Quantitative analysis of stain variability in histology slides and an algorithm for standardization

Babak Ehteshami Bejnordi[a], Nadya Timofeeva[b], Irene Otte-Höller[b],
Nico Karssemeijer[a], and Jeroen AWM van der Laak[b]

[a] Department of Radiology, Radboud University Medical Center, Nijmegen, The Netherlands;
[b] Department of Pathology, Radboud University Medical Center, Nijmegen, The Netherlands;

ABSTRACT

This paper presents data on the sources of variation of the widely used hematoxylin and eosin (H&E) histological staining, as well as a new algorithm to reduce these variations in digitally scanned tissue sections. Experimental results demonstrate that staining protocols in different laboratories and staining on different days of the week are the major factors causing color variations in histopathological images. The proposed algorithm for standardizing histology slides is based on an initial clustering of the image into two tissue components having different absorption characteristics for different dyes. The color distribution for each tissue component is standardized by aligning the 2D histogram of color distribution in the hue-saturation-density (HSD) model. Qualitative evaluation of the proposed standardization algorithm shows that color constancy of the standardized images is improved. Quantitative evaluation demonstrates that the algorithm outperforms competing methods. In conclusion, the paper demonstrates that staining variations, which may potentially hamper usefulness of computer assisted analysis of histopathological images, can be reduced considerably by applying the proposed algorithm.

Keywords: Stain variability, Standardization, Hematoxylin and Eosin, Hue-saturation-density, Computer-aided diagnosis

1. INTRODUCTION

Traditional diagnosis of cancer involves microscopic examination of histological slides acquired from tissue samples. Tissue sections are treated with multiple contrasting dyes to highlight different tissue structures and cellular features. Pathologists make diagnostic interpretations of the histology slides by assessing the cell structures and their spatial arrangement.[1] The task is laborious and time-consuming and prone to subjectivity.[2,3] Computer-aided diagnosis (CAD) can potentially address the issue of subjective interpretations by providing an objective quantitative assessment of digital pathology slides. CAD systems may facilitate cancer diagnosis by identifying abnormal areas on the slide and providing second opinions for patients.[4] However, variations in staining color and intensity complicate quantitative tissue analysis.[4] Such variations are due to inter-patient variation and inconsistencies in the preparation of histology slides (e.g. staining duration, stain concentration, tissue thickness). Although standardizing the staining protocols can minimize these effects, it is infeasible to remove all sources of variation[5] (e.g. tissue samples stained on different days of the week by following the same staining protocol in the same laboratory may result in different staining hue). Different approaches to overcome this problem have been proposed in literature, based on manipulation of the digital image after scanning of the slide. Methods based on normalization rely on extracting stain vectors and decomposing the image into individual stain components via color deconvolution.[6] The appearance of the images is normalized by adjusting the distribution of the color for each stain to a predefined range.[7] Methods based on color standardization match the color distribution of a histology image into a pre-defined template image by mapping its histogram-specific landmarks to the corresponding landmarks of the template image[8,9] . In[9] it was shown that separate standardization of the histogram for different tissue components yields an improved color constancy over global standardization

approaches. However, the result of this standardization scheme relies on the accuracy of the tissue segmentation algorithm without considering the fact that pixels generally contain mixtures of stains.

This study first investigates the contribution of several factors to the color variation of histology slides stained with the widely used hematoxylin and eosin (H&E) staining. Next, this study presents and evaluates a new algorithm for standardization of histology slides based on an initial clustering of the image into two tissue components having different absorption characteristics for different dyes. Standardization of the color distribution is performed by aligning the 2D histogram of color distribution in the previously described hue-saturation-density (HSD) model.[10]

2. METHODS

2.1 Analysis of color variation in hematoxylin and eosin stained histology slides

The H&E staining is the most commonly used staining technique in histopathology. Hematoxylin stains cell nuclei blue, and the counter-stain eosin stains cytoplasmic and non-nuclear components in different shades of pink.[11] An essential first step in the development of a CAD system is robust segmentation of the objects of interest (e.g. cell nuclei). The appearance of nuclei, however, can vary considerably among histological slides, which can cause significant adverse effects on the results of CAD. To investigate the extent and major reasons behind color variations in histological images, a technique was developed to quantify the distribution of color information. This technique applies the HSD model, which was specifically designed for absorption light microscopy.[10] The HSD model transforms RGB data into two chromatic components (c_x and c_y; which are independent of the amount of stain) and a density component (D; linearly related to the amount of stain). Expectation Maximization (EM)[12] algorithm was employed to estimate the parameters of a Gaussian mixture model in the $c_x c_y$ plane of the HSD transform. Each image is clustered by EM into three broad classes, representing different tissue components: nuclei, cytoplasm/stroma, and background. The probability density function for the ith class is represented by its mean μ_i and covariance matrix Σ_i. To understand the extent of the color variation of the hematoxylin stained tissue class, the variations in the mean value (i.e. perceived color) and the average eigenvalue (intra-specimen color variability) of the probability density function for the class representing the nuclei, is quantitatively described.

