BLIND STAIN DECOMPOSITION FOR HISTO-PATHOLOGY IMAGES USING CIRCULAR NATURE OF CHROMA COMPONENTS

Xingyu Li, Konstantinos N. Plataniotis

Multimedia Lab, The Edward S. Rogers Department of Electrical and Computer Engineering
University of Toronto, 10 King's College Road, Toronto, Canada

ABSTRACT

In this paper, we present a novel approach to achieve blind stain decomposition in histo-pathology images. The method is based on stain color estimation, followed by stain absorbing vector generation and matrix computation. Unlike conventional approaches adopting linear processing algorithms to analyze chromatic information in the cylindrical-coordinate color spaces, which may be inappropriate for circular data such as hue, we propose the use of circular thresholding on saturation-weighted hue histogram to compute candidates for stain representative colors. Experimental results suggest that our stain decomposition method is capable to address spectral variation in stains effectively. We compare the proposed method to state-of-the-art blind stain separation algorithms for nuclei segmentation on breast histo-pathology images, and demonstrate that the segmentation scheme adopting our method in its pre-processing step achieves the best segmentation results.

Index Terms— Histo-pathology image, circular thresholding, saturation-weighted hue histogram, stain separation.

1. INTRODUCTION

In digital histo-pathology, color plays a vital role in image analysis due to the use of chemical stains to highlight different histological components. When analyzing such color information, stain decomposition is frequently performed as a pre-processing step in various digital histo-pathology tasks for two purposes. First, stain decomposition is capable to address the stain co-localization (or color mixing) issue for histological component detection and segmentation [1, 2]. Second, appearance variation in histo-pathology images is a key challenge in automatic histo-pathology image analysis [3]. Stain decomposition followed by stain normalization can be used to address this issue [4, 5].

Stain decomposition, or stain separation, is a signal analysis process that estimates stain spectra and corresponding stain proportions in a histo-pathology image according to the Beer-Lambert law [6]. Assuming N types of stains used in tissue staining, image color $I(p) = [r(p), g(p), b(p)]$ at location $p = (x, y)$ generated by transmitted light that is not absorbed by stains can be formulated as $I(p) = I^i exp(-M \times D(p))$, where $I^i = [r^i, g^i, b^i]$ denotes imaging illumination coming from back of tissue samples, M is a N-by-3 spectrum matrix whose j^{th} row, $m_j = [m_{jr}, m_{jg}, m_{jb}]$ for $1 \leq j \leq N$, is the j^{th} stain's spectrum vector quantified by its absorbing factors in the red, green, and blue channels, and $D(p) = [d_1(p), ..., d_N(p)]'$ is so-called stain proportions, or stain density map, describing amount of stains bounded at location p. By moving I^i to the left and taking logarithm on both sides, the more frequently-used optical density (OD) domain equivalent expression is as follows:

$$log(I^i/I(p)) = M \times D(p). \quad (1)$$

Once the spectrum matrix M is determined, stain density map $D(p)$ can be derived by matrix inverse computation. Specifically, in the fixed-matrix (FM) decomposition method [7], matrix M was pre-determined via experiments. To address spectral variation in stains, many adaptive stain separation approaches were proposed, while most of them required human intervention [5, 8, 9]. To achieve fully-automatic stain decomposition, a plane-fitting (PF) method [10] is achieved by singular value decomposition and thresholding for images containing two stains only. Later, blind color decomposition (BCD) proposed to perform expectation-maximization on color distributions in the Maxwell color triangle to estimate matrix M [11]. Though a heuristic randomization function tries to select stable colors for stain estimation, BCD method is prone to be affected by achromatic pixels when estimating spectrum vectors for weak stains.

In this paper, we introduce a novel blind stain decomposition method for histo-pathology images generated by light-absorbing stains (e.g. Hametoxylin and eosin (H&E)). The proposed method generates a pool of candidates for stain representative color, which is a key in our method, followed by stain spectrum vector generation, and achieves optimal stain decomposition in terms of minimizing decomposition residue. Particularly, unlike conventional image processing works that apply linear analysis algorithms to image color which may be measured by circular data, stain color estimation in this paper adopts circular thresholding and

Thanks for Patrick Reding' involvement in this study.

