https://doi.org/10.2352/issn.2169-2629.2021.29.179

Dye Amount Estimation in a Papanicolaou-stained specimen using Multispectral Imaging

Saori Takeyama[†], Tomoaki Watanabe[†], Masahiro Yamaguchi[†], Takumi Urata[††], Fumikazu Kimura[††], and Keiko Ishii[†††]; [†]Department of Information and Communications Engineering, School of Engineering, Tokyo Institute of Technology, Kanagawa, Japan; [††]Department of Biomedical Laboratory Sciences, School of Health Sciences, Shinshu University, Nagano, Japan; [†††]Division of Diagnostic Pathology, Okaya City Hospital, Nagano, Japan

Abstract

Papanicolaou stain is mainly used in cytological diagnosis such as gynecological diseases. In the image analysis of stained histopathology specimens, color unmixing technique, which estimates the dye abundance map, is useful. In this paper, we apply the dye amount estimation method based on color unmixing to Papanicolaou-stained specimen. In the proposed method, we capture the Papanicolaou-stained samples using a multispectral microscope, and then we estimate the amount of dyes from the observation and practically measured the spectral characteristics of the stain. Besides, we construct an application depicting the amount of stain and the bar-graph plot. In the experiments, we verify the feasibility of the proposed method and analyze a precancerous lesion of the uterine cervix using the proposed method.

Introduction

Pathological diagnosis morphologically identifies tissue structures of disease and reveals the cause and the construction. Therefore, it is an essential technique to current medical. Pathological diagnosis has roughly two methods: histology and cytological diagnosis. Histology uses tissue collected from a lesion. In histology, many image analysis methods have been implemented thanks to the great progress of Whole Slide Imaging (WSI) technology, which captures the whole sample by high resolution [1].

In cytological diagnosis, a sample is collected by scraping cells, so it can more easily and less burden be gotten than that of histology. The sample includes thick cells of about $20\mu m$, and analyzing the 3D structure of cells can expose the characteristics of diseases. To suitably scan the 3D structure, the technique is necessary to capture the sample in high resolution for many focuses, and currently, it is difficult at high speed. For the reason, the image analysis methods for cytological diagnosis are not actively addressed, but the techniques are essential. This is because, with the improvement of the scanning technique, the 3D structure would enable it to be captured in near future. In addition, in some applications that do not require 3D information, it is expected to apply image analysis techniques for acquiring quantitative information useful for cytological diagnosis.

Lobular endocervical glandular hyperplasia (LEGH) cells are believed to be a precancerous lesion of minimal deviation adenocarcinoma and adenocarcinoma of gastric types in the uterine cervix and expected to be useful for early detection of them [2]. For this reason, the identification of LEGH cells from normal endocervical (EC) cells is required, but it is a very hard task to morphologically distinguish them by human eyes because LEGH cells lack nuclear atypia [3]. Ishii et al. indicate that LEGH has the neutral mucin in the cytoplasm, and the color is subjectively different from that of EC in [4,5]. In general, EC has acid mucin, so the analysis of the mucin plays an important role in the identification. In [6], it is reported that the Papanicolaou technique stains the mucin of EC and LEGH cells to pink and yellow color, respectively, which is useful to distinguish LEGH cells. However, there are no quantitative diagnostic criteria based on the color of the mucin in the cervical cell specimen. Here, Papanicolaou stain dyes cell nucleus and lipid to hematoxylin (H) and Bismarck brown Y (BY), respectively. Then, it stains cytoplasm to Eosin Y (EY), Light Green SF yellowish (LG), or Orange G (OG), whose strengths are changed by molecular weight, diffusibility, and differentiation. The quantification of color in the Papanicolaou-stained cytological image is not easy because of the device dependence of the RGB values captured by a color camera. In the paper, the authors in [6] suggest the quantification of the color of mucin based on the amount of each dye in the Papanicolaou-stained specimen.

Dye amount estimation has been employed in histopathology for a long time. For example, in H&E (Hematoxylin and Eosin) staining, the image of the H component is used to automatically detect cell nuclei. In immunostaining, DAB (3,3'-Diaminobenzidine) component image is often used to extract the protein expression. For dye amount estimation, linear color unmixing or color deconvolution technique is widely used [7–9]. Nevertheless, in Papanicolaou staining, the five dyes are used, and it is impossible to apply linear color unmixing from RGB three-channel images.