2.2 Standardization of histology slides

The proposed algorithm for standardization of histology images is based on an initial clustering of the image into three classes (nuclei, cytoplasm/stroma, background) in the $c_x c_y$ plane of the HSD transform. Background pixels (devoid of stain) show very low density values ($D < 0.2$). Given that the image background is white, all three RGB values have to exceed 180 to be classified as background. By application of singular value decomposition (SVD) on the remaining data points in the $c_x c_y$ plane, the two singular vectors with largest singular values are computed. The angle between each data point and the second singular vector (perpendicular to the first) was calculated. All the data points having an angle below $90°$ degrees were classified as pixels absorbing the majority of hematoxylin stain and the remaining as pixels absorbing the majority of eosin stain. Figure 1 shows the scatter plot of the classified pixels in $c_x c_y$ space.

The process for standardization of each tissue class i involves: (i) computing mean μ_i and covariance matrix Σ_i of the data distribution for class i by modeling it as a Gaussian; (ii) centering the axis of the data points by translating the mean of the data distribution in class i to the point $(0,0)$ and rotating the entire data distribution along the major eigenvector of Σ_i to maximize the variance along the axes; (iii) computing landmarks $\{c_{min}, c_1, c_{median}, c_{99}, c_{max}\}$ for the rotated c_x and c_y values independently, where c_1, c_{median} and c_{99} denote the $1st$, $50th$ and $99th$ percentiles of the data points belonging to class i, and c_{min} and c_{max} denote the minimum and maximum values of the entire distribution; (iv) interpolating the data distribution to match the predefined corresponding landmarks from the template image $\{t_{min}, t_1, t_{median}, t_{99}, t_{max}\}$ by performing piecewise linear mappings – from $[c_{min}, c_1]$ to $[t_{min}, t_1]$, $[c_1, c_{median}]$ to $[t_1, t_{median}]$, etc.; (v) rotating back the entire distribution along the major eigenvector of the corresponding tissue class in the template image; (vi) translating the entire distribution to the point (c_{xm}, c_{ym}) where c_{xm} and c_{ym} are the mean c_x and c_y values of the data distribution for the corresponding tissue class in the template image.
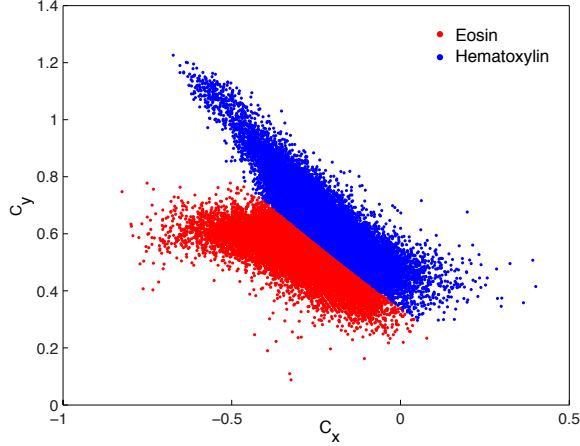
Figure 1: Scatter plot of all the pixels classified as hematoxylin and Eosin in $c_x c_y$ space by application of singular value decomposition.

A weighted sum of the transformed $c_x c_y$ values for each class yields the final $c'_x c'_y$ for the entire distribution. The weight function is computed by first estimating the distribution of the two classes as two Gaussian components $k \in \{1, 2\}$ and then calculating their corresponding posterior probabilities for each observation X (excluding background observations) in the original $c_x c_y$ plane through $P(K \mid X) = \frac{\pi_k N(X|\mu_k, \Sigma_k)}{\sum_{j=1}^{k} \pi_j N(X|\mu_j, \Sigma_j)}$, where π_k is the prior probability and $N(X \mid \mu_k, \Sigma_k)$ is the probability density function of component K for the observation X. This yields the following transformation:

$$c'_x c'_y = c_{x_1} c_{y_1} \times P(K = 1 \mid c_{x_1} c_{y_1}) + c_{x_2} c_{y_2} \times P(K = 2 \mid c_{x_2} c_{y_2}) \tag{1}$$

where $c_{x_1} c_{y_1}$ and $c_{x_2} c_{y_2}$ are the transformed $c_x c_y$ values belonging to each class, and $c'_x c'_y$ is the yielded final transformation. After scaling the density histogram of the image to the density histogram of the template image using 10 evenly spaced percentiles as landmarks, the RGB channels are reconstructed by the reverse HSD transform.[10]

