ALGORITHM THAT USES SELF-ORGANIZING MAP AND MORPHOLOGICAL FEATURES TO EXTRACT HEPATIC FAT DROPLETS

Ayako KATOH†   Emi SHIMADA†   Naoki KOBAYASHI†   Hideki KOMAGATA†
Fumikazu KIMURA‡   Masahiro YAMAGUCHI‡
Tokiya ABE†‡   Akinori HASHIGUCHI†‡   and   Michiie SAKAMOTO†‡

† Faculty of Health and Medical Care, Saitama Medical University
‡ Global Scientific Information and Computing Center, Tokyo Institute of Technology
† ‡ Department of Pathology, School of Medicine, Keio University

ABSTRACT

The accurate evaluation of the number and area of fat droplets in hepatic tissue images is needed for the risk evaluation of hepatocellular carcinomas. The purpose of this study is to develop an algorithm that can extract hepatic fat droplets accurately. Our extraction algorithm, which uses a Self-organizing map and morphological features, is challenged by pathology images of hepatic tissue that have already been scored. The difference between the area of the fat droplets extracted by the method and that judged by a cytologist (F.K.) was taken as the error area. The results show that the maximum ratio of error area to total tissue area was 2.9%. This algorithm provides a useful way of extracting fat droplets in hepatic tissue images with color robustness, and evaluating the changes in fat characteristics. This method is expected to support the study of the relationship between the spatial distribution of fat droplets of each size and cancer risk.

1. INTRODUCTION

Due to recent progress in digital slide scanner technology, demand has risen for the application of image processing technology in pathological diagnostic imaging. In order to achieve this, it is necessary to correctly recognize the cells, nuclei, extracellular matrix, blood vessels, etc. that compose pathological images [1][2].

Key features of hepatocellular carcinomas are nuclear density and internuclear distance, and several methods have been proposed to extract nuclei from an image and quantify nuclear density and internuclear distance. Moreover, the quantitative evaluation of fibrosis [3] or steatosis [4] is expected to develop risk indexes for hepatocellular carcinomas. Since steatosis is also an indicator of non-alcoholic steatohepatitis(or NASH), which is related to a lifestyle-related disease, exact extraction of fat droplets and quantization of steatosis are urgent goals.

We propose an extraction algorithm that uses a Self-organizing map (SOM) and morphological features; we validate it on pathology images of hepatic tissue.

2. MATERIAL AND METHOD

2.1. Fat droplet extraction

To develop an exact extraction algorithm we focus on two points. One is color robustness. The color of pathological images is important in classifying tissue components. However, the color details of Hematoxylin and eosin (H&E) stain tissue images vary with staining conditions etc. As SOM-based algorithms are expected to provide good robustness for pixel classification, we introduce SOM to the proposed extraction algorithm.

The second is morphological feature of fat droplets. As fat droplets on a pathological image are characterized by their whiteness, i.e., the color of the glass slide, discriminating them from sinusoids and blood vessels, both of which are also white, is important.

A flow chart of fat droplet extraction (image processing) is shown in Fig. 1. The extraction procedure is divided into three parts; extracting fat droplet candidates, setting boundaries to separate fat droplets that are connected, and the elimination of extraneous components such as sinusoids.

In order to extract fat droplet candidates by using color and morphological information, an algorithm that assigns each pixel into one of 20 classes was built using a SOM, which is one type of artificial neural network. The SOM input for pixel classification is 8-dimensional data consisting of YCbCr color components of the object pixel, the color components of the object pixel after applying a median filter, the pattern matching results with a medium size of fat droplet image and with a small size of lipid droplet image. The SOM is

trained by inputting the 8-dimensional data of a sufficient number of pixels. The classes indicative of fat droplets are determined by comparison against a manually classified image. Finally the classified image is converted into a binary image by using a threshold to extract fat droplet candidates.

However, the above algorithm sometimes "connects" adjacent fat droplets. These must be separated in order to acquire detailed data. Our solution is to use a Watershed line algorithm. To avoid excessive separation, a low-pass filter is used before the Watershed line algorithm.