Based on the above discussion, we propose a dye amount estimation method from a multispectral (MS) image based on color unmixing. Here, the MS image has information on cells dyed by Papanicolaou stain. For the proposed method, at first, we practically capture cells dyed by each stain, i.e., H, E, LG, or OG using an MS camera and calculate the spectral absorption coefficient matrix of Papanicolaou stain. Then, by using the pseudo-inverse matrix, the proposed method estimates the amount of stain at every pixel of the microscopic image of the cytological specimen.

Besides, to use the proposed method in actual cytological diagnosis, we construct an application software that estimates each dye amount and plots the average dye amounts in a selected area. In the experiments, we illustrate the performance feasibility of the proposed method and investigate the possibility to identify the characteristics of EC and LEGH cells based on the dye amounts of the mucin region in the image.

Proposed Method

In this work, we aim to estimate the amount of stain from an observed MS image in the Papanicolau stain. Color unmixing (spectral unmixing) is a technique that estimates underlying colors from observation, and it is a key technique for our proposed method. By assuming Lambert-Beer ' s law, for a certain wavelength λ, the absorbance of an observation $\alpha(\lambda)$ is modeled by

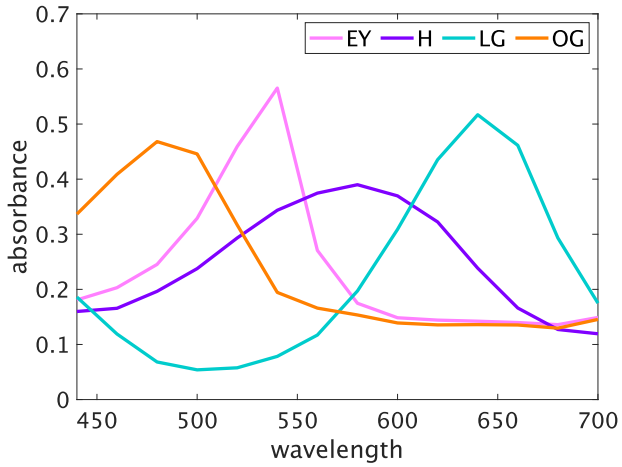


Figure 1. *The absorbance of EY, H, LG, and OG*

Figure 2. *Overview of the proposed application software tool.*

the following formulation:

$$\alpha(\lambda) = \sum_i c_i \varepsilon_i(\lambda), \tag{1}$$

$$I_1(\lambda) = 10^{-\alpha(\lambda)} I_0(\lambda), \tag{2}$$

where c_i is the amount of an ith stain ($i = 1,\ldots,M$, and M is the number of dyes), and $\varepsilon_i(\lambda)$ is the spectral absorption coefficient. Let $I_0(\lambda)$ and $I_1(\lambda)$ be the transmitted and incident lights, respectively. They can be measured from the un-stained (glass) and stained areas in observation, respectively. By letting λ_i ($i = 1,\ldots,N$) be wavelength of an N-bands MS image, we reformulate (1) as follows:

$$\mathbf{a} = \mathbf{H}\mathbf{c} \tag{3}$$

$$\text{s.t. } \mathbf{a} = (\alpha(\lambda_1), \alpha(\lambda_2), \ldots, \alpha(\lambda_N))^\top \in \mathbb{R}^N,$$

$$\mathbf{H} = \begin{bmatrix} \varepsilon_1(\lambda_1) & \varepsilon_2(\lambda_1) & \cdots & \varepsilon_M(\lambda_1) \\ \varepsilon_1(\lambda_2) & \varepsilon_2(\lambda_2) & \cdots & \varepsilon_M(\lambda_2) \\ \vdots & & \ddots & \vdots \\ \varepsilon_1(\lambda_N) & \varepsilon_2(\lambda_N) & \cdots & \varepsilon_M(\lambda_N) \end{bmatrix} \in \mathbb{R}^{N\times M}$$

$$\mathbf{c} = (c_1, c_2, \ldots, c_M)^\top \in \mathbb{R}^M.$$

From the above model, we can estimate the amount of stain $\mathbf{c}$ by the pseudo-inverse of the matrix $\mathbf{H}$, which represents the spectral absorption coefficients of dyes.