2.3 Histology images

The image data used in this study originate from a set of 45 digitized H&E stained histopathology slides of lymph nodes from three different patients. The slides were stained in three different laboratories on different days of the week (15 slides for each lab). All the slides were digitized using a CCD RGB camera mounted on a light microscope with a $40\times$ objective lens (Olympus dotSlide system, Olympus, Japan). Three representative regions of interest (ROI) images were acquired from each slide yielding a total of 135 images. Each image is of size 2300×3300 pixels with square pixels of size $0.16\ \mu m$ in the microscope image plane. Figure 2 shows three sample images stained in different laboratories.

2.4 Experiments

To evaluate the major factors causing stain variations in histology images and to understand the extent of these changes and also to evaluate the performance of the proposed color standardization algorithm two experiments were performed.

The aim of the first experiment was to statistically measure the variations in the mean (c_x, c_y) values (describing absolute hematoxylin color) and the average eigenvalues (describing intra-slide hematoxylin variation) of the probability density function for the class representing the nuclei in the $c_x c_y$ plane of the HSD model. The importance of three potential factors that contribute to staining variations including patient, staining laboratory and staining day of the week was studied. Provided that the staining solutions used by the automated staining
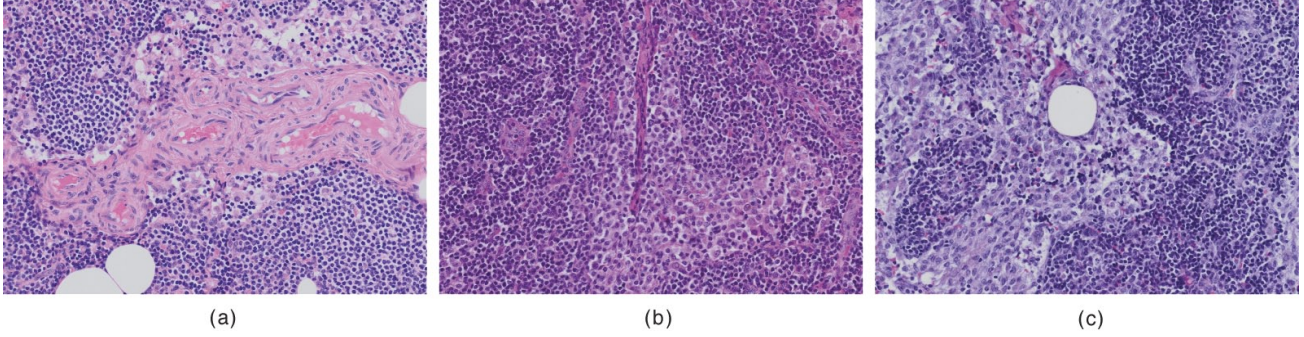
Figure 2: Sample images from the three laboratories. (a), (b), and (c) are three sample ROI images stained by laboratory 1, 2 and 3 respectively.

machines is changed once or twice per week, staining on different days of the week may result in a stain having different spectral characteristics. Multivariate analysis of variance (MANOVA)[13] with Pillai's statistics was used to determine the statistical significance of differences between these variables.

In the second experiment, the performance of the proposed standardization algorithm is both qualitatively and quantitatively evaluated. Qualitative evaluation is performed by visualizing the original image and assessing the visual color constancy of the standardized image with respect to the template image. For quantitative evaluation of the results, the color constancy of the nuclei segmented regions were evaluated and compared to the global standardization (GS) algorithm[8] and the appearance normalization method[7] by calculating the normalized median intensity (NMI).[9, 14] The standard deviation of the NMI values and the coefficients of the variations (i.e. standard deviation divided by mean) were computed for all the images in the dataset before and after standardization using the proposed algorithm. The nuclei were segmented by imposing a threshold of $T = 120$ (determined empirically) on the average of the RGB values for each image.

3. RESULTS

Experiment 1: MANOVA results show that differences between labs may be considerable, and that for certain labs staining results differ significantly between days of the week (Table 1). Comparable results were found for the mean (c_x, c_y) values and the average eigenvalues of the nuclei. No significant color variations were observed between tissues from different patients.

