Color processing in pathology image analysis system for liver biopsy

Yuri Murakami[1], Tokiya Abe[2], Yoshiko Yamashita[3], Masahiro Yamaguchi[1], Masahiro Ishikawa[4], Akinori Hashiguchi[2], Kiyuna Tomoharu[3], Akira Saito[3], Michiie Sakamoto[2]
[1]Global Scientific Information and Computing Center, Tokyo Institute of Technology; Tokyo, Japan
[2]Department of Pathology, School of Medicine, Keio University; Tokyo, Japan
[3]Medical Solutions Division, NEC Corporation; Tokyo, Japan
[4]Faculty of Health & Medical Care, Saitama Medical University, Hidaka-shi, Japan

Abstract

We have been developing a prototype system for the automatic detection of hepatocellular carcinoma (HCC) from whole slide images (WSIs) of liver biopsy based on image analysis techniques. In this paper, we present two color-related topics of this system: color correction and the calculation of color-related features of cytoplasm. A WSI-basis color correction method was implemented for the prototype system. We tested the color correction using more than 300 WISs, and it was confirmed that the color correction works well and stably. In addition, it was found that the success rate of nuclei detection significantly increaseed due to color correction. As for color-related features, we propose a method to calculate the representative color of cytoplasm and the clearness of cytoplasm. It was found that there was slight difference between the distributions of the representative colors of HCC and non-cancer tissues. In addition, there was slight correlation between the clearness of cytoplasm and nuclei density, which implies a promising role of the clearness index in a nuclei-based HCC detection.

Introduction

In pathological diagnosis, pathologists examine tissue slides using optical microscope to determine how cell and tissue morphology has changed, so as to yield the final diagnosis and the subsequent treatment policy. Recently, digital imaging technologies, as typified by whole slide imaging, have been introduced into the field of pathology [1]. One of the advantages of digital imaging in pathology is applicability of various techniques of digital image analysis [2].

We have been developing a prototype system for hepatocellular carcinoma (HCC) detection based on quantitative measurement by means of image analysis techniques (Figure 1) [3]. Targets of this system are whole slide images (WSIs) of liver biopsy specimens stained by Hematoxylin & Eosin (H&E) stain. Once a WSI of liver biopsy is registered, approximately 0.9 mm^2 regions called ROIs (region of interests) are automatically arranged to cover the whole tissue area. Then, for every ROI image, histological components such as nuclei [3], hepatic cell cords [4], and fat droplets are detected, and their morphological, textural, and distributional features are calculated, which are displayed as heat maps and histograms on the system. In addition, based on these measurement results, the system predicts the HCC probability of each ROI.

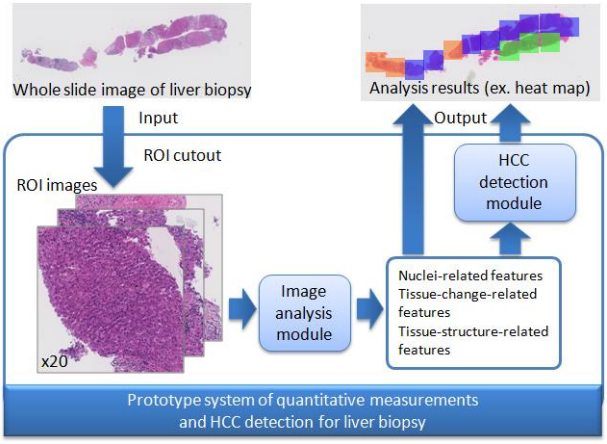

***Figure 1**. Schematic diagram of prototype system of quantitative measurements and HCC detection for liver biopsy.*

This paper introduces color-related topics of the developed prototype system. The first topic is color correction. It is well known that the color of stained tissue slides can be variable depending on protocol of staining [5]. Since image analysis is affected by the color variation, color correction is inevitable for stable image analysis. Therefore, the prototype system we developed is equipped with a function of color correction. In this paper, we introduce the method and results of the color correction.

In addition, the color of a stained tissue sometimes contains the information on the state of the tissue. For instance, the degree of clearness of cytoplasm indicates large stores of glycogen; the color balance of H&E in cytoplasm indicates the acidophile degree. Although it is difficult to separate these meaningful color changes from meaningless color variation, the prototype system try to calculates and output these color-related characteristics. The second topic of this paper is these color-related features calculated by the prototype system.