First, for the spectral calibration of the MS microscope (PerkinElmer Inc., currently sold by Akoya Biosciences Inc. Vectra®3), we capture a color chart slide, and we fit the results to ground-truth spectral transmittance according to [10]. Here, Vectra®3 is a 14-bands MS microscopic imaging system, which can capture both brightfield and fluorescence images, while brightfield mode is used in this experiment. Vectra®3 uses a white LED as a light source and can obtain the spectral information from 440nm to 700nm at 20nm interval. Next, to obtain the spectral absorption coefficients of the dyes included in Papanicolaou stain, we prepare single stain samples and measure them by the MS microscope. Here, we utilize other than BY, i.e., EY, H, LG, and OG because lipid detection is less important in cytological diagnosis and lipid is hard to dye. Then, we crop some areas in 5×5 pixels from the MS images, and each absorbance is averaged and normalized. The absorbance of EY, H, LG, and OG is shown in Figure 1. As a result, we get the spectral absorption coefficient matrix $\mathbf{H}$, so $\mathbf{c}$ can be estimated using the pseudo-inverse matrix of $\mathbf{H}$.

To present the quantified dye amount to pathologists, it is convenient if the estimated amount of stain $\mathbf{c}$ is normalized. For this purpose, well-stained areas are selected from the single stained images to define the reference value of dye amount, i.e., some nuclei for H, and cytoplasm regions for EY, OG, and LG. We measure the average value of well-stained areas for each dye in advance and use it as the reference value for the dye amount. Namely, $\mathbf{c}$ is divided by the reference value to obtain the normalized dye amount.

To test the proposed method by pathologists, we construct an application software tool that depicts the amount of dyes in the selected area. The bar-graph plots the average of each dye amount in the selected area (see (f) and (g) in Figure 4). We show a simple user-interface panel of the application software and the example of the result in Figure 2. In this application, at first, one chooses a file of an MS image of Papanicolaou-stained specimens. Then, one can select the "image show" mode depicting the dye amount image or the "bar graph" mode. In both modes, the application requires the selection of an un-stained background area like the red area in Figure 2. The pixel values are used to remove the background in the dye amount image. This is because the background is slightly dyed by Papanicolaou stain in some cases. The bar graph mode requires selecting an analysis area too. After the above step, one can view the estimated results by our proposed method in a separate window.

Experiments

We conducted three experiments: 1) verifying the performance of the proposed method to single stain, 2) estimating dye amount maps and bar graph using the proposed application, and 3) applying the proposed method to MS images capturing EC and LEGH cells. In all experiments, we utilized 14-band MS images captured by Vectra®3.

1) Verification of the performance

First, we applied the proposed method to single-stain images. Getting the ground-truth dye amounts is very difficult, so we qualitatively verify the performance of the proposed method using single-stain images. The images captured oral cells stained with H, EY, LG, or OG, and we used 6 each stained images. Each estimated dye amount map are shown in Table 1. The amount was calculated by averaging the values contained in the region

Table 1. The amount of dyes on the single stain separation experiments

observation \ amount of stain	H	EY	LG	OG
H-single stain	**0.1075**	0.0002	0.0022	0.0042
EY-single stain	0.0005	**0.2785**	0.0002	0.0022
LG-single stain	0.0059	0.0064	**0.4867**	0.0043
OG-single stain	0.0001	0.0003	0.0002	**0.8493**

