Rough-Fuzzy Circular Clustering for Color Normalization of Histological Images

Pradipta Maji* **and Suman Mahapatra**
Biomedical Imaging and Bioinformatics Lab
Machine Intelligence Unit, Indian Statistical Institute
203 B. T. Road, Kolkata, West Bengal, 700 108, India
{*pmaji,sumancse_r*}*@isical.ac.in*

Abstract. Color disagreement among histological images may affect the performance of computer-aided histological image analysis. So, one of the most important and challenging tasks in histological image analysis is to diminish the color variation among the images, maintaining the histological information contained in them. In this regard, the paper proposes a new circular clustering algorithm, termed as rough-fuzzy circular clustering. It integrates judiciously the merits of rough-fuzzy clustering and cosine distance. The rough-fuzzy circular clustering addresses the uncertainty due to vagueness and incompleteness in stain class definition, as well as overlapping nature of multiple contrasting histochemical stains. The proposed circular clustering algorithm incorporates saturation-weighted hue histogram, which considers both saturation and hue information of the given histological image. The efficacy of the proposed method, along with a comparison with other state-of-the-art methods, is demonstrated on publicly available hematoxylin and eosin stained fifty-eight benchmark histological images.

Keywords: Histological image analysis, color normalization, circular clustering, rough sets.

1. Introduction

Histology is defined as the microscopic study of tissue sections to analyze the manifestation of diseases under consideration. The most important property of histological data is the presence of enormous amount of information, compared to radiological and other imaging modalities. As a result, automated diagnosis based on histological data produces more accurate results than other imaging modalities. To

*Address for correspondence: Machine Intelligence Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata, 700 108, India.

ease pathologists' analysis, tissue samples collected from different biopsy sets are stained with several histochemical reagents, which in turn highlight different cellular structures. Staining not only highlights different cellular components, but also provides useful information in proper diagnosis of several pathological conditions [1]. Hence, in histological image analysis, color plays an important role in identification of different histological components.

The most important problem in histological image analysis is the unacceptable variation in stained tissue color. The color representation of the stained histological images may vary due to various factors, such as, manual tissue specimen preparation, orientation of lens aperture, variation in staining routine, different quality stains from different manufacturer, etc. As a result, numerical features, extracted from these histological images with an observable variation in color representation, may be inaccurate and may lead to difficulty in image interpretation by computer-aided systems [2, 3]. Hence, deviation in color representation affects proper detection, trustworthy diagnosis and gradation of the diseases under consideration. The color disagreement creates difficulty in image interpretation by software trained on a particular color appearance.

In digital histology, normalization algorithms play a pivotal role to reduce the color disagreement among the images. As the color normalization methods try to reduce color deviation among the images, similar cellular structures in different images should exhibit similar color representation. Thus, the resulting images facilitate subsequent image segmentation and feature extraction, eventually improving the accuracy of the computer-aided diagnosis system. The most popular approach in this domain is the histogram specification approach, which maps color representation of a source/query image to that of a pre-specified template image [4]. In another variation of histogram specification [5], color map quantile matching is used for color normalization. As these methods do not investigate the actual reasons associated with color variation problem, histological information is not expected to be kept intact after color normalization. Another group of color normalization approaches follows the color transfer technique, where each image is converted from RGB color space to L$\alpha\beta$ color space, as referred in [6], and corresponding to each image channel, mean and standard deviation are computed. Based on the source and template image channel statistics, source image is standardized [7].

Color normalization approach based on stain separation considers local information of the image content and tries to deal with the actual reasons associated with color variation among the images. To normalize the color of histological images, a non-linear Bezier curve-fitting based mapping approach is used in [2]. In this method, global stain color descriptor and local pixel-level information are fed to a relevance vector machine for classifying each pixel to a particular stain class. In [8], each RGB pixel is first transformed to corresponding optical density (OD) values and the OD values are projected onto a plane, spanned by the eigenvectors corresponding to two largest eigenvalues computed from singular value decomposition of the OD pixels. In [9], the histological images within the set are subjected to an illumination normalization module to get rid of color variation due to illumination inconsistency. After converting the RGB image to corresponding HSI counterpart, saturation-weighted hue histogram is computed, and finally k-means clustering is performed to accomplish the pixel clustering task. A sparse non-negative matrix factorization (NMF) based blind stain estimation procedure is proposed in [10], where the imposed sparsity on the NMF objective function ensures the uniqueness of the solution.