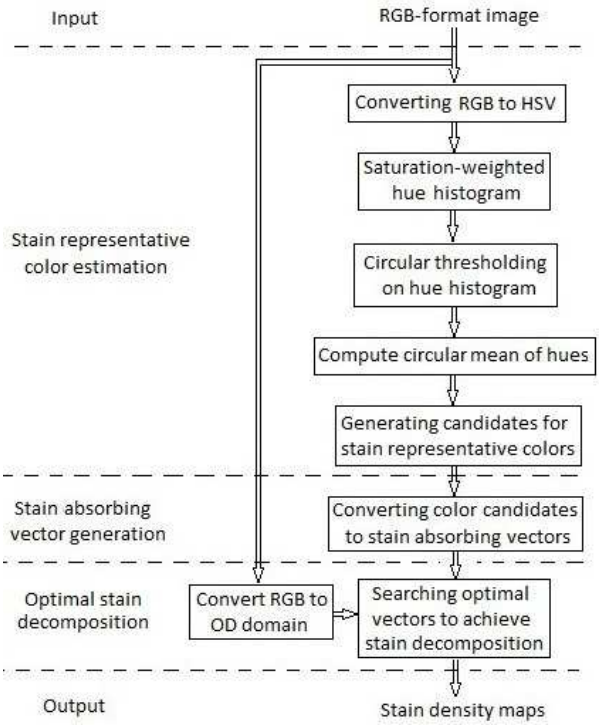

Fig. 1. Block diagram of the proposed blind stain decomposition method.

saturation-weighted hue histogram to address the periodic nature of hue and close relationship between saturation and hue components in color description, respectively. Experimentation demonstrates that the proposed method addresses spectral variation in stains effectively and improves nuclei segmentation in breast cancer images.

The rest of this paper is organized as follows. Section 2 presents the proposed stain decomposition method in detail. Evaluation protocols and corresponding experimental results are given in Section 3, followed by conclusions in Section 4.

2. PROPOSED STAIN DECOMPOSITION

The block diagram of our blind stain decomposition method is illustrated in Fig.1. Given a query image, candidates for stain representative colors are generated by circular thresholding on saturation-weighted hue histogram in the HSV color space. Then each color candidate is converted to a light-absorbing vector in the OD domain. Finally, optimal spectrum vectors are selected from resulting vector pools to form spectrum matrix M, and stain density map $D(p)$ is obtained via matrix inverse operation.

The input of the proposed stain decomposition approach is a linear RGB-format histo-pathology image, which is assumed to 1) contain light-absorbing stains only, and 2) be imaged under a standard illumination I^i, such as CIE illuminant D_{65} [12], so that image colors are mainly determined by stains' light-absorbing activities. The outputs of our scheme are density maps, a series of single-stain images.

2.1. Stain Representative Color Estimation

Though the RGB color space is widely used in image sensing and display, it is not suitable for image analysis due to its less coherent to human visual perception. To estimate stain representative colors from a query image, we convert an RGB-format input to the HSV color space, where colors are described by hue and saturation on the chromatic plane.

In histo-pathology images, colors are generated by transmitted light that is not absorbed by stains [13]. A hue histogram of an image actually shows what spectra of light are not absorbed, and thus can be used to estimate stains' absorbing spectra. Remind that in the HSV color model, a hue value of an achromatic pixel, or a low-saturated pixel, is meaningless for color description. In particular, less saturated pixels in histo-pathology images are usually noise or tissues bounded by few stains. To limit impacts of such achromatic colors on stain color estimation, this work uses a saturation-weighted hue histogram [14] defined as follows:

$$H_\theta^{sw} = \sum_p s(p)\delta(\theta, h(p)), \tag{2}$$

$$\text{where} \quad \delta(\theta, h(p)) = \begin{cases} 1 & \text{if } \theta = h(p) \\ 0 & \text{otherwise} \end{cases}.$$

$\theta \in [0°, ..., 360°)$ represents a bin in histogram, $s(p)$ and $h(p)$ are the saturation and hue at pixel p in the HSV color space. Fig.2 (b) and (c) depict 360-bin standard histogram and saturation-weighted histogram for an H&E stained image in Fig.2 (a). The standard hue histogram has more spikes contributed by achromatic pixels; while the saturation-weighted histogram shows smooth dominant colors which are contributed by reliable image pixels and thus more coherent to human's color perception.