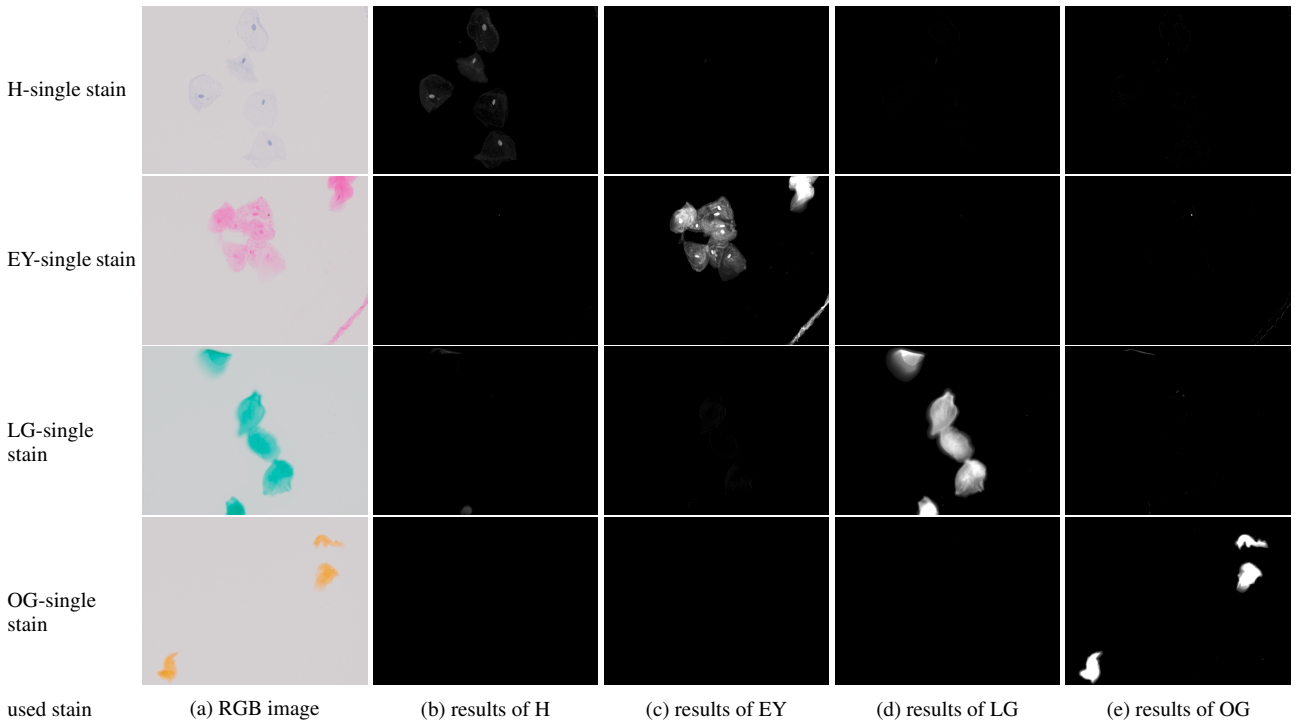


Figure 3. *The estimated results of single-stain images by the proposed method.*

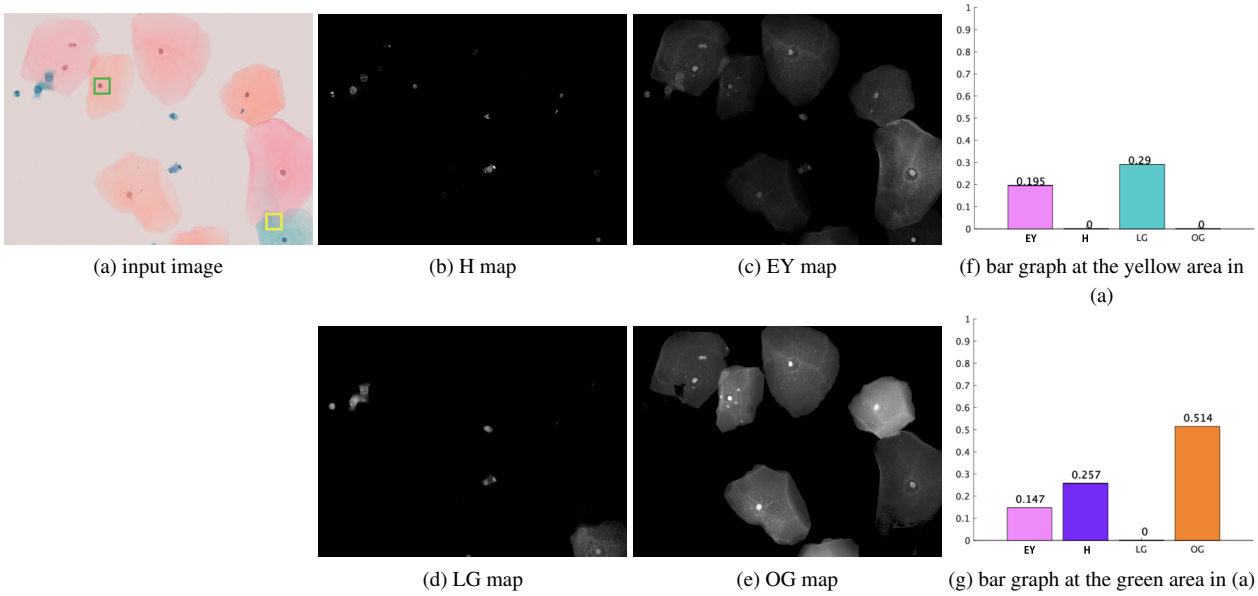
where any stain is non-zero, and then averaging that for each stain. In the table, one can see that the proposed method achieves high estimated performance for all stains. Figure 3 shows the estimated results of single stain images. Here, the RGB images are generated by appropriately transforming the MS images using CIE 1931 color matching function and sRGB standard. The proposed method can estimate H, EY, OG. For LG cases, most area are effectively estimated, but one can see that the results of H and OG have a few dye amounts. This would be because the noise and the limited dynamic range produce distortion to absorbance. The problem will be resolved in future by modeling noise and applying a transformation considering dynamic range limitation.