Table 1: Multivariate analysis of variance with Pillai's statistics

Variables	c_x, c_y	Average eigenvalues
Patients	generally no significant effect	no effect
Laboratory	significant differences between $L3$ and $L1/L2$ ($p < 0.004$) weak difference between $L1$ and $L2$	significant differences between $L3$ and $L1/L2$ ($p < 0.004$) weak difference between $L1$ and $L2$
Days of the week	$L1$ ($p < 0.003$) $L2$ not significant $L3$ ($p < 0.004$)	$L1$ ($p < 0.02$) $L2$ not significant $L3$ ($p < 0.002$)
$L1$: Laboratory 1, $L2$: Laboratory 2, $L3$: Laboratory 3		

Figure 3 shows boxplots of the average eigenvalue distributions of the nuclei segmented regions of the slides

stained in different laboratories on different days of the week. The mean and standard deviation of the average eigenvalues for different days of the week are considerably higher for the third laboratory.
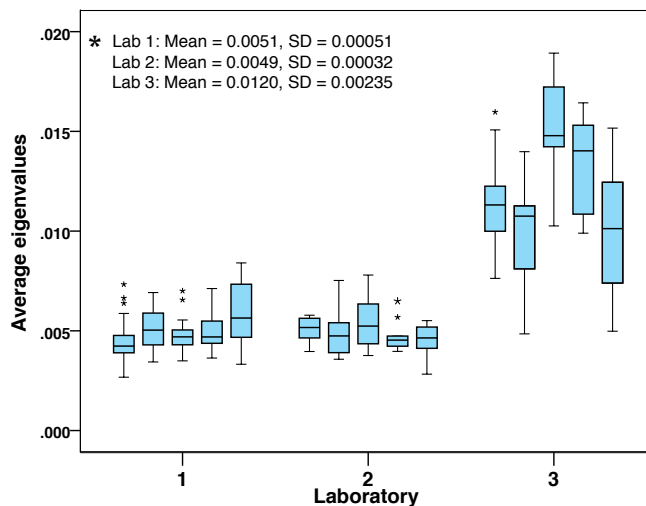

Figure 3: Boxplots of average eigenvalue distributions for the three laboratories for different days of the week. The mean and standard deviation have been calculated for the five weekday distributions for each laboratory.

Experiment 2: Figure 4 shows the result of the standardization by the proposed algorithm. The original images in Figure 2b and Figure 2c are standardized by using the image in Figure 4a as the template image. The standard deviation (SD) of the normalized median intensity and its coefficient of variation (CV) for the segmented nuclei are presented in Table 2 for the three laboratories. The proposed method outperforms GS algorithm and the appearance normalization method by yielding the lowest SD and CV of normalized median intensity.
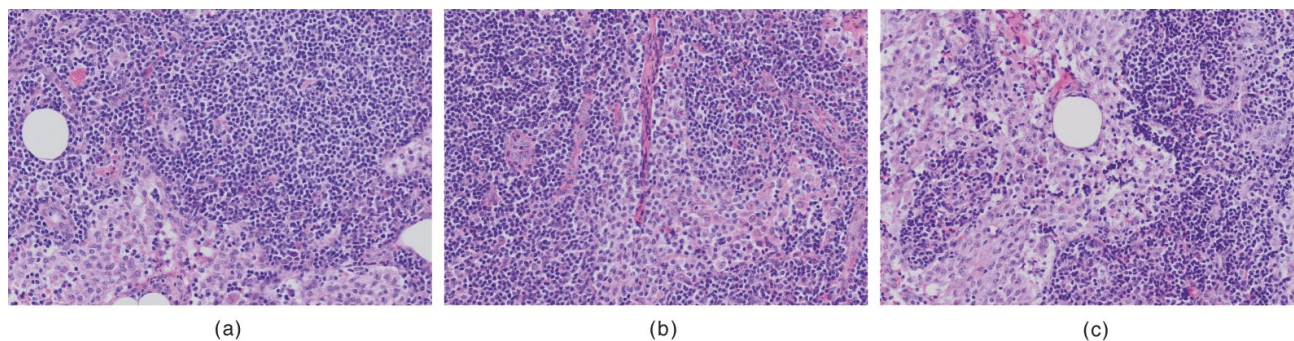

Figure 4: Standardization of H&E stained histopathology images. (a) The template image used for extracting the parameters of the standardization algorithm. (b) The standardized image of the image shown in Figure 2b (c) The standardized image of the image shown in Figure 2c.

Table 2: Standard deviation and coefficient of variation of NMI for all the images in the three laboratories.