Observing that clusters in the saturation-weighted histogram associate with different stains contained in a query image, we propose the use of circular thresholding [15], which addresses the angular nature of hue, to divide a histogram into non-overlapping stain regions. Specifically, hue is a periodic measurement in the range of $[0, 360)$. Dividing hue histogram into N stain regions requires N thresholds; whereas conventional linear thresholding methods, such as the Otsu's method [16], generate $N-1$ thresholds only. An implicit assumption when applying linear thresholding to circular data is that the angular value of 0 is always a threshold. However, this assumption hardly holds as pixels whose hues are around 0 and 360 may be all observed red. Take the image in Fig.2 (a) as an example. To highlight the angular attribute of hue, the circular equivalent 360-bin saturation-weighted histogram of the image is depicted in Fig.2 (d). By applying circular thresholding to it, two thresholds labeled as th_1 and th_2 divide the hue histogram into H region and E region.

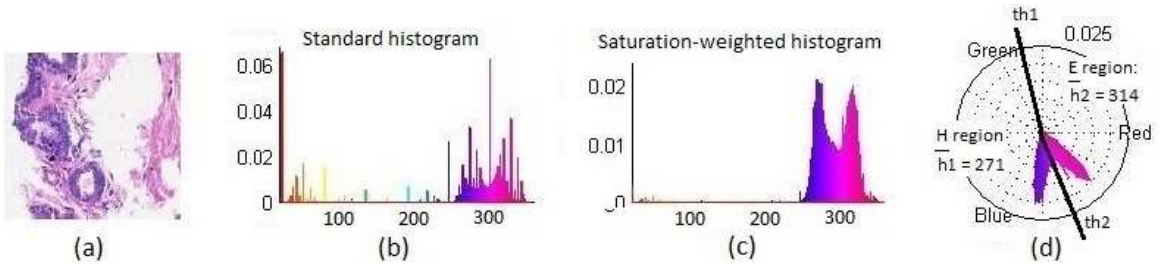

Fig. 2. Example of stain representative hue estimation. (a) Original query H&E stained breast cancer image; (b) Standard hue histogram of the query image; (c) Saturation-weighted hue histogram; (d) Circular thresholding on the circular equivalent saturation-weighted histogram.

Assuming that a hue histogram has L bins, representative hue values of the N stain regions can be derived by formula of circular mean as follows:

$$\overline{h}_k = \arctan\left(\frac{\sum_{\theta=th_k}^{th_{k+1}-1} H_\theta^{sw}\cos\theta}{\sum_{\theta=th_k}^{th_{k+1}-1} H_\theta^{sw}\sin\theta}\right), 1 \le k < N, \quad (3)$$

$$\overline{h}_N = \arctan\left(\frac{\left[\sum_{\theta=th_N}^{L-1} + \sum_{\theta=0}^{th_1-1}\right] H_\theta^{sw}\cos\theta}{\left[\sum_{\theta=th_N}^{L-1} + \sum_{\theta=0}^{th_1-1}\right] H_\theta^{sw}\sin\theta}\right). \quad (4)$$