In general, the stain classes are overlapping with each other. This overlapping nature of multiple contrasting histochemical stains introduces the uncertainty in stain class definition. Most of the aforementioned methods do not consider the overlapping characteristics of histochemical reagents. In [11], a circular mixture modeling based blind stain decomposition method is recently proposed, which considers

the physical constraints such as non-negativity of the decomposed matrices as well as overlapping property of the histochemical stains. However, this method does not address the uncertainty due to vagueness and incompleteness in stain class definition.

In this regard, a new circular clustering algorithm, termed as rough-fuzzy circular clustering, is proposed for color normalization of histological images. The proposed method incorporates the merits of both rough sets and fuzzy sets. While rough sets deal with the uncertainty, vagueness and incompleteness in stain class definition, fuzzy sets address the overlapping nature of histochemical reagents. The proposed method also integrates saturation-weighted hue histogram to reduce the impact of less-saturated achromatic pixels on stain color estimation. The efficiency of the proposed method, along with a comparison with other state-of-the-art methods, is demonstrated on fifty-eight standard histological images.

2. Proposed Method

The stain color appearance map and stain density map, corresponding to each histology image, can be estimated using a stain decomposition method. In the proposed method, the stain decomposition is performed following the transmission light microscopy based image model [9]. Since the stain vectors intersect among each other at the origin in the OD domain, at first each image is mapped to OD space and then decomposed to extract the stain color appearance map and corresponding stain density map. The stain color appearance map or the spectrum matrix is estimated using a circular clustering, referred to as rough-fuzzy circular clustering, which follows next.

2.1. Rough-Fuzzy Circular Clustering

In the proposed rough-fuzzy circular clustering algorithm, each stain class is designated by a cluster prototype, a crisp lower approximation region, and a fuzzy boundary region. Let us suppose that the lower approximation and boundary region, corresponding to the i-th stain class β_i, are represented by $\underline{A}(\beta_i)$ and $B(\beta_i)$, respectively. The upper approximation $\overline{A}(\beta_i)$ of stain class β_i is then defined by $\overline{A}(\beta_i) = \{\underline{A}(\beta_i) \cup B(\beta_i)\}$.

According to the notions of lower approximation and boundary region of rough sets, if the j-th hue x_j lies in the lower approximation $\underline{A}(\beta_i)$ of the i-th stain class β_i, then $x_j \notin \underline{A}(\beta_k), \forall k \neq i$ and also $x_j \notin B(\beta_i), \forall i$. That is, hue x_j certainly belongs to the stain class β_i. Hence, the weights of the hue values, belonging to the lower approximation region of a particular stain class, should be independent of other centroids and stain classes, and should not be coupled with their similarity with respect to other centroids. Also, the hue values lying in lower approximation of a particular stain class should have identical influence on respective centroid and stain class. On the other hand, if a hue value x_j belongs to the boundary region $B(\beta_i)$ corresponding to the i-th stain class β_i, then x_j possibly lies in β_i and potentially belongs to another stain class. Hence, the hue values lying in boundary regions should have different influence on the centroids and stain clusters.

To estimate the stain class representative vectors, the proposed rough-fuzzy circular clustering algorithm uses a saturation-weighted (SW) hue histogram H^{sw}, instead of standard hue histogram. It has been shown in [12] that, the standard hue histogram exhibits a number of sharp ridges, attributed by achromatic pixels. Thus, the hue values corresponding to the achromatic region are ill-defined and insignificant for the pixels possessing low saturation values. As a result, color distribution analysis, based on standard hue histogram, produces insignificant and erroneous results. The effects of the achromatic

pixels on the standard hue histogram can be reduced by computing saturation-weighted hue histogram H^{sw} as follows:

$$H^{sw}(x) = \sum_{k \in I} s_k \delta(x, h_k); \text{ where } \delta(x, h_k) = \begin{cases} 1 & \text{if } x = h_k \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Here, I denotes a digital histological image. The hue and saturation values, corresponding to the k-th pixel in the HSI color space, are represented by h_k and s_k, respectively. The proposed rough-fuzzy circular clustering algorithm clusters the saturation-weighted hue histogram H^{sw}, corresponding to a histological image I, into c stain classes (an achromatic region and $(c-1)$ stain regions) by minimizing the following objective function, defined over a parameter set ψ:

$$J(\psi) = \sum_{i=1}^{c} \left[\omega \times J_i^L(\psi) + (1-\omega) \times J_i^B(\psi) \right]; \quad (2)$$

where the parameters ω and $(1-\omega)$ control relative importance between lower approximation and boundary region corresponding to each stain class, and

$$J_i^L(\psi) = \sum_{x_j \in \underline{A}(\beta_i)} [1 - \cos(x_j - v_i)] H^{sw}(x_j); \quad (3)$$

$$J_i^B(\psi) = \sum_{x_j \in B(\beta_i)} \mu_{ij}^m [1 - \cos(x_j - v_i)] H^{sw}(x_j). \quad (4)$$