Materials

For developing and testing the prototype system, mainly three sets of WSIs of H&E-stained liver specimens were used; they are called Sets A-C. Set A consists of specimens from the portions of liver removed by surgery. Sets B and C consists of liver biopsy specimens. Every liver specimen was collected from HCC patients, and prepared and stained at a single medical facility, but at

different time by different operator. A WSI of each specimen was acquired using the same slide scanner, the NanoZoomer2.0HT® (Hamamatsu Photonics K.K., Hamamatsu, Japan) at ×20 magnification. The difference of colors of WSIs is larger among sets compared to within a set.

Color Correction of WSIs

Color correction method

Histological components in H&E-stained tissue specimens have specific colors respectively: nuclei are blue-purple of H, cytoplasm is pink-purple of the combination of H&E, fibers are pink of E, and so on. Therefore, color is one of the important information for detecting each component. However, depending on the protocol of staining, the color distribution shifts overall. Therefore, the purpose of the color correction is to match the color distribution of a target WSI to the color distribution which the detector supposes.

The color correction was performed per each WSI but for each ROI. Since a single ROI corresponds to approximately a 0.9 mm^2 area, there is a limit to the variety of the components included in a ROI. While on the other hand, a WSI likely to include various components as well as various types of tissue. Therefore, if we use the colors sampled from a WSI, it is easy to expect the accurate color distribution of the WSI which reflects the staining condition.

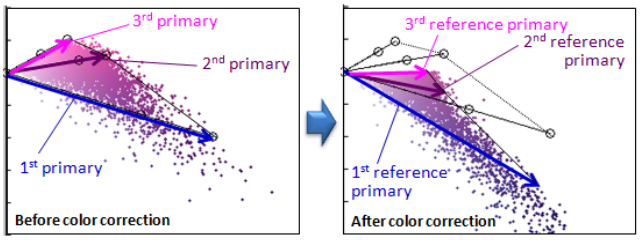

Figure 2. *Color distribution of example WSI of H&E-stained liver biopsy samples before and after color correction. Colors are plotted on a plane in logarithm RGB space and the three primary vectors of target WSI and reference are shown by arrows, which are used for color correction.*

There have been several reports on color correction methods for H&E stained tissue slide images [6-8]. A promising approach is based on a model that the colors are approximately represented