2) Estimation of dye amount maps and bar graph display

Next, we confirmed the performance of the proposed application. We inputed a Papanicolaou-stain MS image in the application, and then the dye amount maps and bar graphs were obtained using the software tool. The results are shown in Figure 4. From the H, EY, LG, and OG maps, one can visually confirm that the application achieves the separation of dye abundance and reasonably estimates the dye amount. In this experiment, we choose two areas to depict bar graphs, which seem to include different stains. The bar graphs show that the yellow area includes EY and LG, and the green area mainly is dyed to OG and H. From the results, the application would help to analysis the sample dyed by Papanicolaou stain.

3) Application to the color analysis of EC and LEGH cells

We adopted the proposed method to the color analysis of EC and LEGH cell images, where we used 32 EC cell images and 37 LEGH cell images. Some results are shown in Figure 5. The top two rows in the figure are the results on EC cell images, and the bottom two are on LEGH cell images. Figure 5 shows that the mucin of the EC cells is scarcely stained by OG, and EY and LG stain the mucin. For LEGH cells, one can see that the mucin is stained with EY, LG, and OG. For quantitative evaluation, we plot the amount of stain of all results in Figure 6 with p values and effect sizes by Mann-Whitney U test. The amount of dyes was calculated like the previous experiments; average values in the areas excluding the glass regions. The figure shows that the amounts of H and LG are almost the same between EC and LEGH cells. On the other hand, that of EC cells is higher than that of LEGH cells in the EY case, and that of EC cells is lower than that of LEGH cells in the OG case. By Mann-Whitney U test, it is shown that no significant differences are found in H and LG, whereas there are significant differences regarding EY and OG. From the above results, the mucin of EC and LECH cells has cytoplasm of different molecular weight, and one can probably identify them using EY and OG maps. In other words, the mucin of LEGH appears relatively yellowish because of higher



(a) input image (b) H map (c) EY map (f) bar graph at the yellow area in (a)

(d) LG map (e) OG map (g) bar graph at the green area in (a)

Figure 4. *The results on the proposed application software tool.*

concentration of OG and lower concentration of EY.

Conclusion

We have proposed a dye amount estimation method based on color unmixing and analyzed the characteristics of EC and LEGH cells of Papanicolaou stain. The method requires the spectral absorption coefficient of Papanicolaou stain, so we measure it using a 14-bands MS microscopic camera. We also implemented a software tool that shows the estimated dye amount map and the bar graph for the selected area. This application tool will help pathologists to analyze samples using the proposed method. In the experiments, we demonstrate the performance of the proposed method and analyze the characteristics of EC and LEGH cells to Papanicolaou stain. The results show that the mucin of EC cells is more strongly stained with EY, and that of LEGH cells is stained with OG. In the future, we will investigate the possibility of the use of an RGB color image for this purpose.

References

[1] R. S. Weinstein, A. R. Graham, L. C. Richter, et al., "Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future," Human pathology, vol. 40, no. 8, pp. 1057-1069, 2009.

[2] M. R. Nucci, P. B. Clement, R. H. Young, "Lobular endocervical glandular hyperplasia, not otherwise specified: a clinicopathologic analysis of thirteen cases of a distinctive pseudoneoplastic lesion and comparison with fourteen cases of adenoma malignum," The American journal of surgical pathology, vol. 23, no. 8, pp. 886, 1999.