Here, x_j represents the j-th hue and v_i denotes the i-th centroid corresponding to the i-th stain class β_i, $m \in [1, \infty)$ represents the fuzzifier and $\mu_{ij} \in [0, 1]$, represents the fuzzy membership of the j-th hue into i-th stain class β_i. Hence, the parameter set ψ is given by $\psi = \{v_i, \mu_{ij}\}$.

Combining (2) and (4), the objective function $J^B(\psi)$, corresponding to the boundary regions, can be written as follows:

$$J^B(\psi) = \sum_{i=1}^{c} \sum_{x_j \in B(\beta_i)} \mu_{ij}^m [1 - \cos(x_j - v_i)] H^{sw}(x_j). \quad (5)$$

Optimization of $J^B(\psi)$ with respect to μ_{ij}, subject to the constraint $\sum_{i=1}^{c} \mu_{ij} = 1, \forall j$, is transformed to the following Lagrangian function:

$$\mathcal{J}^B(\psi) = \sum_{i=1}^{c} \sum_{x_j \in B(\beta_i)} \mu_{ij}^m \left[1 - \cos(x_j - v_i)\right] H^{sw}(x_j) + \lambda \sum_{x_j \in B(\beta_i)} \left[1 - \sum_{i=1}^{c} \mu_{ij} \right] H^{sw}(x_j); \quad (6)$$

where λ denotes Lagrange's multiplier. Differentiating $\mathcal{J}^B(\psi)$ partially with respect to μ_{ij}, we get

$$\frac{\partial \mathcal{J}^B(\psi)}{\partial \mu_{ij}} = 0;$$

$$\Rightarrow m\mu_{ij}^{m-1}d_{ij}H^{sw}(x_j) - \lambda H^{sw}(x_j) = 0 \Rightarrow \mu_{ij} = \left(\frac{\lambda}{md_{ij}}\right)^{\frac{1}{m-1}}. \quad (7)$$

Here, $d_{ij} = [1 - \cos(x_j - v_i)]$ represents the cosine distance. So,

$$\sum_{i=1}^{c}\mu_{ij} = 1 \Rightarrow \left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}} = \left(\sum_{k=1}^{c}\left(\frac{1}{d_{kj}}\right)^{\frac{1}{m-1}}\right)^{-1}. \quad (8)$$

Substituting (8) in (7), the membership function μ_{ij} can be estimated as follows:

$$\mu_{ij} = \left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}}\left(\frac{1}{d_{ij}}\right)^{\frac{1}{m-1}}; \quad \Rightarrow \mu_{ij} = \left(\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{1}{m-1}}\right)^{-1}. \quad (9)$$

It is clearly observed from (9) that the computation of membership function μ_{ij} is dependent on the fuzzifier value m.

The computation of cluster prototype depends on the weighted average of crisp lower approximation and fuzzy boundary region corresponding to each stain class. The effects of both fuzzy membership, and lower approximation and boundary regions are incorporated in the computation of the centroid. The centroid v_i, corresponding to the i-th stain class β_i, is computed by differentiating (2) partially with respect to v_i, as follows:

$$\frac{\partial J(\psi)}{\partial v_i} = 0;$$

$$\Rightarrow \omega \times \sum_{x_j \in \underline{A}(\beta_i)} \sin(x_j - v_i)H^{sw}(x_j) + (1-\omega) \times \sum_{x_j \in B(\beta_i)} \mu_{ij}^m \sin(x_j - v_i)H^{sw}(x_j) = 0;$$

$$\Rightarrow v_i = \arctan\left[\frac{\omega \times \mathcal{P}_1 + (1-\omega) \times \mathcal{P}_2}{\omega \times \mathcal{Q}_1 + (1-\omega) \times \mathcal{Q}_2}\right]; \quad (10)$$

$$\text{where } \mathcal{P}_1 = \sum_{x_j \in \underline{A}(\beta_i)} \sin(x_j)H^{sw}(x_j), \mathcal{P}_2 = \sum_{x_j \in B(\beta_i)} \mu_{ij}^m \sin(x_j)H^{sw}(x_j);$$

$$\mathcal{Q}_1 = \sum_{x_j \in \underline{A}(\beta_i)} \cos(x_j)H^{sw}(x_j), \text{and } \mathcal{Q}_2 = \sum_{x_j \in B(\beta_i)} \mu_{ij}^m \cos(x_j)H^{sw}(x_j).$$