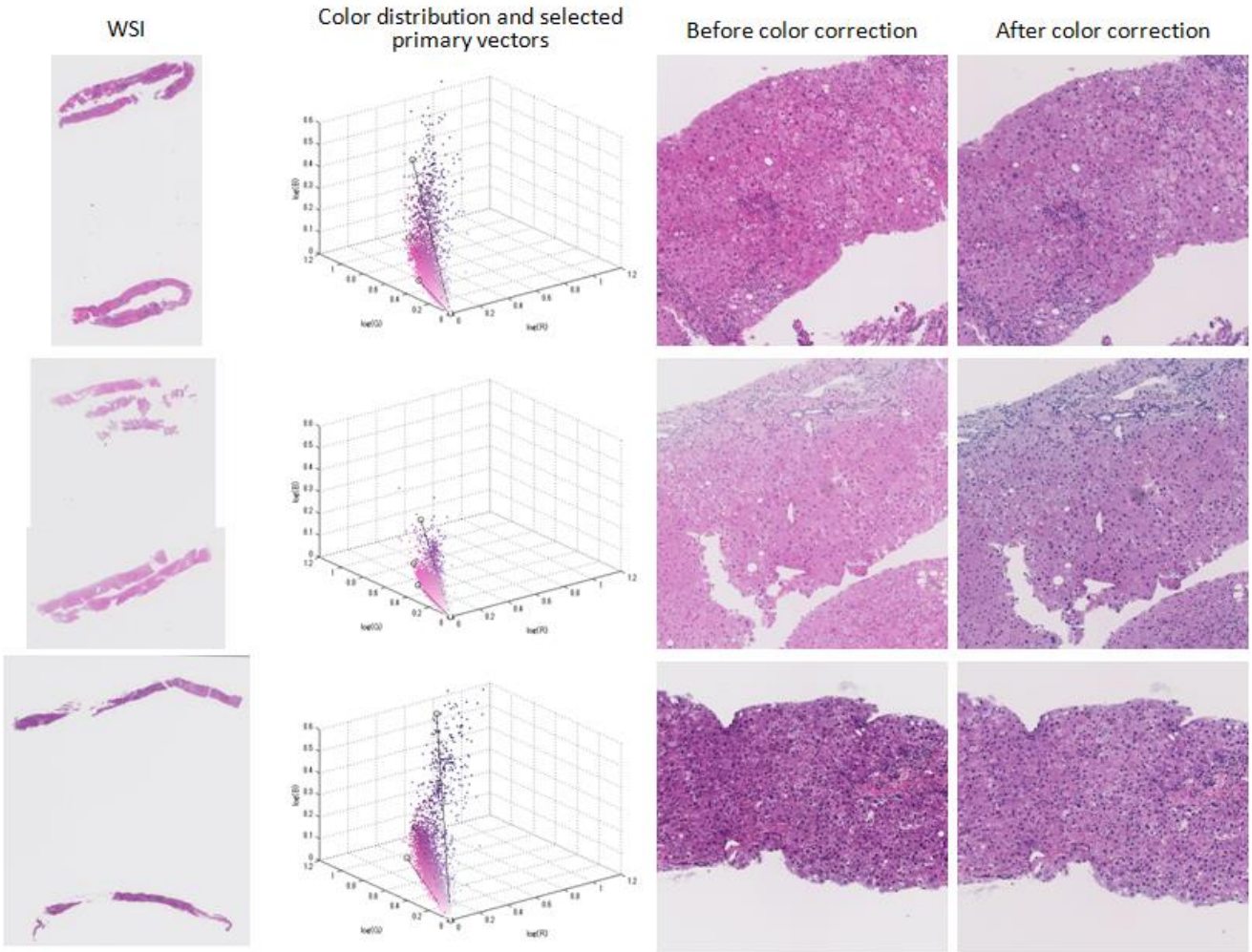

Figure 3. *Three example WSIs with their color distribution and ROI images before and after color correction. Top and middle are from set B, and bottom is from set C. Color distribution is plotted in logarithm RGB space and the extracted three primary vectors are shown as black lines.*

by the weighted sum of two vectors representing H and E dyes in logarithm RGB color space [6]. In the prototype system, a similar model was adopted; however, three vectors, called primary vectors, are used because it is difficult to identify the vector which directly corresponds to H or E dye. More specific procedure of the color correction is explained below:

- First of all, RGB signal values of glass area are obtained as 'white' of the target WSI. Then, white balance adjustment is applied to the target WSI.
- Using a low-magnification WSI, several ROIs that are suitable for color sampling are selected. Then, colors are sampled from high-magnification ROI images corresponding to the selected ROIs. This two-step procedure realizes effective and correct color samplings [9].
- Based on the sampled colors, a color-distributing plane is detected in logarithm RGB color space. Then, three primary vectors on the plane are identified based on the distribution peak and wide. The set of primary vectors is obtained per WSI.
- Next is the color correction phase. The color of each pixel in logarithm RGB color space is represented by the weighted sum of the two primary vectors out of three, and re-represented by exchanging the primary vectors for the reference primary vectors.
- Finally, the white is adjusted to the white of reference WSI.

Figure 2 shows the color distribution of an example WSI. The color distribution is presented on a plane in a logarithm RGB color space. The left panel shows the color distribution before color correction and the three primary vectors extracted by the system. The right panel shows the color distribution after color correction and the three primary vectors of reference. We can see the shift of the color distribution by the color correction.

Color correction results

Figure 3 shows the example results of color correction. The left-side two columns show WSIs and its color distribution accompanied with extracted three primary vectors in a logarithm RGB color space. The right two columns show example ROI images picked up from the WSIs before and after color correction. The top two WSIs are from Set B and the bottom WSI is from Set C. We can see that the color of WSIs are different each other. In addition, it is confirmed that the ROI images after correction have similar tendency of color.

Let us introduce one of the results which explains the effectiveness of the color correction. We performed nuclei detection for the image of Sets B and C, where the nuclei detection module of the prototype system had been developed by using Set A images. Supposing that nuclei detection is regarded as a success when approximately more than 80% nuclei are detected, we counted the number of the success ROIs. As a result, only 20% ROIs of Set B succeeded while 98% ROIs of Set C succeeded without color correction. This result means that the color of Set B images is largely different from Set A images. However, if we apply nuclei detection after color correction, the success rate increases to 83% for Set B, where the reference of the color correction was made from the images of Set A. An example result of nuclei detection before and after color correction are shown in Figure 4. Although still 17% ROI images of Set B was not succeeded, the staining condition of these images is severe bad caused by color fading. We are considering to adding the function to detect and alert for the WSIs of such inacceptable staining conditions.

