the trabecular regions and (b) states of the cord labels.



**Fig. 11** Thickness of individual elements along a branch: (a) original image of a trabecular region, (b) a branch along which the thickness is computed, and (c) thinning image of a distance image.

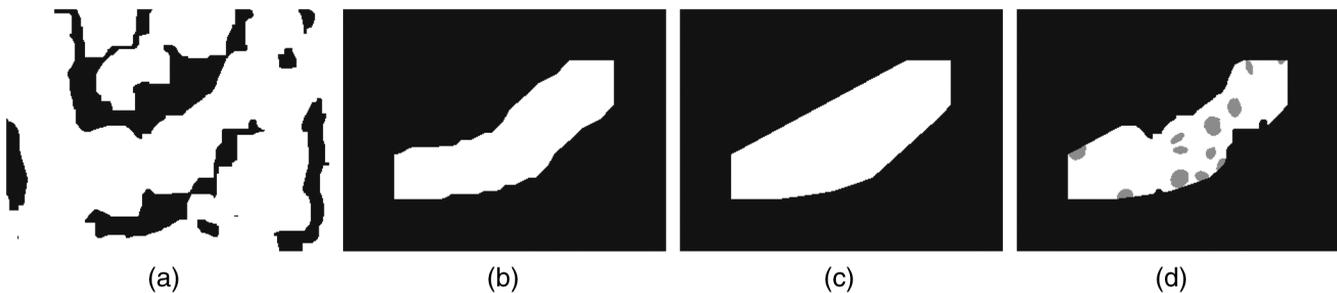


**Fig. 12** Thickness along the branch depicted in Fig. 11(b).

ends of A and B are excluded if the distance along the centerline to a branching point or an endpoint is smaller than the average thickness. For example, the ranges indicated by the gray segments in Fig. 12 are excluded from the calculation of the trabecular thickness representative value. The median thickness is then considered as the thickness of the cord. The trabecular thickness feature in a region of interest (ROI) is the average thickness of the trabeculae in the ROI.

### 4.3 Nuclear Density and Nuclear–Cytoplasmic Ratio

Nuclear density and N/C ratio are two important features in the diagnosis of cancer. While nuclear density defines the number of hepatocyte nuclei per unit area, the N/C ratio specifies the ratio of the areas occupied by hepatocyte nuclei and cytoplasm. In this study, the nuclear density and N/C ratio were computed for every cord label. Therefore, we extracted the cytoplasm around each cord label from the consecutive cytoplasm region. Figure 13 illustrates how to achieve this. Figure 13(a) illustrates the cytoplasm extracted from Fig. 13(b) using the technique explained in Sec. 3. The centerline is expanded proportionally to the thickness of the cytoplasm, as shown in Fig. 13(b), by morphological dilation operations. Moreover, because Fig. 13(b) differs from actual trabecular regions, its convex hull is calculated [see Fig. 13(c)], and the cytoplasm region of the trabecula is then extracted from Fig. 13(a), limited by the extraction of the convex hull region shown in Fig. 13(c). The final result is shown in Fig. 13(d). In this image, the gray spots are the extracted cell nuclei. The nuclear density and the N/C ratio are then, respectively, computed by



**Fig. 13** Cytoplasm segmentation: (a) cytoplasm, (b) dilation image of the trabecular centerline, (c) convex hull of the dilated image (b), and (d) result of trabecular extraction.

$$\text{Nuclear density} = \frac{\text{Number of nuclei}}{\text{Per unit area}}, \quad (1)$$

$$\text{N/C ratio} = \frac{\text{Area of nuclear}}{\text{Area of cytoplasm}}. \quad (2)$$

#### 4.4 Calculation of the Number of Layers

The trabecular thickness is often counted as the number of layers,<sup>22</sup> which is considered to be the number of cells aligned orthogonally to the centerline of the trabeculae. For example, the cytoplasm depicted in Fig. 13(d) has one or two layers.

To numerate the number of layers for individual cord labels, a simple model is used in which the cell arrangement is approximated by a rectangular box of thickness  $t$  and length  $l$  (see Fig. 14).

