DIAGNOSTIC COLOR ESTIMATION OF TISSUE COMPONENTS IN PATHOLOGY IMAGES VIA VON MISES MIXTURE MODEL

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 diagnostic color estimation for histological objects in pathology images. The method is based on a von Mises mixture model for hue histogram, followed by implicit pixel clustering via maximum likelihood estimation and representative color computation. Unlike conventional approaches adopting linear processing algorithms to analyze hue histogram which is characterized by a nature of periodicity, we build a circular cluster model composed of multiple von Mises distributions to address the directional nature of hue. Experimental results on synthetic circular data suggest that the proposed circular model outperforms both classical linear thresholding methods and the state-of-art circular thresholding approach in terms of cluster parameter estimation. The color estimation experiment on publicly-accessible cytopathology images demonstrates that our method is capable to accurately estimate object's diagnostic color, which can be used for subsequent image analysis.

Index Terms— Pathology image, von Mises mixture model, circular data, hue histogram, color estimation.

1. INTRODUCTION

Color vision cue is an important information source in human visual system, as it helps people to sense and to segment objects. It is believed that objects in their diagnostic colors are recognized more readily [1]. Based on the functionality of color in human vision processing, many color analysis algorithms are developed for computer vision applications. Particularly in pathology image analysis, color cues attributed to chemical staining of biopsy samples provide useful and reliable information and are widely used to detect histological objects from background. For instance, diagnostic colors of stroma, lumen, epithelial nuclei, and epithelial cyoplasm in H&E stained images are widely used to label histological structures for prostate gland segmentation and classification [2, 3, 4]. It should be noted that though chemical dyes have their own typical colors, color variation, which refers to color difference for the same type of stains, is usually observed in pathology images due to stains' distinct manufacturers, storage conditions, and concentrations [5]. Consequently, color-based pathology images analysis may generate inaccurate results due to color variation in images. To take advantages of color cues in pathology images and meanwhile to address the color variation issue, blind estimation of object's diagnostic color is significant for subsequent quantitative analysis.

Hue is an efficient color feature, independent of intensity attribute [6]. Hence, in literature of color estimation for object's detection and segmentation, multilevel thresholding on hue histogram followed by representative color computation in resulting divided regions is widely used. Depending on thresholding algorithms, we categorize various histogram-based color estimation methods into non-parametric and parametric approaches. In nonparametric methods, hue histogram is divided into regions by thresholds determined by optimizing certain objective functions [7, 8, 9]. Particularly, Otsu's method [10] and its variants, which aim to minimize intra-class variance for thresholding separation, are usually used to cluster hue components so that each cluster corresponds to one object [9]. However, applying these linear thresholding methods on circular data, such as hue, may lead to inappropriate separation, finally resulting inaccurate color estimation. Later, circular histogram thresholding were achieved by exhaustive search on the hue circle using Otsu's criterion as an objective function [11, 12]. However, their computation complexity increases dramatically as thresholding level increases. Recently, efficient circular thresholding proposed in [13] reduces the computation complexity of binary circular thresholding from quadratically to linearly. By contrast, parametric approaches model hue distribution using various statistical cluster models, e.g. the Gaussian mixture distribution [14]. By computing parameters of the mixture model on the basis of hue distribution in histogram, image pixels are clustered into groups for color estimation [15, 16]. However, as classical statistical models, including the widely used Gaussian mixture model, describe distributions of linear variables, they are incapable to address the periodic nature of hue, which may introduce inaccurate, or even irrelevant color estimations. Hence, an efficient circular cluster model is needed when processing directional data, such as hue histogram.

In this paper, we address the issue of object's diagnostic color estimation in pathology image analysis and introduce

a novel blind color estimation method based on circular data clustering. The proposed method models image saturation-weighted hue histogram using a statistical distribution composed of multiple von Mises distributions, followed by pixel clustering and color information estimation. Particularly, unlike conventional parametric color estimation methods that apply linear mixture models to image hue components, the von Mises mixture model used in this paper addresses the periodic nature of hue and achieves optimal pixel clustering via maximum likelihood estimation. Experimentation suggests that our von Mises mixture model outperforms classical linear thresholding methods and the state-of-art circular thresholding algorithm in terms of circular data clustering. Diagnostic color of histological components estimated from pathology images demonstrates the effectiveness of the proposed method.

The rest of this paper is organized as follows. Section 2 presents the proposed diagnostic color estimation method in detail. Experimental results and discussion are given in Section 3, followed by conclusion in Section 4.