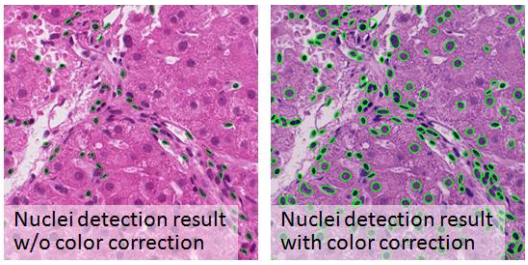

***Figure 4**. Example results of nuclei detection. Detected nuclei are presented by green lines. Without color correction, nuclei detection is almost failed(left) while every nucleus can be detected by applying color correction (right).*

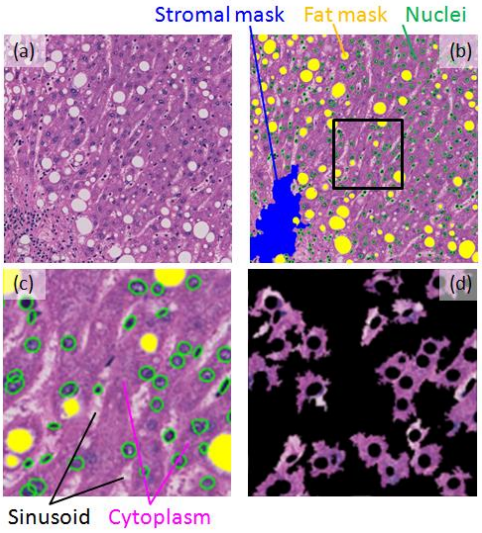

***Figure 5**. (a) Original image, (b) masked image, (c) magnified masked image, and (d) extracted cytoplasm region. Black rectangle in (b) shows the area of (c) and (d).*

Quantification of color features of cytoplasm

The color of cytoplasm sometimes reflects the condition of tissue; for instance, strong E indicates acidophile state. However, the color of cytoplasm changes depending on not only the tissue condition but also the staining condition. In addition, it is difficult to separate these two meaningful and meaningless color changes.

At present, prototype system performs color correction without distinguishing these color changes. Therefore, the color-corrected images might lose the meaningful color information. However, since color correction is performed on a WSI basis, the color difference contained in a single WSI can be saved even after color correction. Moreover, in the future, the meaningful color changes can be maintained if color correction is performed on a group-of-WSIs basis, where the group of WSIs consists of the WSIs which are prepared and stained under the same condition.

Below, we'll explain the extraction of cytoplasm regions and two kinds of color-related features obtained from them. The calculations were performed from color-corrected images. Therefore, as mentioned above, some of the original color information might be lost through the color correction procedure; however, we'll show that there is still meaningful information.

Extraction of cytoplasm

A tissue section of liver mainly consists of hepatocyte and stroma. Hepatocyte consists of cytoplasm and nuclei, and sometimes contain fat droplets (white circles in images). The gap appears between hypatocyte called a sinusoid, which corresponds to the capillary of liver. We have developed method to detect stroma [4], fat droplets, and nuclei [3] before. Therefore, by excluding these regions, we can obtain the area which only consists of cytoplasm and sinusoids. As described below, there is a large color variation in cytoplasm especially for its lightness. Therefore it is difficult to use color information to stably extract cytoplasm. On the other hand, it is sufficient to partially extract cytoplasm regions for analyzing color features of cytoplasm. In addition, cytoplasm is surrounding nuclei in most cases. Considering these facts, we took the approach to extract the surrounding regions of nuclei as cytoplasm region from masked image with stroma and fat masks. More specifically, a masked image is divided into superpixels and the only superpixels adjacent to nuclei are selected, which makes sinusoid less likely to be extracted by mistake. Figure 5 shows an example image, its masked image, and extracted cytoplasm regions.

Color of cytoplasm

We identify the color of cytoplasm of each ROI image as follows. The average of the color within each extracted superpixel is calculated, and then, three-dimensional histogram in RGB color space is calculated. The color with highest frequency is decided as the representative color of cytoplasm of the ROI.

Figure 6 shows a bubble plot of the representative colors of cytoplasm calculated from non-cancer ROIs (blue) and HCC ROIs (red) from Set B. The top-right panel is the plot for showing the colors corresponding to the bubble centers. The size of the bubbles indicates the number of the ROI images which have the color as their representative colors of cytoplasm. We can see the slight difference in the distributions of HCC and non-cancer; the distribution of HCC is slightly shifts to blue direction, which means less containing E dye.