The area occupied by a single cell is estimated as follows:

$$a^2 = \frac{t \times l}{m}, \quad (3)$$

where  $a$  is the cell diameter, and  $m$  is the number of hepatocyte nuclei, respectively. The number of layers  $N_{lay}$  is, therefore, given by

$$N_{lay} = \frac{t}{a} = \sqrt{\frac{m \times t}{l}}. \quad (4)$$

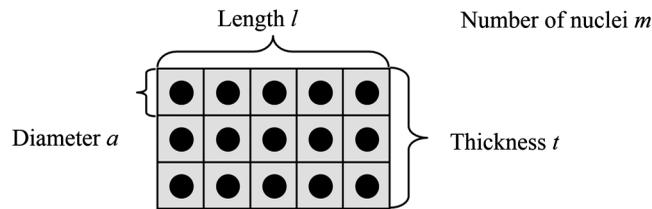


Fig. 14 Schematic of cell arrangement in local cytoplasm.

Figure 15 shows examples of determining the number of layers. The results are almost identical to those obtained by visual inspection.

#### 4.5 Feature Set for the Tissue in an Image

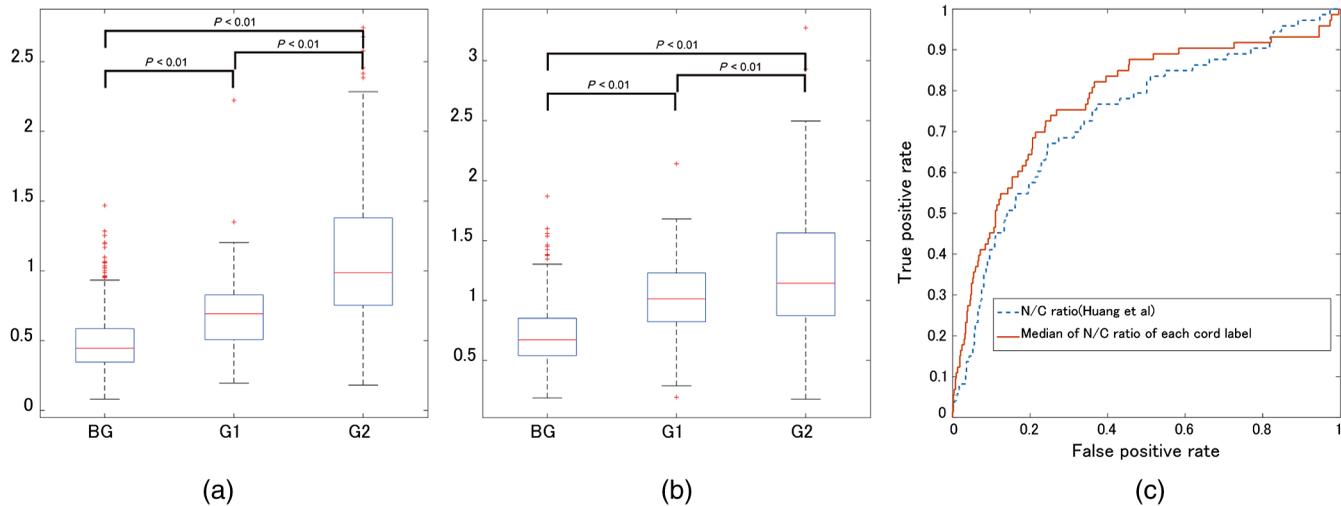
Using the methods described from Sec. 4.1 to Sec. 4.4, the features are calculated for each cord in a tissue. The feature set includes N/C ratio, nuclear density, and the number of layers. It is then possible to calculate the distributions of those tissue features in an image. As a feature descriptor for the image, we consider the median of each feature in the image. A feature vector for the tissue structure is derived for an image along with a feature vector for the nuclear morphology.<sup>23</sup> The features of N/C ratio and nuclear density have previously been used in conventional histopathology image analysis research;<sup>9,13</sup> however, their accuracy can be enhanced by aggregating them over all cords in the tissue image, because the sinusoid and stromal regions are excluded from the calculation.