2. PROPOSED COLOR ESTIMATION SCHEME

The block diagram of the proposed color estimation method is depicted in Fig.1. Given a query color image, its saturation-weighted hue histogram in the HSV color space is modeled via a von Mises mixture distribution, and image pixels are clustered into groups, each corresponding to one type of cytological objects. Then complete diagnostic color in each cluster is estimated from the query image. The inputs of the proposed estimation method are a color pathology image and the number of stained tissue components contained in the image. The outputs of our scheme are diagnostic colors for different cytological objects. In this work, we assume that a query image is imaged under a standard illuminant, such as CIE D_{65} [18], so that image background appears white.

It is noteworthy that though this work focuses on diagnostic color estimation for histological objects in pathology images, the proposed method is applicable to color images in other computer vision applications.

2.1. Saturation-Weighted Hue Histogram

As tissue samples are almost transparent, even though they are stained by chemical dyes, their corresponding images may still contain a number of achromatic pixels. For instance, the pathology image in Fig. 1 has a large area of white background, which is composed of achromatic pixels. Since achromatic color characterized by a small saturation component in the HSV color space is less meaningful, or even meaningless, for color description, impacts of low-saturated pixels on object's color estimation should be avoided. To this end, a color image is converted to the HSV model and color estimation is performed on saturated image pixels only. In

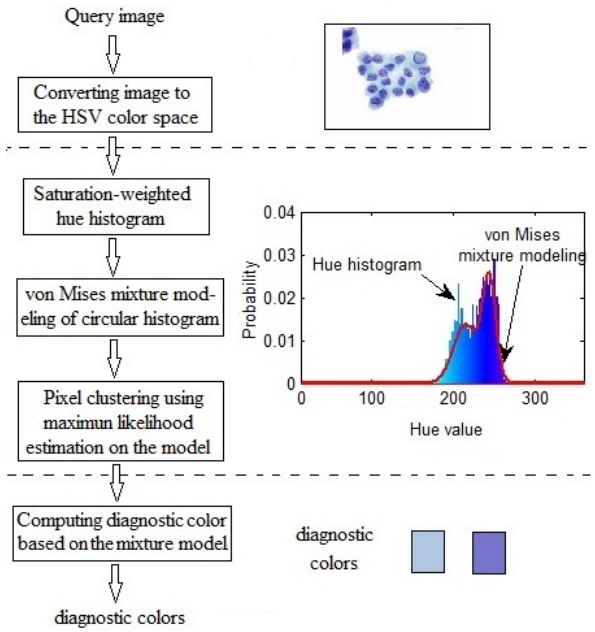

Fig. 1. Block diagram of the proposed object's color estimation method. The color cytopathology image is from the thyroid image atlas [17].

specific, to select reliable pixels for object's color estimation, we follow the work in [19] and use saturation-weighted hue histogram [20] defined as follows:

$$H_\theta^{sw} = \sum_p s(p)\delta(\theta, h(p)), \tag{1}$$

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

$\theta \in [0°, ..., 360°)$ represents a bin in histogram, $s(p)$ and $h(p)$ are the saturation and hue at pixel p in the HSV color space. In Eqn. (1), saturation components are used as filter parameters to compute histogram. Consequently, less-saturated pixels have small contributions to the hue histogram.

2.2. Von Mises Mixture Modeling of Hue Histogram

In linear processing, values of 0 and 360 are separated apart and their average is 180. By contrast, hue is an angular measurement of color in the chromatic plane, and $h(p) = 360$ is equivalent to $h(p) = 0$. Hence, the average hue value of 0 and 360 on the chromatic plane is still 0 (or 360). The different mean values of linear and circular data suggest that applying linear processing to directional data is inappropriate. To avoid analysis errors, this work uses a circular distribution, rather than linear statistical models, to model hue histogram.

As demonstrated in Fig. 1, the resulting hue histogram is composed of multiple hue clusters, each corresponding to one type of stained tissue components in the query image. Without any prior knowledge on color distribution, we propose

the use of a circular unimodel, called von Mises distribution $vM(\mu, \kappa)$ [21], to model each cluster of hue in the resulting saturation-weighted histogram. Specifically, the probability density function (pdf) of $vM(\mu, \kappa)$ is defined as follows:

$$vM(\mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa cos(x-\mu)}, \quad (2)$$

where x is a circular random variable, $0 \leq \mu < 2\pi$ and $\kappa \geq 0$ represent its mean direction and concentration, and $I_0(\kappa)$ is the zero-order modified Bessel function. When $\kappa = 0$, $vM(\mu, \kappa)$ converges to the uniform distribution on the unit circle, and when $\kappa \to \infty$, $vM(\mu, \kappa)$ tends to an impulse at the direction μ. The reason for selecting the von Mises distribution is because $vM(\mu, \kappa)$ is the 'nature' analogue on the circle of the Gaussian distribution in linear statistics [21].