[3] M. R. Nucci, "Pseudoneoplastic glandular lesions of the uterine cervix: a selective review," International Journal of Gynecological Pathology, vol. 33, no. 4, pp. 330-338, 2014.

[4] K. Ishii, T. Katsuyama, H. Ota, et al., "Cytologic and cytochemical features of adenoma malignum of the uterine cervix," Cancer Cytopathology: Interdisciplinary International Journal of the American Cancer Society, vol. 87, no. 5, pp. 245-253, 1999.

[5] K. Ishii, T. Kumagai, M. Tozuka, et al., "A new diagnostic method for adenoma malignum and related lesions: latex agglutination test with a new monoclonal antibody, HIK1083," Clinica chimica acta; international journal of clinical chemistry, vol. 312, no. 1-2, pp, 231-233, 2001.

[6] F. Kimura, T. Kobayashi, R. Kanai, et al., "Image quantification technology of the heterochromatin and euchromatin region for differential diagnosis in the lobular endocervical glandular hyperplasia," Diagnostic cytopathology, vol. 47, no. 6, pp. 553-563, 2019.

[7] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," Analytical and quantitative cytology and histology, vol. 23, no. 4, pp. 291-299, 2001.

[8] K. Fujii, M. Yamaguchi, N, Ohyama, and K. Mukai, "Development of support systems for pathology using spectral transmittance: the quantification method of stain conditions," Medical Imaging 2002: Image Processing, vol. 4684, pp. 1516-1523, International Society for Optics and Photonics, 2002.

[9] R. M. Levenson, P. J. Cronin, and K. K. Pankratov, "Spectral imaging for brightfield microscopy," Spectral Imaging: Instrumentation, Applications, and Analysis II, vol. 4959, pp. 27-33, International Society for Optics and Photonics, 2003.

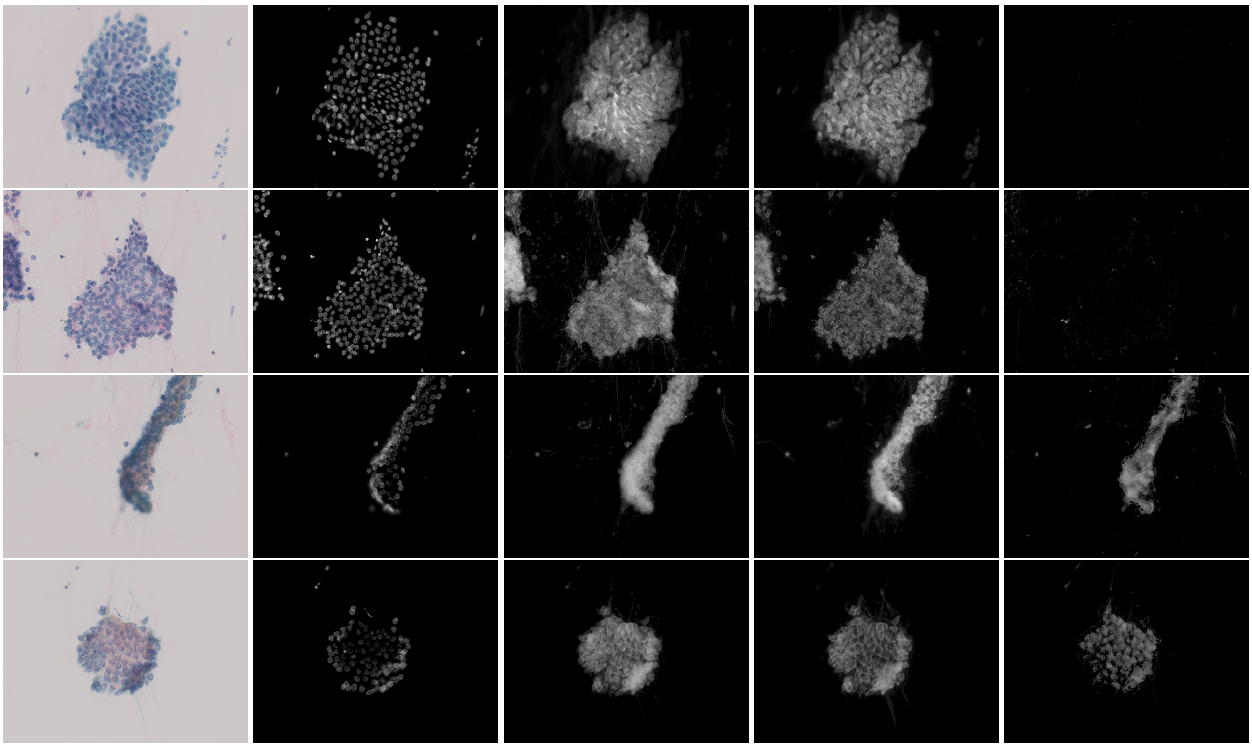
[10] Y. Murakami, H. Gunji, F. Kimura, et al., "Color correction in whole slide digital pathology," Color and Imaging Conference, Society for Imaging Science and Technology, pp. 253-258, 2012.