## 5 Experiment

The experiment was performed on a collection of histopathology images supplied by the Department of Pathology, Keio University, Japan. This study was approved by the ethics committee of Keio University School of Medicine. To evaluate the feature values derived in the proposed method, we calculated them using a number of images of HE-stained liver specimens. The ROI images were  $2174 \times 2174$  pixels in size, which is equivalent to a  $1 \times 1$ -mm area. All slide images were captured using the NanoZoomer 2.0HT slide scanner (Hamamatsu Photonics K. K., Hamamatsu, Japan) at  $20\times$  (equivalent to  $0.46 \mu\text{m}/\text{pixel}$ ) as NanoZoomer Digital Pathology Images. In this experiment, 551 images were used, comprising 225 images from background (BG) tissues as noncancer samples, and 326 images of HCC. Based on the Edmondson grading, 80 and 246 of these images were graded as G1 and G2, respectively. The grades were annotated by a consensus of expert pathologists. The features described in Sec. 4 were calculated for each

Number of layers	Original image	Detected trabecula
1.46		
2.13		
3.45		

Fig. 15 Layer number calculation using the extracted trabecula images.



**Fig. 16** Boxplot of the (a) N/C ratio (Huang et al.), (b) the median of N/C ratio of each cord label, and (c) ROC curve.

cord in the image and the median was used as the representative value.

## 5.1 Relationship Between the Feature Values and Edmondson Grading

### 5.1.1 Nuclear–cytoplasmic ratio for each cord label

Figure 16(a) shows a boxplot of the “mean nucleus-to-cytoplasm ratio” of Huang et al.,<sup>9</sup> and Fig. 16(b) shows a boxplot of the N/C ratio distribution calculated using the method explained in Sec. 4. To calculate the result shown in Fig. 16(b), the N/C ratio of the per-unit area ( $2174 \times 2174$  pixels) was derived. The median N/C ratio for each cord label ( $p$ -value  $< 0.01$ ) and mean nucleus-to-cytoplasm ratio ( $p$ -value  $< 0.01$ ) were evaluated by the Kruskal–Wallis test. Figures 16(a) and 16(b) present the result of a multiple comparison test, which shows that the difference between the cancer and noncancer labels was significant. To verify the separation of BG and G1 by the proposed method, we computed the receiver operating characteristic (ROC) curve for classification by a logistic regression of the N/C ratio and proposed method [see Fig. 16(c)]. The area under the curve (AUC) was significantly higher for the median N/C ratio for each cord label than it was for the mean nucleus-to-cytoplasm ratio ( $p$ -value  $< 0.05$ ). The median of the N/C ratio calculated for all cords is more clearly separated than the mean nucleus-to-cytoplasm ratio calculated using the total areas of nuclear and cytoplasmic regions. This indicates

that the proposed method improves the separation of well-differentiated HCC from noncancer tissue. Table 1 shows the AUC of the N/C ratio and median of N/C ratio for each cord label.