It is evident from (10) that the computation of centroid depends on selection of the parameter ω. As the hue values corresponding to lower approximation region certainly belong to a specific stain class, they are supposed to be assigned a higher weightage ω compared to a lower weightage $(1-\omega)$, assigned to the hues belonging to the boundary region.

2.2. Proposed Algorithm and Color Normalization

The optimization of $J(\psi)$ using the proposed rough-fuzzy circular clustering algorithm is based on an alternating optimization method, referred as Picard iteration, through (9) and (10). The procedure starts by initializing the parameters $v_i^{(0)}$ and $\mu_{ij}^{(0)}$ corresponding to $(c-1)$ stain regions and an achromatic

region. Depending on the circular nature of hue, circular thresholding method of [13] is utilized to produce approximate partitions corresponding to the achromatic region and different stain regions. The initial cluster prototype $v_i^{(0)}$, corresponding to both stain and achromatic regions, can be computed from the saturation-weighted hue histogram H^{sw} using the initial parameter estimation method, proposed in [11]. The fuzzy memberships, associated to all hues, are then computed using (9). The basic steps of the proposed rough-fuzzy circular clustering algorithm are outlined as follows:

1. Initialize $v_i^{(0)}$ $(1 \leq i \leq c-1)$ for $(c-1)$ stain regions and $v_{ac}^{(0)}$ for the achromatic region.
2. Select values for fuzzifier m, and pre-specified thresholds ϵ and δ. Set iteration counter $\kappa = 1$.
3. Compute μ_{ij} for c classes and n hues using (9).
4. If μ_{ij} and μ_{kj} represent the highest and second highest memberships of x_j, respectively, and $(\mu_{ij} - \mu_{kj}) > \delta$, then $x_j \in \underline{A}(\beta_i)$; otherwise $x_j \in B(\beta_i)$ and $x_j \in B(\beta_k)$.
5. Update μ_{ij} considering the lower approximation and boundary regions for c classes and n hues.
6. Compute new centroid v_i as per (10).
7. Repeat steps 3 to 6, by incrementing κ, until $|\mu_{ij}^{(\kappa)} - \mu_{ij}^{(\kappa-1)}| < \epsilon$.
8. Stop.

The spectrum matrix can be computed from the cluster centroids representing each stain class. The cluster prototype, v_i, $i = 1, 2, \cdots, c$, estimated from the proposed circular clustering algorithm, is considered to be the representative hue h_i for the i-th stain class β_i. The saturation s_i and intensity value y_i, corresponding to the representative hue h_i, are computed using the saturation-weighted statistics, described in [9]. The HSI domain representative stain vector $I_i^{hsi} = [h_i, s_i, y_i]^t$, corresponding to the i-th stain class β_i, is first transformed to RGB color space: $I_i^{rgb} = G_{hsi}^{rgb}(I_i^{hsi})$, where the function $G_{hsi}^{rgb}(.)$ converts a HSI domain stain vector to corresponding RGB domain vector. Thus, the stain vectors corresponding to the c stain classes are utilized to form the representative spectrum matrix of dimension $3 \times c$. The RGB domain color appearance vector I_i^{rgb} is then converted to corresponding OD domain vector represented as: $W_i = \log(I^b / I_i^{rgb})$, where I^b denotes the intensity value corresponding to a background pixel. The stain density value corresponding to a pixel p can be evaluated using the following equation:

$$X(p) = W^{-1}[\log(I^b / I(p))]. \tag{11}$$

As the spectrum matrix and corresponding stain density map should satisfy the element-wise non-negativity constraint, the NMF is applied to refine the decomposition solution:

$$\{W^*, X^*\} = \min_{W \succeq 0, X \succeq 0} || \log(I^b / I) - W \times X ||^2 . \tag{12}$$

Here, $\succeq 0$ represents element-wise non-negativity corresponding to the factor matrices. Using the template image stain color appearance map W^*, the source image color is normalized maintaining the stain density information X^* associated with each pixel of the source image. The block diagram of the proposed rough-fuzzy circular clustering based color normalization method is depicted in Figure 1.