In the case of Set B, liver biopsy specimens obtained from HCC and non-cancer portions of liver are placed on the same glass slide, and scanned as a single WSI. Therefore, if there is any color difference between HCC and non-cancer tissues, the difference of them can be seen as in Figure 6 even after color correction.

Clearness degree of cytoplasm

When cytoplasm contains glycogen, the color of cytoplasm becomes faint. The top row of Figure 7 shows three tissue images composed of cell of non-clear cytoplasm, moderately clear cytoplasm, and clear cytoplasm from left to right. We tried to quantify the degree of the clearness of cytoplasm based on the extracted cytoplasm regions.

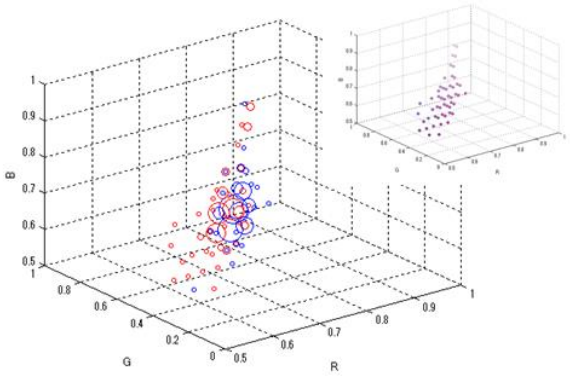

***Figure 6**. Bubble plots of representative color of cytoplasm from non-cancer ROIs (blue) and HCC ROIs (red). Top-right panel shows colors corresponding to the bubble center.*

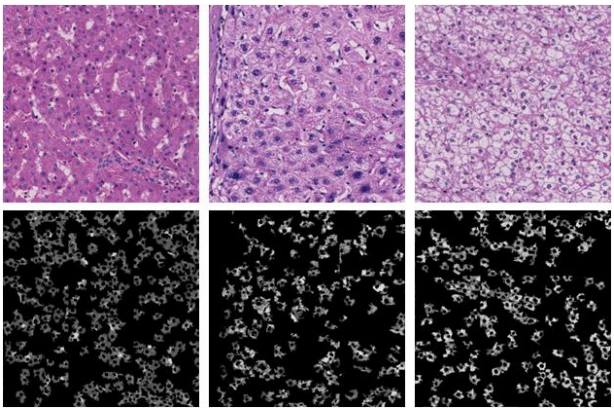

***Figure 7**. Top row shows tissue images composed of cells of non-clear cytoplasm (left), moderately clear cytoplasm (middle), and clear cytoplasm (right).Bottom row shows corresponding green channel images in which only extracted cytoplasm regions are presented.*

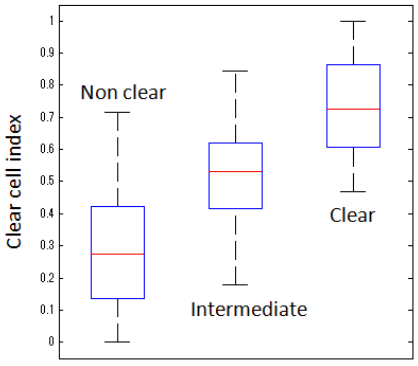

***Figure 8**. Boxplot of clear cell indices calculated for ROI images grouped into three categories: non clear, intermediate, and clear.*

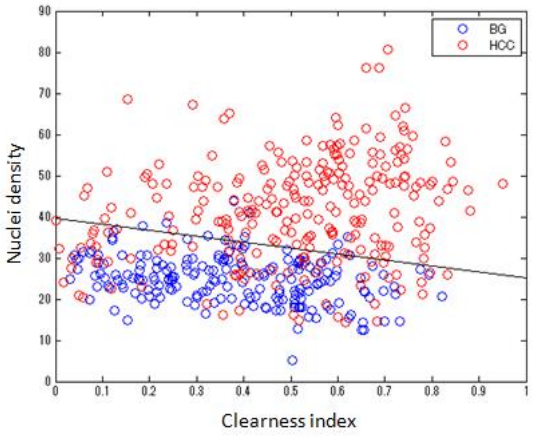

***Figure 9**. Scatter plot of clearness index and nuclei density for HCC and non-cancer (BG).Black line is linear discriminate line between HCC and BG on this plane.*

The bottom row of Figure 7 shows the green channel images in which only extracted cytoplasm regions are presented. We can see that the variance of the signal value increases as clearness increases. Then, we focus on the cumulative histogram of the signal values of these pixels, and derive the clearness index based on it.