### 5.1.2 Nuclear density for each cord label

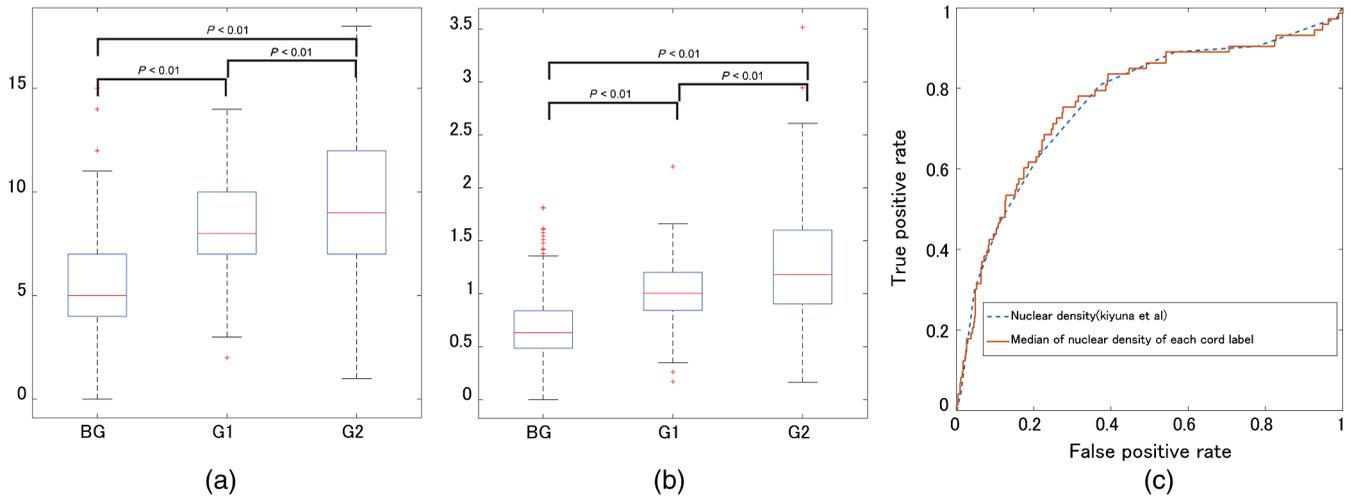
Figure 17(a) shows a boxplot of the median nuclear density method of Kiyuna et al.,<sup>13</sup> and Fig. 17(b) shows a boxplot of the nuclear density distribution calculated using the method explained in Sec. 4. The median nuclear density of each cord label ( $p$ -value  $< 0.01$ ) and nuclear density of Kiyuna et al. ( $p$ -value  $< 0.01$ ) were evaluated by the Kruskal–Wallis test. Figures 17(a) and 17(b) present the result of a multiple comparison test, which shows that the difference between the cancer and noncancer labels was significant. To verify the separation of BG and G1 by the proposed method, we computed the ROC curve for classification using a logistic regression of the nuclear density and proposed methods [see Fig. 17(c)]. An ROC curve was obtained for the median nuclear density and median nuclear density distribution for each cord label threshold. In this figure, the true positive rate is plotted on the y-axis versus the false positive rate on the x-axis. There was no significant difference between nuclear density of the proposed method and Kiyuna et al.’s method ( $P = 0.836$ ). Table 1 shows the AUCs of the nuclear density and median of nuclear density of each cord label. The method of Kiyuna et al. calculates nuclear density per unit area. The proposed method calculates nuclear density per unit trabeculae. Therefore, if the image includes substances other than cells, like stroma or regions of glass, the results of the proposed and conventional methods would be different. No significant difference was observed in this experiment because an image whose entire space was occupied by cells was used.

**Table 1** AUC and  $p$ -value for each feature.

	N/C ratio	Nuclear density	Trabeculae feature (BG versus G1)	Trabeculae feature (BG versus G1, G2)
AUC				
Conventional method	0.738	0.766	0.522	0.489
Proposed method	0.781	0.768	0.587	0.644
$p$ -value	$p < 0.05$	$p = 0.836$	$p = 0.141$	$p < 0.01$

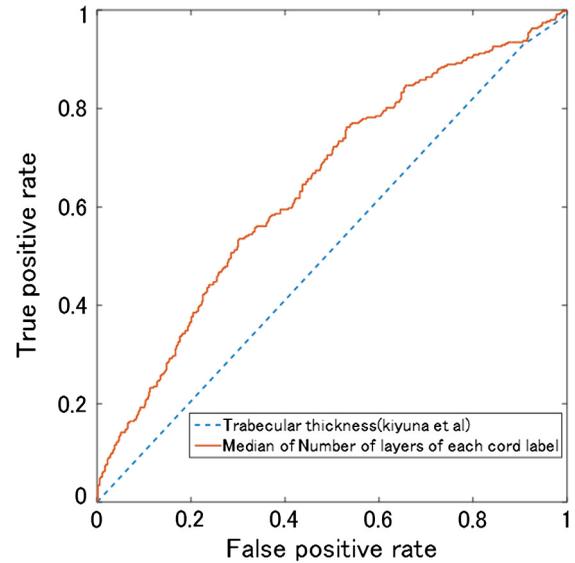
### 5.1.3 Number of layers for each cord label

Figure 18(a) shows a boxplot of the trabecular thickness of Kiyuna et al.,<sup>13</sup> and Fig. 18(b) shows a boxplot of the distribution of the number of layers, calculated using the method explained in Sec. 4. The median number of layers of each cord label ( $p$ -value  $< 0.01$ ) and median trabecular thickness feature of Kiyuna et al. ( $p$ -value = 0.478) were evaluated by the Kruskal–Wallis test. Figures 18(a) and 18(b) present the result of