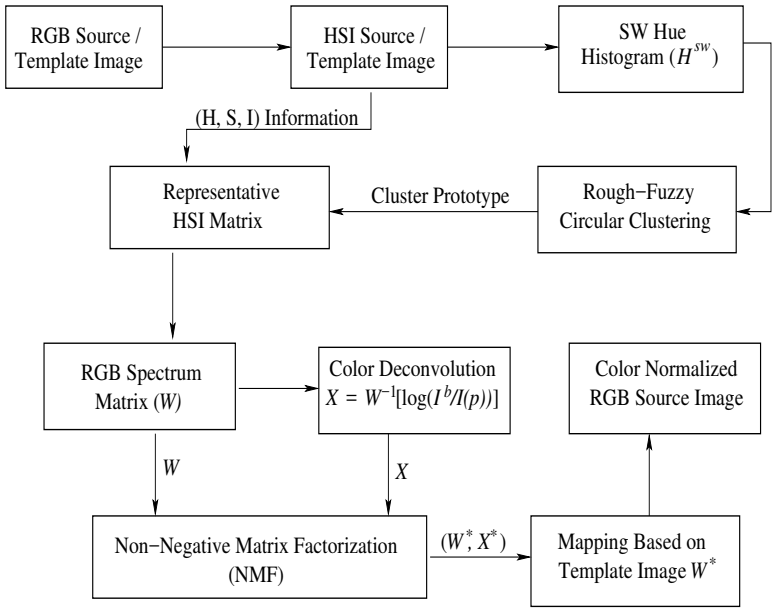

Figure 1. Block diagram of the proposed color normalization method

3. Experimental Results and Discussions

The performance of the proposed rough-fuzzy circular clustering based stain separation method is studied extensively and compared with that of different existing methods, such as plane fitting (PF) [8], HTN method as referred in [9], structure-preserving color normalization (SPCN) [10] and expectation-maximization (EM) based circular mixture modeling [11]. The performance of the proposed method in color normalization is also compared with several color normalization techniques, namely, color transfer (CT) technique [7], stain color description (SCD) [2] as well as PF [8], HTN [9], and SPCN [10]. In the proposed rough-fuzzy circular clustering method, experimentally it has been observed that the algorithm performs best for $\delta = 0.1$ and weight parameter $\omega = 0.55$.

3.1. Description of Data Set

To assess the performance of several stain decomposition and color normalization approaches, publicly available UCSB Breast Cancer Cell Data, published by University of California, Santa Barbara [14], is used. This histology image set consists of 58 images: 32 non-cancerous benign cell and 26 cancerous malignant cell images, collected from ten hematoxylin (H) and eosin (E) stained breast cancer biopsy sets. As UCSB data set comprises of images stored in 24 bit nonlinear RGB format, each image in this set is transformed to linear RGB space before analysis [15]. Each image has dimension 896×768.

3.2. Illustrative Example

Figure 2 presents an illustrative example of the proposed color normalization algorithm, considering ytma23_022103_benign2_ccd of biopsy set ytma23_022103 and ytma10_010704_benign1_ccd of biopsy set ytma10_010704 as template and query/source images, respectively. While Figure 2(a) represents the

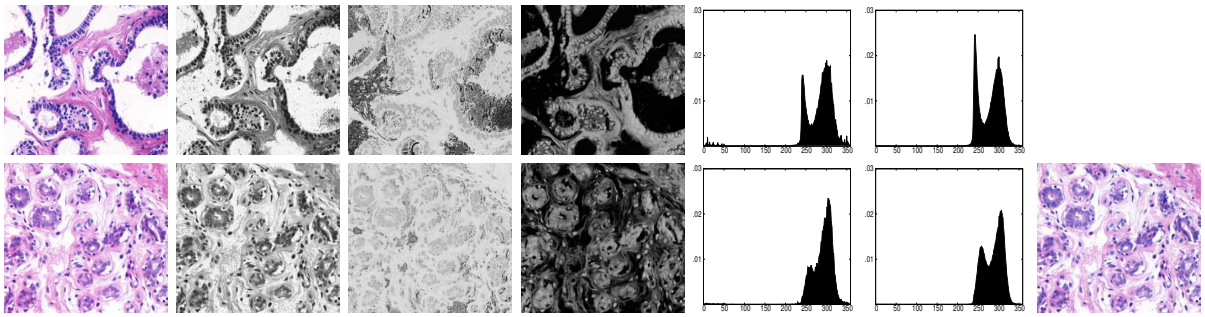

Figure 2. Example of color normalization (top: template; bottom: source): (a) original; (b) intensity; (c) hue; (d) saturation; (e) hue histogram; (f) saturation-weighted hue histogram; (g) color normalized source