Then we proceed to compute value and saturation components in the HSV color space for each stain, finally generating candidates for stain representative color. For this end, we exploit the normalization constraint of a spectrum matrix, and derive the value and saturation components as follows. Assume that a stain is observed in a color $c_{hsv} = [h, s, v]$ in the HSV color space, where h is derived following Eqn.(3) or Eqn.(4). Let $f_{hsv}^{rgb}(\cdot)$ denote the function to convert a color in the HSV model to the RGB color space. Then $c_{rgb} = [r, g, b] = f_{hsv}^{rgb}(c_{hsv})$ represents the corresponding stain color in the RGB color space, where $r, g, b \in [0, 1]$. As any stain's absorbing vector in the spectrum matrix has a unit length, it requires $\|log(I^i/c_{rgb})\|^2 = log^2(r^i/r) + log^2(g^i/g) + log^2(b^i/b) = 1$, which can be expressed as

$$\|log(I^i/f_{hsv}^{rgb}(c_{hsv}))\|^2 = 1. \quad (5)$$

Given a value v, cooperating with an estimated hue h from Eqn.(3) or Eqn.(4), a value of saturation can be obtained by solving Eqn.(5). Then, the derived color c_{hsv} is converted to the RGB color space via $c_{rgb} = f_{hsv}^{rgb}(c_{hsv})$. Since v lies in the range of $[0, 1]$, we build a color pool for each stain by letting $v = k\delta$, where the step-size $\delta \in [0, 1]$ and $k = 1, ..., floor(1/\delta)$.

2.2. Stain Absorbing Vector Generation

An absorbing vector in the spectrum matrix characterizes the deduction amount of light in the OD domain when light transmits through a stain [7]. Hence, for each stain whose representative color is c_{rgb}, its corresponding absorbing vector in the OD domain is $m = log(I^i/c_{rgb})$. As we build a color pool for each stain, each stain possesses a pool of candidates for stain absorbing vectors.

2.3. Optimal Stain Decomposition

For a biopsy sample stained by N types of stains, the N-by-3 spectrum matrix M is formed by using $m_j, j = 1, ..., N$ as the j^{th} column. Since each stain has a pool of candidates for its absorbing vector, we need to find one appropriate vector from each pool to form the matrix M. In this work, taking into account the physical constraint that stain proportions $D(p)$ must be non-negative, we propose the search of stains' optimal absorbing vectors in terms of minimizing the expected decomposition mean-squared residue over all pixels p:

$$\arg \min_{m_j, 1\le j\le N} \mathbf{E}_p\left[\|log(I^i/I(p)) - M \times D(p)\|^2\right] \quad (6)$$

$$\text{where } \begin{cases} M = [m_1; ...; m_N] \\ D(p) = \max(M^{-1} log(I^i/I(p)), 0). \end{cases}$$

With the search of optimal absorbing vectors m_j for $j = 1, ..., N$, a spectrum matrix M and stain density maps D are obtained simultaneously.

3. EXPERIMENTAL RESULTS

In this work, the UCSB breast cancer cell dataset [17], which consists of 58 H&E stained breast cancer histo-pathology images stored in 24-bit RGB format, is selected as our evaluation dataset for two reasons. First, spectral variation in H&E stains in this dataset enables us to assess performance of the proposed method against stain variation. Second, groundtruth of nuclei segmentation in region of interest for each image is provided in this dataset. Note, we take H&E stained images as evaluation data in this work because H&E staining is the dominant staining protocol in histo-pathology; However, our method is applicable to histo-pathology images stained by other light-absorbing chemical dyes.

To evaluate the proposed stain decomposition method qualitatively and quantitatively, three experiments are conducted on the UCSB image set as follows.

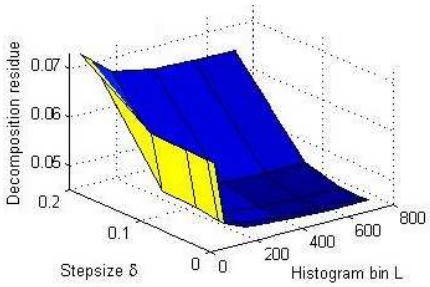

Fig. 3. Parameters analysis of our stain decomposition.