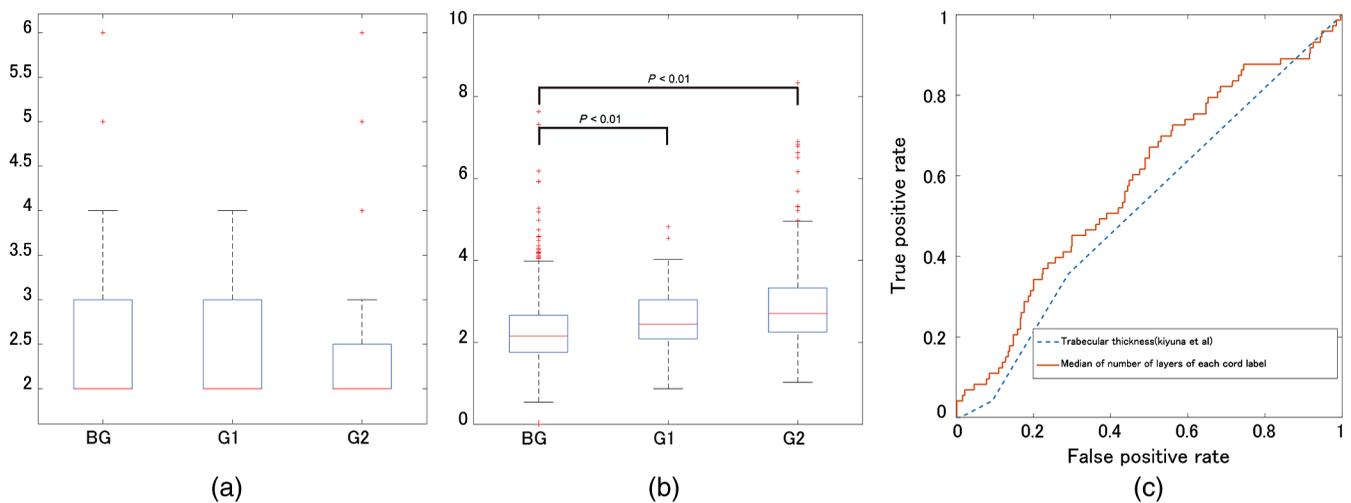


**Fig. 17** Boxplot of the (a) nuclear density, (b) median of nuclear density for each cord label, and (c) ROC curve.

a multiple comparison test, which shows that the difference between the cancer and noncancer labels was significant. To verify the separation of BG and G1 by the proposed method, we computed the ROC curve for classification by a logistic regression of the trabecular thickness and proposed method [see Fig. 18(c)]. An ROC curve was obtained for trabecular thickness and median number of layer distributions for each cord label threshold. In this figure, the true positive rate is plotted on the y-axis versus the false positive rate on the x-axis. There was no significant difference between the nuclear density of the proposed method and Kiyuna et al.'s method ( $P = 0.141$ ). Table 1 shows the AUCs for trabecular thickness and the median of the number of layers of each cord label. However, the number of layers of the Kruskal–Wallis test had a significant difference for BG and G1 as well as for BG and G2. Therefore, to verify the separation of BG and G1,G2 by the proposed method, we computed the ROC curve for classification by a logistic regression of the trabecular thickness and proposed method (see Fig. 19). The AUC was significantly higher for the median N/C ratio of each cord label than it was for the mean nucleus-to-cytoplasm ratio ( $p$ -value  $< 0.01$ ).



**Fig. 19** ROC curve of trabecula thickness and number of layers of BG and G1,G2.



**Fig. 18** Boxplot of the (a) trabecular thickness, (b) number of layers for each cord label, and (c) ROC curve.

The trabecular thickness method of Kiyuna et al. is not highly accurate. Although the trabecular thickness is a classification feature for HCC, it is difficult to diagnose HCC using trabecular thickness alone. Therefore, Kiyuna et al. proposed the use of a combination of various features such as nuclear density and NC ratio. The proposed method obtains a higher accuracy than the trabecular thickness method. It is believed that classification based on the number of layers alone is difficult because the structural changes in G1 compared to normal tissue are very small. As the proposed method is an approach that quantifies the number of layers, which is just one part of the diverse structural changes that occur in liver cells, it is not useful for all situations. Therefore, it is believed that the number of layers would be best used in combination with various other features such as nuclear density and NC ratio. In fact, in a study conducted by Aziz et al.,<sup>23</sup> the accuracy of the difficult classification of G1 was improved from 65% to 77% by combining nuclear features with the proposed method.