original image of both template and source, Figures 2(b)-(d) depict the corresponding intensity, hue, and saturation, respectively, of both images. Figures 2(e) and (f) show the hue histograms and corresponding saturation-weighted hue histograms. Comparing Figures 2(e) and (f), it can be seen that the saturation-weighted hue histogram can eliminate the sharp ridges, which are mainly attributed by achromatic pixels and exhibited in standard hue histogram. Three cluster centroids obtained using the proposed rough-fuzzy circular clustering algorithm corresponding to the source image, considering saturation-weighted hue histogram, are $v_{ac} = 20.907$, $v_H = 262.089$, and $v_E = 303.857$. Figure 2(g) depicts the color normalized source image obtained using the proposed color normalization algorithm.

3.3. Relevance of Saturation-Weighted Hue Histogram

In order to ensure the usefulness of saturation-weighted hue histogram over the standard hue histogram, extensive analysis is performed on UCSB data set, and corresponding results are reported in Table 1. The UCSB data set consists of 10 biopsy sets: 9 of them contain 6 images each, whereas another contains only 4 images. It can be assumed that the images within the same biopsy set have undergone similar type of staining treatment. As a result, the images within the same biopsy set are expected to produce similar or consistent spectrum vectors. Based on this intuition, the standard deviation (σ) values of the estimated stain vectors of all the images, corresponding to the same biopsy set, are summarized in Table 1 (values are represented in 10^{-2} scale). As the spectrum vectors, corresponding to the same biopsy set, must exhibit low standard deviation, the smallest σ value, associated to respective matrix element, is marked in bold. It can be observed from Table 1 that the saturation-weighted hue histogram attains lowest σ values in 21 and 17 cases, out of 30 cases each for H-stain and E-stain, respectively, while the standard hue histogram achieves lowest σ in only 10 and 13 cases.

Moreover, the consistency of the color distribution, after color normalization, is evaluated with respect to normalized median intensity (NMI) [16] measure. The comparative performance analysis of two hue histograms is performed by means of tables of mean, median, standard deviation, and p-values computed using both paired-*t* (one-tailed) and Wilcoxon signed rank (one-tailed) tests. From the results reported in Table 2, it can be observed that the saturation-weighted hue histogram produces highest mean and median values for NMI and performs significantly better than the standard hue histogram, considering 0.05 as the level of significance.

Table 1. Performance of Standard and Saturation-Weighted Histograms on UCSB Data

Different Biopsy Sets	Different Channels	Standard Histogram		SW Histogram	
		σ_{H}	σ_{E}	σ_{H}	σ_{E}
ytma10_010704	Red	6.31	**2.68**	**2.39**	3.65
	Green	3.80	1.08	**1.39**	**1.03**
	Blue	2.22	**2.20**	**1.59**	2.55
ytma12_010804	Red	3.99	**5.26**	**1.89**	5.73
	Green	3.16	**1.63**	**1.17**	2.03
	Blue	1.25	1.46	**0.94**	**1.19**
ytma23_022103	Red	**1.61**	**5.12**	1.67	5.31
	Green	1.33	**1.53**	**1.24**	1.80
	Blue	0.69	1.49	**0.64**	**1.38**
ytma49_042003	Red	3.98	3.07	**3.28**	**2.36**
	Green	2.41	**0.45**	**1.99**	0.54
	Blue	1.02	1.51	**0.69**	**0.94**
ytma49_042203	Red	7.53	5.77	**2.24**	**3.07**
	Green	4.09	5.90	**1.49**	**0.87**
	Blue	1.65	12.86	**0.64**	**1.89**
ytma49_042403	Red	2.28	**3.29**	**1.72**	3.80
	Green	1.38	**0.91**	**0.94**	1.10
	Blue	**0.71**	1.28	0.76	**0.72**
ytma49_072303	Red	1.45	4.24	**1.41**	**2.86**
	Green	**1.14**	1.19	**1.14**	**1.09**
	Blue	**0.29**	1.08	0.30	**0.67**
ytma49_111003	Red	**1.89**	**3.24**	2.79	5.34
	Green	**1.51**	**1.31**	2.39	2.69
	Blue	**1.01**	**1.44**	1.08	1.93
ytma49_111303	Red	**1.70**	6.29	2.29	**6.09**
	Green	**1.43**	**2.44**	2.06	2.54
	Blue	**0.79**	1.82	0.80	**1.80**
ytma55_030603	Red	4.18	6.46	**1.91**	**3.84**
	Green	3.21	3.57	**1.52**	**1.62**
	Blue	0.80	10.41	**0.77**	**1.67**