Non-cancer ROI images from Set A are divided into three categories by visual observation: non clear, intermediate, and clear. The number of ROI images in each categories is 327, 135, and 70. Then, clearness index was calculated for every ROI image. Figure 8 shows boxplot of the results. We can see that the derived index well represents the clearness of the cytoplasm.

Next, we examined the effectiveness of the clearness index on HCC detection. More specifically we examined the relationship between the clearness index and nuclei features which are main features in HCC detection. Figure 9 shows the relationship between clearness index and nuclei density calculated using the ROI images in Set B. Although we can see the obvious difference in nuclei density between HCC and non-cancer, it is also confirmed that discrimination of HCC from non-cancer becomes easy by adding the clearness index to nuclear density.

Conclusions

In this paper, we present the color-related topics of prototype system for image analysis of liver biopsy. First, a color correction implemented in the system was presented. Although the color correction does not adopt novel techniques, we confirmed that it works well and stably for most WSIs expect for some WSIs suffered from severe color fading. In the present system primary vectors used for color correction are extracted on a WSI basis. However, if we perform it on a group-of-WSIs basis, it could be possible to correct meaningless color changes while meaningful color changes are maintained. Second, we presents the color-related features of cytoplasm calculated by the system. In pathology diagnosis, morphological features of tissue, such as the increase of nuclei density, are important. As well as these morphological features, we present the promising features which represent the state of tissues based on color information. We are supposed to examine the effectiveness of these features in HCC detection in detail in the future.

ACKNOWLEDGEMENT

This research is supported by New Energy and Industrial Technology Development Organization (NEDO) under the Research and Development Project for pathological image recognition.

References

[1] R. S. Weinstein, A. R. Graham, L. C. Richter, G. P. Barker, E. A. Krupinski, A. M. Lopez, Y. Yagi, and J. R. Gilbertson, "Overview of Telepathology, Virtual Microscopy, and Whole Slide Imaging: Prospects for the Future," Human Pathology, 40, 1057 (2009).

[2] M. N. Gurcan, L. E. Boucheron, A Can, A. Madabhushi, N. M. Rajpoot, B. Yener, "Histopathological Image Analysis: A Reivew," IEEE Review in Biomedical Engineering, 2, 147 (2009).

[3] T. Kiyuna, A. Saito, A. Marugame, Y. Yamashita, M. Ogura, E. Cosatto, T. Abe, A. Hashiguchi, M. Sakamoto, "Automatic classification of hepatocellular carcinoma images based on nuclear and structural features", SPIE Medical Imaging 2013, 2013

[4] M. Ishikawa, S. T. Ahi, F. Kimura, M. Yamaguchi, H. Nagahashi, A. Hashiguchi, M. Sakamoto, "Segmentation of Sinusoids in Hematoxylin and Eosin Stained Liver Specimens Using an Orientation-Selective Filter," Open Journal of Medical Imaging, 3(4), 144-155 (2013).

[5] Y. Yagi, "Color Standardization and Optimization in Whole Slide Imaging," Diagn. Pathol., 6 Suppl 1, S15 (2011).

[6] Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, et al. A method for normalizing histology slides for quantitative analysis. Proceedings of the 6th IEEE International Conference on Symposium on Biomedical Imaging: From Nano to Macro; 2009 Jun 28-1; Boston, Massachusetts. NJ: IEEE Press Piscataway; 2009.

[7] D. Magee, D. Treanor, D. Crellin, M. Shires, K. Smith, K. Mohee, et al., "Colour normalisation in digital histopathology images," Proceedings of Optical Tissue Image Analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop); 2009 Sep 20-24. London, UK.

[8] S. Tani, Y. Fukunaga, S. Shimizu, M. Fukunishi, K. Ishii, K. Tamiya, "Color standardization method and system for whole slide imaging based on spectral sensing," Anal Cell Pathol (Amst) 2012;35:107-15.

[9] Y. Murakami, T. Abe, A. Hashiguchi, M. Yamaguchi, M. Sakamoto, "Color correction for automatic fibrosis quantification in liver biopsy specimens," J Pathol Inform 2013;4:36 (2013).

Author Biography

Yuri Murakami received her M. S. and Ph. D. degrees from Tokyo Institute of Technology (Tokyo Tech), Japan in 1998 and 2005. She was an assistant professor at Tokyo Tech in 2000-2005 and in 2012-2014. Her research interests include multispectral imaging, color image reproduction.