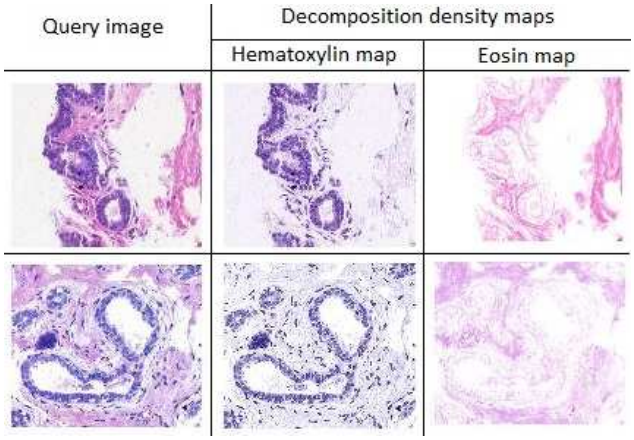

Fig. 4. Examples of stain separation by our method. Color variation in images is attributed to spectral variation in stains.

System Robustness to Parameter Settings: There are two parameters pre-determined in our stain separation method: the number of bins in a hue saturation-weighted histogram, L, and the search stepsize of v in the HSV color space, δ. We perform a parameter analysis using the H&E stained image in Fig.2(a) to check different combinations of L and δ, which is shown in Fig.3. The vertical axis represents the expected decomposition residue computed by $\mathbf{E}_p[\|log(I^i/I(p)) - M \times D(p)\|^2]$. The small decomposition residues in order of 10^{-2} imply that our method is insensitive to system settings, especially when $L \geq 180$ and $\delta \leq 0.1$. For decomposition accuracy and computational efficiency, we set $L = 360$ and $\delta = 0.05$ in all our experiments.

Qualitative Evaluation of Stain Decomposition: In this experiment, we qualitatively examine capability of the proposed method to address the stain variation issue which is often observed among histo-pathology images. As similar observations are obtained, two stain separation examples are presented in Fig. 4. Though nuclei and cytoplasm in the upper H&E stained image appear more magenta and more pink, respectively, compared to histological components in the lower query image, our method works on both images successfully.

Nulcei Segmentation After Stain Decomposition: Segmentation of nuclei in histo-pathology images is a key task for cancer diagnosis. This experiment quantitatively evaluates performance of a nuclei segmentation algorithm [2] on the UCSB dataset when its color unmixing step exploits different stain separation methods: the FM method [7], the PF approach [10], the BCD approach [11], and the proposed method. Specifically, we first tune parameters of the segmentation algorithm particularly on a reference image in the UCSB dataset. Then the parameter set is applied to other images in the dataset. Segmentation performance is evaluated using the dice coefficient [2], which measures the agreement between groundtruth X and segmentation result Y via $D(X,Y) = 2|X \cap Y|/(|X| + |Y|)$. $D(X,Y) \in [0,1]$ and $D(X,Y) = 1$ indicates identical segmentation.

Table 1. Dice coefficients of nuclei segmentation using various stain decomposition approaches over the UCSB dataset

Methods	Mean (Std)	Min	Max
FM method [7]	0.7264 (0.0069)	0.4746	0.8571
PF method [10]	0.7465 (0.0034)	0.5783	**0.9019**
BCD method [11]	0.7309 (0.0062)	0.5366	0.8727
Proposed method	**0.7469** (0.0046)	**0.5814**	0.8728

The statistics of dice coefficients of nuclei segmentation over the UCSB dataset are shown in Table 1, where the largest dice coefficients are marked black. Taking performance of the FM method [7] as comparison baseline, we get two observations. First, all three blind decomposition methods improve nuclei segmentation as they can address stain variation in different degrees. Particularly, the minimum dice coefficient over the UCSB image set is improved more than 0.1 by our method. It should be note that in medical-target tasks, 0.01 improvement may mean significance. Second, though the proposed method and the PF approach [10] achieve close dice coefficients, our method outperforms the PF approach since the PF approach works on images containing two stains only, whereas the proposed method has no such constraints.