For the entire hue histogram that contains N clusters of hue, we build a von Mises mixture model whose pdf is

$$g(x) = \sum_{i=1}^{N} \pi_i vM(\mu_i, \kappa_i), s.t. \sum_{i=1}^{N} \pi_i = 1. \quad (3)$$

Based on the mixture model and the specific hue histogram H_θ^{sw}, maximum likelihood estimation (MLE) of unknown parameters $\Psi = \{\mu_i, \kappa_i, \pi_i\}$ for $i = 1, ..., N$ is performed. We formulate this MLE problem in Eqn. (4).

$$\Psi^{mle} = \arg\max_{\Psi} g(H_\theta^{sw}|\Psi), s.t. \sum_{i=1}^{N} \pi_i = 1. \quad (4)$$

Due to the complexity of the mixture model in the MLE problem, solving Eqn. (4) directly by the Lagrangian method is difficult. Hence, we apply the expectation-maximization (EM) algorithm [22] to obtain Ψ^{mle} iteratively. An example of the proposed circular mixture model with MLE parameters is illustrated in the diagram scheme in Fig.1. It should be noted that when approaching a solution to Eqn. (4), the EM algorithm iteratively updates pixel clustering information, which can be used for subsequent quantitative analysis.

2.3. Diagnostic Color Computation

In the optimal solution to the MLE problem expressed in Eqn. (4), we obtain mean direction $h_i = \mu_i^{mle}$ for the i^{th} cluster in the hue histogram. Then we proceed to calculate complete diagnostic color information for each cluster, which includes representative value and saturation components in the HSV space. Specifically, for each cluster of hue that corresponds to one stained cytological component, the two components of a diagnostic color are estimated using following formulas [23] to mitigate effects of achromatic pixels on the estimation:

$$v_i = [\sum_p v(p)\delta(h_i, h(p))]/[\sum_p \delta(h_i, h(p))], \quad (5)$$

$$s_i = [\sum_p s(p)\delta(h_i, h(p))]/[\sum_p \delta(h_i, h(p))], \quad (6)$$

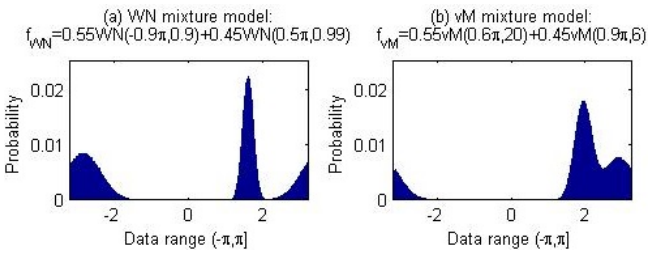

Fig. 3. Pdf of the two statistical models in the first experiment. (a) wrapped Normal mixture distribution; (b) von Mises mixture distribution.

where $\delta(x, y)$ is an indicator function, i.e. $\delta(x, y) = 1$ when $x = y$, otherwise $\delta(x, y) = 0$. Finally, complete diagnostic color of a stained histological object in a query image is described by a color vector $[h_i, s_i, v_i]$ in the HSV model.

3. EXPERIMENTAL RESULTS

In this work, two experiments are performed to evaluate the proposed method. In the first experiment, the proposed von Mises mixture model is assessed and compared to Otsu's method [10], Gaussian mixture model [14], and the state-of-art circular thresholding [13] on synthetic circular data. In the second experiment, the proposed color estimation approach is applied to publicly-accessible pathology images which are stained by different chemical dyes.

3.1. Effectiveness of Von Mises Model On Circular Data

This experiment quantitatively evaluates effectiveness of the von Mises mixture model on circular data generated following different mixture statistical distributions in the form of $\pi f(\Theta_1) + (1 - \pi) f(\Theta_2)$, where $f(\Theta)$ is a certain circular unimodel distribution with a parameter set Θ.

Testing Data: Two mixture models are used to generate synthetic data in this experiment. The first distribution is defined by pdf of $f_{WN} = 0.55WN(-0.9\pi, 0.9) + 0.45WN(0.5\pi, 0.99)$, where $WN(\Theta = \{\mu, \kappa\})$ denotes the wrapped Normal distribution [21]. The other distribution for data generation has pdf of $f_{vM} = 0.55vM(0.6\pi, 20) + 0.45vM(0.9\pi, 6)$. We depict pdf of the two mixture models in Fig. 3. Following each distribution, 20 sets of data are generated, each set containing 10000 data points.