## 6 Conclusion

To analyze the structure of histopathological hepatic images, we isolated the trabeculae by extracting the sinusoids, fat droplets, and stromata. We then measured the morphological features of the extracted trabeculae, divided the image into cords, and calculated the feature values of the local cords. We proposed a method of calculating N/C ratio, nuclear density, and the number of layers in trabecula thickness using the local cords. The results of the experiment clearly show that the method was able to effectively quantify the Edmondson grade with respect to N/C ratio, nuclear density, and the number of trabecular layers. Quantifying the organization structure used in pathological diagnosis is a critical technology from the viewpoint of improving the precision of computer-aided diagnosis.

## Acknowledgments

This study was supported by the New Energy and Industrial Technology Development Organization of Japan.

## References

1. R. S. Weinstein et al., "Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future," *Hum. Pathol.* **40**(8), 1057–1069 (2009).
2. M. N. Gurcan et al., "Histopathological image analysis: a review," *IEEE Rev. Biomed. Eng.* **2**, 147–171 (2009).
3. H. Edmondson and P. Steiner, "Primary carcinoma of the liver. A study of 100 cases among 48,900 necropsies," *Cancer* **7**(3), 462–503 (1954).
4. R. Jagoe, C. Sowter, and G. Slavin, "Shape and texture analysis of liver cell nuclei in hepatomas by computer aided microscopy," *J. Clin. Pathol.* **37**(7), 755–762 (1984).
5. H. Irshad et al., "Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential," *IEEE Rev. Biomed. Eng.* **7**, 97–114 (2014).
6. S. Doyle et al., "Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer," *BMC Bioinf.* **13**, 282 (2012).
7. L. Gorelick et al., "Prostate histopathology: learning tissue component histograms for cancer detection and classification," *IEEE Trans. Med. Imaging* **32**(10), 1804–1818 (2013).
8. K. Nguyen et al., "Automatic glandular and tubule region segmentation in histological grading of breast cancer," *Proc. SPIE* **9420**, 94200G (2015).
9. P. Huang and Y. Lai, "Effective segmentation and classification for HCC biopsy images," *Pattern Recognit.* **43**(4), 1550–1563 (2010).
10. C. Atupelage et al., "Computational grading of hepatocellular carcinoma using multifractal feature description," *Comput. Med. Imaging Graph.* **37**(1), 61–71 (2013).
11. K. G. Ishak, P. P. Anthony, and L. Sobin, *Histological Typing of Tumours of the Liver*, Springer, New York (2012).
12. S. G. Hubscher et al., *MacSween's Pathology of the Liver*, Churchill Livingstone, London (2011).
13. T. Kiyuna et al., "Automatic classification of hepatocellular carcinoma images based on nuclear and structural features," *Proc. SPIE* **8676**, 86760Y (2013).
14. M. Ishikawa et al., "Automatic segmentation of hepatocellular structure from HE-stained liver tissue," *Proc. SPIE* **8676**, 867611 (2013).
15. M. Ishikawa et al., "Segmentation of sinusoids in hematoxylin and eosin stained liver specimens using an orientation-selective filter," *Open J. Med. Imaging* **3**(40), 144–155 (2013).
16. M. Ishikawa et al., "An accurate method of extracting fat droplets in liver images for quantitative evaluation," *Proc. SPIE* **9420**, 94200Y (2015).
17. A. Katoh et al., "Algorithm that uses self-organizing map and morphological features to extract hepatic fat droplets," in *Proc. IEEE Image Electronics and Visual Computing Workshop* (2012).
18. B. Turlin et al., "Assessment of hepatic steatosis: comparison of quantitative and semiquantitative methods in 108 liver biopsies," *Liver Int.* **29**(4), 530–535 (2009).
19. R. Lee et al., "Automated quantification of liver steatosis by WSI based image analysis," in *1st Congress of the Int. Academy of Digital Pathology, Analytical Cellular Pathology*, Vol. 34, p. 188 (2011).
20. R. Achanta et al., "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(11), 2274–2282 (2012).
21. L. Lam, S.-W. Lee, and C. Y. Suen, "Thinning methodologies—a comprehensive survey," *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(9), 869–885 (1992).
22. S. Manuel et al., "Histopathology of hepatocellular carcinoma," *World J. Gastroenterol.* **20**(43), 15955–15964 (2014).
23. M. A. Aziz et al., "Enhancing automatic classification of hepatocellular carcinoma images through image masking, tissue changes and trabecular features," *J. Pathol. Inform.* **6**, 026 (2014).