To discriminate sinusoids, we use two circularity-based metrics. One is the standard deviation of distances from the center position to all points on the periphery of the object and the other is calculated from area and periphery length. The 1st metric yields rough sinusoid discrimination because after the watershed line algorithm is applied, sinusoids might be divided into several objects. The 2nd metric is used for final discernment of the fat droplets.
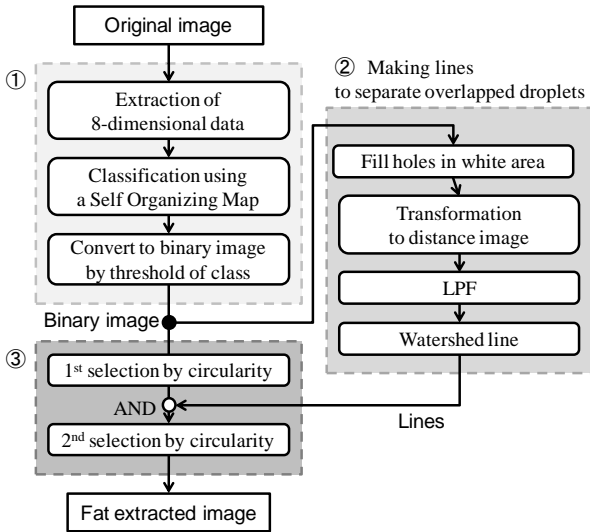

Fig.1 The flowchart of fat droplet extraction

2.2. Materials

Seven pathology image of hepatic tissue with histological scores were prepared. The scoring system we used is the steatosis component of the non-alcoholic fatty liver disease (NAFLD) activity score (NAS) designed by the Nonalcoholic Steatohepatitis (NASH) Clinical Research Network. The prepared images are evaluated on low- to medium- power evaluation of parenchymal involvement by steatosis Score 0 <5%; 1 5%-33%; 2 >33%-66%; 3 >66%. Table 1 shows the scoring system and prepared sample number.

In experiment1 a sample ROI was chosen from the one of the pathology images and a color changed image was prepared to assess color robustness. In experiment2, five ROIs that demonstrated the fat droplet pattern typical of each score were selected to verify fat extraction accuracy. ROI size was 256 pixels square. Experiment3 used all seven whole slide images.

Table 1   Steatosis score conponent of Histological Scoring System for Nonalcoholic Fatty Liver Disease (NAFLD) activity score

Score	Extent	Difinition	Number of sample
0	<5%	Refers to amount of surface area involved by steatosis as evaluated on low to medium power examination; minimal steatosis (<5%) receives a score of 0 to avoid giving excess weight to biopsies with very little fatty change	2
1	5-33%		3
2	>33-66%		1
3	>66%		1

2.3. Experimental settings

A SOM was trained using a pathology image of hepatic tissue different from the candidate for the experiment; the end result was an SOM that could classify the features of H&E stain tissue images into 20 classes. Referring to the cytologist's (F.K.'s) notation, the four classes of 18-19 were determined as fat droplet indicators.

The threshold values used in 1st and 2nd selection are shown in Table 2.

Table 2   Threshold value used at fat selection

	Standard deviation of distances from center to contour	Circularity calculated by area and contour
1st selection	$\leqq 0.5$	$\geqq 0.7$
2nd selection	$\leqq 0.2$	$\geqq 0.7$

3. RESULTS

3.1. Experiment1 – Color robustness -

We examined the color robustness of the proposed method in experiment1. Fig.2 shows a comparison of the fat candidate extraction results yielded by the proposed method and conversion with fixed threshold. Green areas are extraction results for fat droplet candidates before selection by circularity. The proposed method yields virtually the same output for both the original image and the color changed image. Conversely, the fixed threshold method is sensitive to color change.
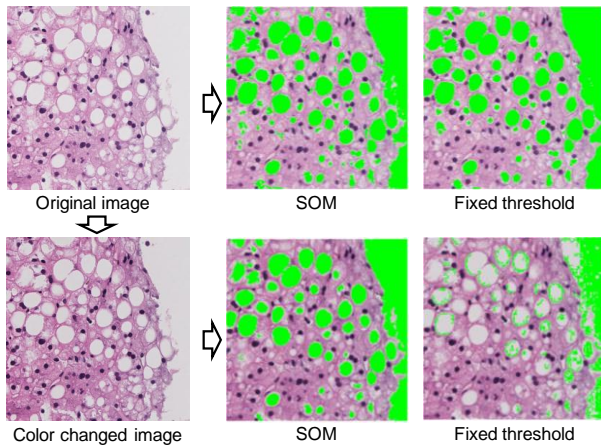
Fig.2 Color robustness assessment: The green areas are the fat candidate extraction results

3.2. Experiment2 - Verification of extraction accuracy -

An example of the fat droplets extraction results is shown in Fig. 3. White areas other than fat droplets were eliminated in the 1st selection (Fig.3 (4) A). The fat droplets formed by several droplets were separated by the separation procedure, and these could be extracted as fat droplets due to the circularity selection. (Fig.3 (3) B, (4) B). In the 2nd selection, objects that were not round such as clear cells and sinusoids were eliminated and fat droplets were extracted (Fig.3 (4) C). However, excessive separation yielded several erroneous results (Fig.3 (3) D, (4) D).