Experimental Design: In this experiment, Otsu's method [10], Gaussian mixture model [14], binary circular thresholding [13], and the proposed von Mises mixture model are used to estimate directional mean for each set of circular data. As 20 sets of data are generated following one distribution, statistics of the 20 mean values are summarized and compared to the true means used for data generation.

Results and Discussion: Statistics (mean and standard deviation) of circular mean estimated from synthetic data are

Table 1. Statistics of mean estimation on synthetic circular data using different thresholding methods

Statistics models		Linear processing		Circular processing	
Name	Means	Otsu's method [10]	Gaussian mixture [14]	Circular thresholding [13]	von Mises mixture
f_{WN}	-0.9π	$-0.836\pi \pm 0.001$	$-0.758\pi \pm 0.245\pi$	$-0.889\pi \pm 0.002\pi$	$-0.901\pi \pm 0.002\pi$
	0.5π	$0.600\pi \pm 0.002\pi$	$0.538\pi \pm 0.173\pi$	$0.509\pi \pm 0.002\pi$	$0.500\pi \pm 0.001\pi$
f_{vM}	0.6π	$0.698\pi \pm 0.001\pi$	$0.696\pi \pm 0.001\pi$	$0.614\pi \pm 0.001\pi$	$0.600\pi \pm 0.002\pi$
	0.9π	$1.083\pi \pm 0.002$	$1.080\pi \pm 0.003\pi$	$0.945\pi \pm 0.002\pi$	$0.900\pi \pm 0.004\pi$

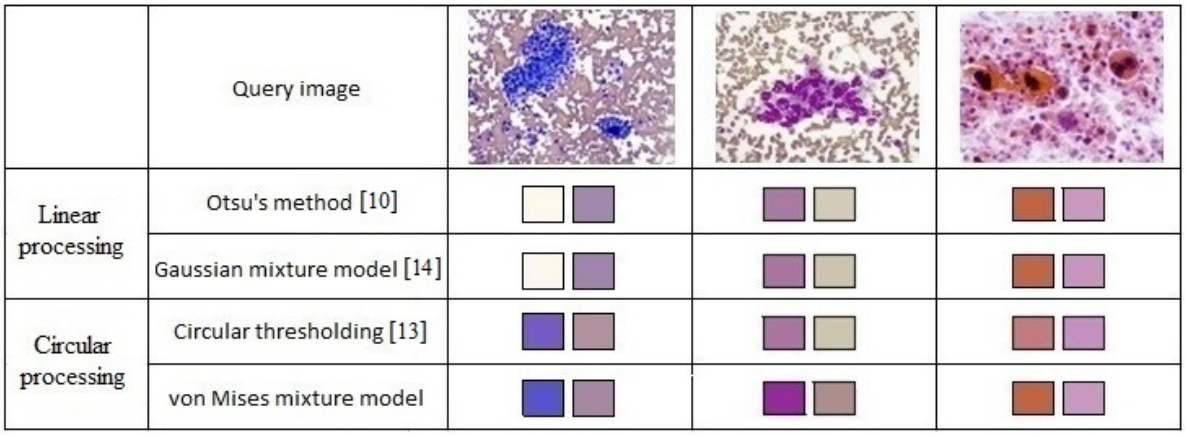

Fig. 2. Examples of color estimation using different thresholding methods for pathology images.

presented in Table 1. We compare the results to the true means used for data generation and obtain three observations. First, estimations generated by linear processing (Otsu's method and Guassian mixture model) are poor, because periodicity of circular data is ignored. Second, binary circular thresholding proposed in [13], which is an extension of Otsu's method for directional data, may generate non-optimal thresholding results. Third, mean estimations by the von Mises mixture model are accurate and stable, even though the synthetic data follows non-von Mises distributions.

3.2. Qualitative Evaluation of Color Estimation

Testing Data: Cytopathology images of thyroid lesions in the thyroid image atlas [17] published by Papanicolaou Society of Cytopathology are selected for the following reason. Unlike conventional pathology image sets consisting of images stained by the same types of chemical dyes, lesion images in this thyroid dataset are stained following different staining protocols, which makes our evaluation reliable and solid.

Experimental Design: Images in the thyroid image atlas are used as input to the proposed color estimation method. For comparison, Otsu's method [10], Gaussian mixture model [14], state-of-art circular thresholding [13] are also performed on resulting hue histograms, replacing the von Mises mixture model to generate diagnostic hue for histological objects.