**Masahiro Ishikawa** is an assistant professor at the Faculty of Health and Medical Care, Saitama Medical University. He received his MS and PhD degrees in engineering from Niigata University, Niigata Japan, in 2004 and 2007, respectively. His research interests are medical image processing, pathology image recognition, and computer-aided diagnosis.

**Yuri Murakami** received her BS degree in mathematical and physical science from Japan Women's University at Tokyo, Japan in 1996 and her ME degree in information processing from the Tokyo Institute of Technology at Yokohama, Japan, in 1998. Her research interests are color image reproduction using multispectral imaging and multispectral image compression.

**Sercan Taha Ahi** received his MS degree in electrical and electronics engineering from the Middle East Technical University and his PhD degree in computational intelligence and systems science from the Tokyo Institute of Technology in 2007 and 2011, respectively. His research interest is identifying textural differences in histopathological images.

**Masahiro Yamaguchi** is a professor at the Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology. He received his BS, MS, and PhD degrees in engineering from the Tokyo Institute of Technology in 1987, 1989, and 1994, respectively. From 1996 to 2011, he was an associate professor at the Imaging Science and Engineering Laboratory in the same institute. From 1999 to 2006, he was concurrently a project subleader in Akasaka Natural Vision Research Center, TAO (currently NICT), Japan.

**Naoki Kobayashi** is a professor at the Faculty of Health and Medical Care, Saitama Medical University. He received his BS and ME degrees from the Tokyo Institute of Technology, Tokyo, Japan, in 1979 and 1981, respectively, and his PhD from Niigata University, Niigata, Japan, in 2000. His research interests are image processing, image compression, and bio-signal processing.

**Tomoharu Kiyuna** is a senior expert at the Medical Solutions Division of NEC Corporation. He received his MS degree from the University of the Ryukyus in 1990 and his PhD from Kyushu University in 2008. He joined NEC Corporation in 1990 where he worked on research of the inverse problem of electroencephalography (EEG). Since 2003, he has been engaged in the research and development of a pathology decision support system, e-Pathologist(R).

**Yoshiko Yamashita** received her BE and ME degrees in information science and engineering from Chuo University, Tokyo Japan, in 1998 and 2000, respectively. Since 2000, she has been working at NEC Corporation, Japan. Since 2006, she has been engaged in the research and development of a pathology decision support system. She is also a doctor course student of the Tokyo Institute of Technology. Her research interests are image processing and machine learning.

**Akira Saito** received his PhD in medical science from Hiroshima University in 1992. He worked on the development of pathological image analysis and cancer detection system from 2001 to April 2015 at NEC Corporation. He moved to Tokyo Medical University on 2015 July as research professor. Now he is researching on clinical evidences by quantitative measured data from pathological and immunological images.

**Tokiya Abe** is a project instructor in the Department of Pathology, Keio University School of Medicine. He received his PhD in 2005 from the Tokyo Institute of Technology, Yokohama Japan. He joined the Research Center for Frontier Medical Engineering, Chiba University. From 2008 to 2009, he worked as a postdoctoral researcher in the Department of pathology at Massachusetts General Hospital. His research interests are multispectral and histo-pathological image processing.

**Akinori Hashiguchi** is an assistant professor in the Department of Pathology, Keio University School of Medicine and is engaged in research on digital pathology and pathology informatics. He received his MD and PhD degrees from the Keio University School of Medicine, Tokyo, Japan, in 1995 and 2011.

**Michiie Sakamoto** is a professor in the Department of Pathology, Keio University School of Medicine since 2002. He received his PhD from Keio University School of Medicine in 1991. He is an Editor-in-chief of Pathology International, an associate editor of Cancer Science and Hepatology Research, and on the Editorial Board of many kinds of international journals. His research interests are Molecular Pathology of Cancer: multistep-carcinogenesis, early cancer, invasion and metastasis, and Pathology Informatics.