A comparison against the number of cells extracted by the cytologist is shown in Table 3. The average sensitivity of fat droplet extraction was 0.77. A small isolated fat droplet in a image with a lower steatosis component of NAS Score that appears similar to a small white area in a clear cell might produce a false-negative. Many false-positive results were clear cells and small sinusoids. Large sinusoids in a high steatosis component of NAS Score image that have similar index value of circularity to a large number of fat droplet agglomerations might produce false-positives.

Table 4 compares the areas of fat droplets extracted by the proposed algorithm to the areas based on the judgment of the cytologist. The difference between the areas was calculated and used as the error area. The results show that the maximum ratio of error area to the total tissue area was 2.9%.

(1) Original image   (2) Classified result

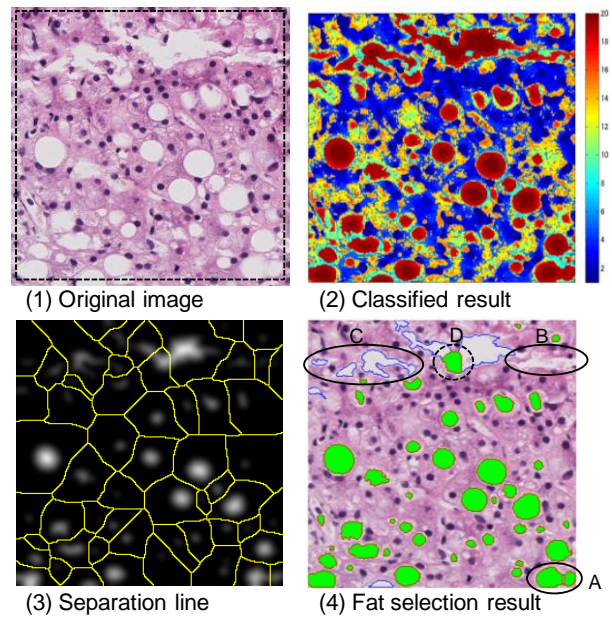(3) Separation line   (4) Fat selection result

Fig.3 Fat extraction result: The green areas framed by red line are the final extraction result. The droplets framed by a blue line were output by 1st selection but eliminated in 2nd selection.

Table 3   The number of fat droplets

Steatosis index of NAS Score	0	1	2	3
SUM	38	80	185	251
True-positive	32	41	162	234
False-positive	54	22	51	39
False-negative	6	39	23	17
Sensitivity	0.84	0.51	0.88	0.93
Positive predictive value	0.37	0.65	0.76	0.86

Table 4   The area of fat droplets

Steatosis index of NAS Score	0	1	2	3
Average area by a cytologist [pixels]	838	1607	6047	14674
Average area by the method [pixels]	1350	1775	7539	16542
Differences between the area by the method and by a cytologist [%]	0.79	0.88	2.29	2.87

3.3. Experiment3 - application to whole slide images –

Fig.4 and Fig.5 shows the results of the fat droplet extraction algorithm when applied to whole slide