Results and Discussion: Three examples of color estimation are presented in Fig. 2. As tissue samples are stained by different chemical dyes, color in the three images has distinct distributions. As shown in the figure, linear processing approaches including both Otsu's method and the Gaussian mixture modeling may fail to obtain diagnostic color, as linear processing cannot address periodic nature of hue. Compared to the query images, diagnostic colors generated by circular thresholding [13] are a little faded due to its non-optimal thresholding. As the proposed method is capable to estimate object's diagnostic color accurately, it can be used for subsequent pathology image analysis, such as tissue component detection and segmentation.

4. CONCLUSION

This paper introduced a blind object's color estimation method for pathology images, which particularly addresses the circular nature of the chromatic component in analysis. The introduced approach was able to select reliable pixels from a query image for subsequent estimation, and obtained accurate color information based on clustering results achieved by maximum likelihood parameter estimation. Especially, to accurately model hue distribution of a color image, which is a key in our method, we made use of a circular mixture model that was composed of multiple von Mises distributions. Experimental results demonstrate that our method is effective in diagnostic color estimation for histological objects in pathology images.

5. REFERENCES

[1] J. Tanaka, D. Weiskopf, and P. Williams, "The role of color in high-level vision," *Trends in cognitive sciences*, vol. 5, no. 5, pp. 211 – 215, May. 2001.

[2] S.K. Tai, Y.C. Wu, C.Y. Li, Y.J. Jan, and S.C. Liu, "Computer-assisted detection and grading of prostatics cancer in biopsy image," in *International Multi-Conference of Engineers and Computer Scientists*, 2010.

[3] K. Nguyen, A.K. Jain, and R.L. Allen, "Automated gland segmentation and classification for gleason grading of prostate tissue images," in *Proc. IEEE Int. Conf. Pattern Recogn.*, 2010.

[4] K. Nguyen, A. Sarkar, and A.K. Jain, "Structure and context in prostatic gland segmentation and classification," in *Proc. Med. Image Comput. Comput.-Assisted Intervention*, 2012.

[5] 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.

[6] D-C. Tseng and C-H. Chang, "Color segmentation using perceptual attributes," in *Proc. IEEE Int. Conf. Pattern Recogn., Vol. III, Conference C*, 1992.

[7] H-D. Cheng and Y. Sun, "A hierarchical approach to color image segmentation using homogeneity," *IEEE Trans. Image Process.*, vol. 9, no. 12, pp. 2071 – 2082, Dec.. 2000.

[8] S. Sural, G. Qian, and S. Pramanik, "Segmentation and histogram generation using the hsv color space for image retrieval," in *Proc. IEEE Int. Conf. Image Process.*, 2002.

[9] J. Wu, P. Zeng, Y. Zhou, and C. Olivier, "A novel color image segmentation method and its application to white blood cell image analysis," in *Proc. IEEE Int. conf. Signal process.*, 2006.

[10] 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.

[11] D-C Tseng, Y-F Li, and C-T Tung, "Circular histogram thresholding for color image segmentation," in *Proc. IEEE Int. Conf. Document Analysis and Recognition*, 1995.

[12] D. Dimov and L. Laskov, "Cyclic histogram thresholding and multithresholding," in *Proc. ACM Int. conf. Computer systems and technologies*, 2009.

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

[14] T. Kurita, N. Otsu, and N. Abdelmalek, "Maximum likelihood thresholding based on population mixture models," *Pattern Recognition*, vol. 25, no. 10, pp. 1231 – 1240, 1992.

[15] C. Bouvier, P-Y. Coulon, and X. Maldague, "Unsupervised lips segmentation based on roi optimisation and parametric model," in *Proc. IEEE Int. Conf. Image Process.*, 2007.

[16] N. Henderson, R. King, and R.H. Middleton, "An application of gaussian mixtures:colour segmenting for the four legged league using hsi colour space," in *Proc. RoboCup 2007: Robot Soccer World Cup XI*, 2008.

[17] Papanicolaou Society of Cytopathology, "Thyroid image atlas," `http://www.papsociety.org/atlas.html`, 2004, [Online:access 5-Jan.-2015].

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

[19] X. L and K.N. Plataniotis, "Blind stain decomposition for histo-pathology images using circular nature of chroma components," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2015.

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

[21] N.I. Fisher, *Statistical analysis of circular data*, Cambridge University Press, 1996.

[22] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. Royal Statist. Soc. Series B*, vol. 39, pp. 1 – 38, 1977.

[23] X. L and K.N. Plataniotis, "A complete color normalization approach to histo-pathology images using color cues computed from saturation-weighted statistics," *IEEE Trans. Biomed. Eng.(in press)*, 2015.