4. CONCLUSION

This paper introduced a blind stain decomposition method for histo-pathology images, which was capable of addressing spectral variation in stains. The introduced approach was able to select reliable pixels from a query image for subsequent image analysis, and generated optimal spectrum matrix to minimize decomposition residue using an optimization search algorithm. Especially, to achieve accurate stain color estimation which is a key in our method, we made use of saturation-weighted histogram to limit impacts of achromatic colors, and adopted circular thresholding to address the periodicity of hue properly when processing color information. Experimental results demonstrate superiority of our method in nuclei segmentation over state-of-the-art blind stain decomposition methods.

5. REFERENCES

[1] E. Ozdemir and C. Gunduz-Demir, "A hybrid classification model for digital pathology using structural and statistical pattern recognition," *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 474 – 483, Feb. 2013.

[2] M. Veta1, P. J. Van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J.P. W.Pluim, "Automatic nuclei segmentation in h&e stained breast cancer histopathology images," *PloS One*, vol. 8, no. 7, Jul. 2013.

[3] M. Veta, J.P.W. Pluim, P.J. van Diest, and M.A. Viergever, "Breast cancer histopathology image analysis: A revier," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1400 – 1411, May 2014.

[4] C.C. Bilgin, J. Rittscher, R. Filkins, and A. Can, "Digitally adjusting chromogenic dye proportions in brightfield microscopy images," *J. Microsc.*, vol. 245, no. 3, pp. 319 – 330, 2011.

[5] A.M. Khan, N. Rajpoot, D. Treanor, and D. Magee, "A non-linear mapping approach to stain normalization in digital histopathology images using image-specic color deconvolution," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1729 – 1738, Jun. 2014.

[6] J.D.J. Ingle and S.R. Crouch, *Spectrochemical Analysis*, Prentice Hall, New Jersey, 1988.

[7] A.C. Ruifrok and D.A. Johnston, "Quantification of histochemical staining by color deconcolution," *Anal. Quant. Cytol. Histol.& Int. Acad. Cytol. Amer. Soc. Cytol*, vol. 23, no. 4, pp. 291 – 299, 2001.

[8] A.C. Ruifrok, R.L. Katz, and D.A. Johnston, "Comparison of quantification of histochemical staining by hue-saturation-intensity (hsi) transformation and color-deconvolution," *Appl. Immunohistochem. Mol. Morphol.*, vol. 11, no. 1, pp. 85 – 91, 2003.

[9] S. Tani, Y. Fukunaga, S. Shimizu, M. Fukunishi, K. Ishii, and K. Tamiya, "Color standardization method and system for whole slide imaging based on spectral sensing," *Analytical Cellular Pathology*, vol. 35, no. 2, pp. 107 – 115, 2012.

[10] M. Macenko, M. Niethammer, J. Marron, D. Borland, J. Woosley, X. Guan, C. Schmitt, and N.E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro*, 2009.

[11] M. Gavrilovic, J.C. Azar, J. Lindblad, C. Wahlby, E. Bengtsson, C. Busch, and I.B. Carlbom, "Blind color decomposition of histological images," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 983 – 994, Jun. 2013.

[12] N. Ohta and A. R. Robertson, *CIE Standard Colorimetric System. Colorimetry: Fundamentals and Applications*, Wiley, 2006.

[13] D. Magee, D. Treanor, D. Crellin, M. Shires, K. Smith, K. Mohee, and P. Quirke, "Color normalization in digital histopathology images," in *Proc. Opt. Tissue Image Anal. Microsc., Histopathol. Endosc.*, 2009.

[14] A. Hanbury, "Circular statistics applied to color images," in *8th Computer Vision Winter Workshop*, 2003, vol. 91, pp. 53 – 71.

[15] Y-K Lai and P. Rosin, "Efficient circular thresholding," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 992 – 1001, Mar. 2014.

[16] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. on Syst., Man Cybern.*, vol. 9, no. 1, pp. 62 – 66, Jan. 1979.

[17] E.D. Gelasca, J. Byun, B. Obara, and B.S. Manjunath, "Evaluation and benchmark for biological image segmentation," in *Proc. IEEE Int. Conf. Image Process.*, 2008.