Author Biography

Saori Takeyama received the B.E. degree in engineering in 2016 and the M.E. and Ph.D. degree in information and communications engineering in 2018 and 2021 from the Tokyo Institute of Technology, respectively. From 2018 to 2021, she was a Research Fellow (DC1) of the Japan Society for the Promotion of Science (JSPS). Since then she has worked as an Assistant Professor in the Department of Information and Communications Engineering, School of Engineering, Tokyo Institute of Technology.

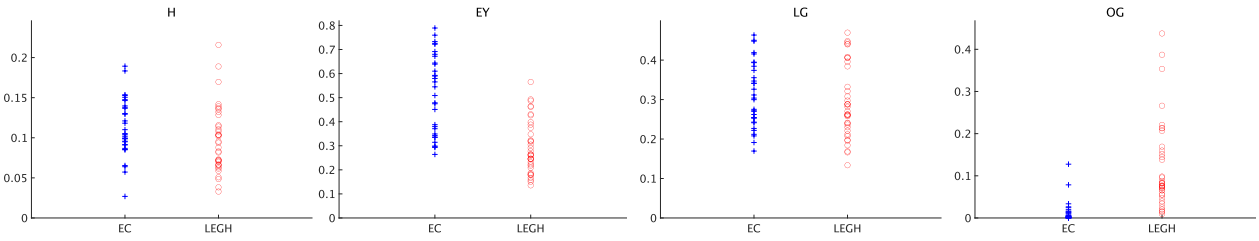
Tomoaki Watanabe received the B.E. degree in engineering in 2019 and the M.E. degree in information and communications engineering in 2021 from the Tokyo Institute of Technology.

Masahiro Yamaguchi is a professor in the School of Engineering, Tokyo Institute of Technology. He received D. Eng. from Tokyo Institute of Technology in 1994. Since 1989 he has been a faculty member of the same Institute, from 2011 to 2015 as a full professor at Global Scientific Information and Computing Center. His research includes color



(a) RGB image (b) results of H (c) results of EY (d) results of LG (e) results of OG

Figure 5. *The estimated results of EC (the top two rows) and LEGH (the bottom two) images by the proposed method.*



$p = 2.147 \times 10^{-7} < 0.001$
effect size: 0.62

$p = 6.050 \times 10^{-2} > 0.05$
effect size: 0.23

$p = 0.2336 > 0.05$
effect size: 0.14

$p = 3.672 \times 10^{-17} < 0.001$
effect size: 1.01

Figure 6. *The histogram of the results with each p value and effect size by Mann-Whitney U test.*

and multispectral imaging, holography, and pathology image analysis. In 2020-2021, Editor in Chief, Optical Review. From 2018, the convener of CIE RF-01, spectral imaging.

Takumi Urata is a Master's student at the School of Health Sciences, Shinshu University. His current research is relevance of cell cycle-related proteins in gynecological diseases.

Fumikazu Kimura received the Ph.D. degree in Medical Science from Department of Clinical Cytology, Graduate School of Medical Sciences, Kitasato University in 2010. He was a Research Fellow of the Global Scientific Information and Computing Center, Tokyo Institute of Technology from 2010 to 2015. Since 2015, he is with the Department of Biomedical Laboratory Sciences, School of Health Sciences, Shinshu University, where he is currently a Junior Associate Professor.

Keiko Ishii received a Medical Doctor and Ph.D. from Shinshu University School of Medicine. She is working as a Pathologist at the Division of Diagnostic Pathology, Okaya City Hospital. Her research includes Lobular endocervical glandular hyperplasia, Gynecological Pathology.