images. The fat droplet areas were correlated with the steatosis component of NAS Score. Moreover, a detailed distribution of fat droplets over each whole slide was calculated automatically. Fig.5 (a) shows the original whole slide image. Green areas are extraction results (fat droplets) yielded by this method in Fig.5 (b). Fat droplet area for each segment (size of 240pixels wide and 252 pixels high) is shown in Fig.5 (c) where red represent large areas and blue represent small areas.
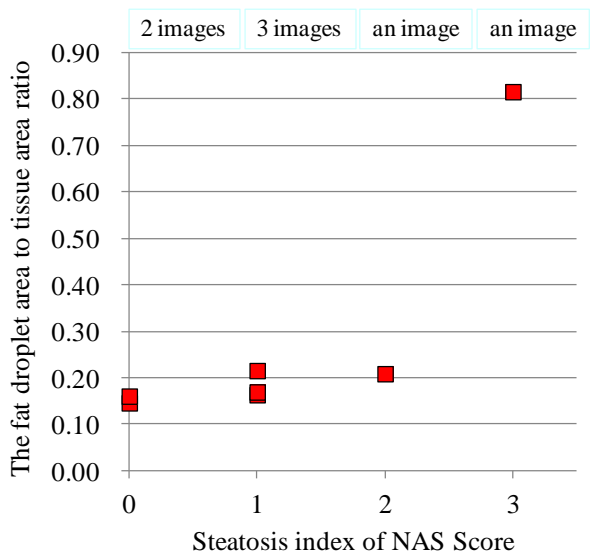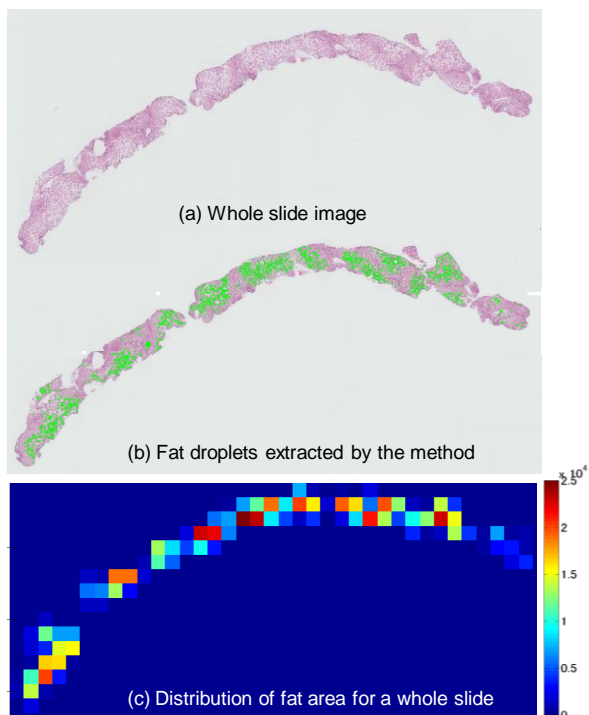

Fig.4 Fat droplet area

(a) Whole slide image

(b) Fat droplets extracted by the method

(c) Distribution of fat area for a whole slide

Fig.5 Fat droplet area over whole slide

4. DISCUSSION

We have developed a method that can automatically segment fat droplets from H&E stain tissue images. The algorithm is based on SOM and morphological features. The method could extract fat droplets and calculate the area of each fat droplet. Thus a histogram of fat droplet area can be calculated. The proposed method will be useful in performing fat droplet quantification and analysis.

A comparison of the proposal to manual judgment of the same image showed that extraction accuracy needs to be improved for practical use. In our method, selection is performed in two stages. The threshold value of the circularity indexes used in the 1st selection was set loosely in this study. Images with high steatosis component of NAS scores often contain as many as 10 fat droplets that have run into each other. Before separation by the watershed line algorithm, the circularity index values of these agglomerations are similar to those of large sinusoids. Using information of the surrounding area might solve this problem.

The proposed method is more robust than the fixed threshold method with regard to color information and morphological features. By processing the input data to identify boundaries and the morphological features of fat droplets and cytoplasm, the accuracy of extracting fat droplet candidates may be increased.

5. CONCLUSIONS

An effective method of evaluating the fat characteristics of patients with hepatocellular carcinoma and NASH will become increasingly important in assessing lifestyle-related disease. The method offers three key functions: (1) Extracting fat droplets individually, (2) Counting the number of fat droplets and calculating the area of each droplet and (3) Visualization of the distribution of fat droplets over the whole slide image. Experiments on actual pathology images of hepatic tissue confirm that the algorithm introduced in this study is useful in evaluating changes in fat characteristics with its image color robustness. This method is expected to support the analysis of the relationship between the spatial distribution of fat droplets of each size and cancer risk.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] R. S. Weinstein, A. R. Graham, L. C. Richter, G. P.Barker, E. A. Krupinski, A. Maria Lopez, K. A. Erps, A. K. Bhattacharyya, Y. Yagi, J. R. Gilbertson, "Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future," *Human Pathology*, vol. 40, pp. 1057–1069, 2009

[2] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, B. Yener, "Histopathological Image Analysis: A Review," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147-171, 2009

[3] T. Abe, K. Yamazaki, A. Hashiguchi, H. Saito, and M.Sakamoto, "Quantification of liver fibrosis by whole slide image analysis," Abstracts of the 1st Congress of the International Academy of Digital Pathology, *Analytical Cellular Pathology*, vol. 34, p. 189, 2011

[4] Roy Lee, Maristela Onozato, Nathan Shores, and Yukako Yagi "Automated quantification of liver steatosis by WSI based image analysis", *Abstracts of the 1st Congress of the International Academy of Digital Pathology, Analytical Cellular Pathology*, vol. 34, p. 188, 2011