	Laboratory 1		Laboratory 2		Laboratory 3	
	NMI SD	NMI CV	NMI SD	NMI CV	NMI SD	NMI CV
Original	0.0363	0.0450	0.0260	0.0317	0.0367	0.0484
Macenko[7]	0.0251	0.0275	0.0207	0.0228	0.341	0.4520
Global standardization[8]	0.0240	0.0277	0.0203	0.0238	0.0219	0.0256
Proposed algorithm	**0.0068**	**0.0081**	**0.0077**	**0.0093**	**0.0094**	**0.0123**

4. CONCLUSIONS AND DISCUSSION

This paper investigated the effect of several factors on the color variation of hematoxylin and eosin stained histology slides. The experimental results demonstrate that staining protocols in different laboratories and staining on different days of the week are the major factors causing color variations in histopathology images. This paper also presented a new algorithm for reducing stain variations in histology slides. Qualitative assessment of the results show the efficacy of the algorithm in maintaining color constancy of the histology images. The empirical results show that the proposed algorithm outperforms the global standardization algorithm and the appearance normalization method (based on automatic derivation of stain vectors) by yielding a lower standard deviation and coefficient of variation of the normalized median intensity. The algorithm presented in this paper can be applied to other histological stains and tissues. This will be the subject of future work.

REFERENCES

[1] Ghaznavi, F., Evans, A., Madabhushi, A., and Feldman, M., "Digital imaging in pathology: Whole-slide imaging and beyond," *Annual Review of Pathology: Mechanisms of Disease* **8**, pp. 331–359 (2013).

[2] Ismail, S. M., Colclough, A. B., Dinnen, J. S., Eakins, D., Evans, D., Gradwell, E., O'Sullivan, J. P., Summerell, J. M., and Newcombe, R. G., "Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia.," *BMJ: British Medical Journal* **298**(6675), p. 707 (1989).

[3] Andrion, A., Magnani, C., Betta, P., Donna, A., Mollo, F., Scelsi, M., Bernardi, P., Botta, M., and Terracini, B., "Malignant mesothelioma of the pleura: interobserver variability.," *Journal of clinical pathology* **48**(9), pp. 856–860 (1995).

[4] Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., and Yener, B., "Histopathological image analysis: A review," *Biomedical Engineering, IEEE Reviews in* **2**, pp. 147–171 (2009).

[5] Niethammer, M., Borland, D., Marron, J., Woosley, J., and Thomas, N. E., "Appearance normalization of histology slides," in *Machine Learning in Medical Imaging*, pp. 58–66, Springer (2010).

[6] Ruifrok, A. C. and Johnston, D. A., "Quantification of histochemical staining by color deconvolution.," *Analytical and quantitative cytology and histology/the International Academy of Cytology [and] American Society of Cytology* **23**(4), pp. 291–299 (2001).

[7] Macenko, M., Niethammer, M., Marron, J., Borland, D., Woosley, J. T., Guan, X., Schmitt, C., and Thomas, N. E., "A method for normalizing histology slides for quantitative analysis," in *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, pp. 1107–1110, IEEE (2009).

[8] Bagcı, U. and Bai, L., "Registration of standardized histological images in feature space," in *Proc. SPIE Medical Imaging 6914*, pp. 69142V–1 (2008).

[9] Basavanhally, A. and Madabhushi, A., "Em-based segmentation-driven color standardization of digitized histopathology," in *Proc. SPIE Medical Imaging*, pp. 86760G–86760G, International Society for Optics and Photonics (2013).

[10] van der Laak, J. A., Pahlplatz, M. M., Hanselaar, A. G., and de Wilde, P., "Hue-saturation-density (hsd) model for stain recognition in digital images from transmitted light microscopy," *Cytometry* **39**(4), pp. 275–284 (2000).

[11] Fischer, A. H., Jacobson, K. A., Rose, J., and Zeller, R., "Hematoxylin and eosin staining of tissue and cell sections," *Cold Spring Harbor Protocols* **2008**(5), pp. pdb–prot4986 (2008).

[12] Bishop, C. M. and Nasrabadi, N. M., [*Pattern recognition and machine learning*], vol. 1, pp. 430–439, springer New York (2006).

[13] Rencher, A. C. and Christensen, W. F., [*Methods of multivariate analysis*], vol. 709, John Wiley & Sons (2012).

[14] Nyúl, L. G., Udupa, J. K., and Zhang, X., "New variants of a method of mri scale standardization," *Medical Imaging, IEEE Transactions on* **19**(2), pp. 143–150 (2000).