Note: σ_{X} represents the standard deviation of the estimated stain vectors of X-stain, corresponding to all images of the same biopsy set; X $\in$ {H, E}

3.4. Significance of Rough Sets

To establish the significance of the proposed rough-fuzzy circular clustering algorithm over its hard and fuzzy counterparts, the results on UCSB data set are reported in Table 3. It can be seen from the results of Table 3 that the proposed rough-fuzzy circular clustering algorithm achieves lowest σ values in 15 and 18 cases for H-stain and E-stain, respectively, while hard and fuzzy clustering attain lowest σ in 10 and 14, and 13 and 3 cases, respectively. Moreover, the proposed circular clustering algorithm provides

Table 2. Comparison of Different Histograms and Clustering Algorithms Using NMI

Different Histograms /Clustering Methods	Normalized Median Intensity (NMI)				
	Mean	Median	StdDev	Wilcoxon Test	Paired-t Test
Standard Histogram	0.700852	0.697148	0.062673	9.29E-09	2.42E-06
Hard Clustering	0.701822	0.696991	0.061359	3.49E-09	1.22E-07
Fuzzy Clustering	0.702271	0.696991	0.060684	9.48E-09	2.18E-07
Proposed	**0.710847**	**0.707624**	**0.057697**		

significantly better NMI values, compared to both hard and fuzzy clustering as observed in Table 2. The proposed rough-fuzzy circular clustering algorithm performs significantly better than its hard and fuzzy counterparts as the incorporation of rough sets can efficiently handle uncertainty in stain class definition due to incompleteness and vagueness.

3.5. Performance of Different Algorithms

In order to establish the efficacy of the proposed rough-fuzzy circular clustering based stain estimation method, the performance of the proposed algorithm is compared with that of different state-of-the-art stain separation methods, namely, PF [8], HTN [9], SPCN [10], and EM [11] and the results are summarized in Table 4. It is observed from Table 4 that the proposed method achieves lowest standard deviation (σ) values in 11 and 12 cases, out of 30 cases each, corresponding to H-stain and E-stain, respectively. On the other hand, the existing PF, HTN, SPCN, EM exhibit lowest σ values in 1 and 7, 8 and 6, 3 and 4, and 7 and 1 cases, respectively, corresponding to H-stain and E-stain. The results reported in Table 4 ensure that the proposed algorithm outperforms the state-of-the-art blind stain separation and stain estimation methods.

The color consistency within the normalized images is next evaluated for the proposed method as well as other color normalization methods using the NMI index, and corresponding results are reported in Table 5. From the results reported in Table 5, it is evident that the proposed method produces highest mean and median values as compared to other color normalization methods. Moreover, the paired-t and Wilcoxon signed rank tests ensure that the proposed method performs significantly better than the existing color normalization approaches. The qualitative comparison of different color normalization methods is presented in Figure 3. The results reported in Table 5 and Figure 3 establish the fact that the proposed method outperforms state-of-the-art color normalization methods as per the color consistency after color normalization is concerned.

4. Conclusion

In histological image analysis, color normalization is one of the most important and foremost tasks, as the presence of color disagreement among the images may affect the performance of automated histological image analysis. In this regard, the contributions of this paper are mainly two-fold: (i) the development of a new circular clustering algorithm, designed for color normalization of histological images, which judiciously integrates the merits of rough sets and fuzzy set theory; and (ii) demonstrating the efficiency of the proposed method, along with a comparison with other state-of-the-art methods, on standard H&E

Table 3. Comparative Performance of Different Clustering Algorithms on UCSB Data

Different Biopsy Sets	Different Channels	Hard		Fuzzy		Proposed	
		σ_{H}	σ_{E}	σ_{H}	σ_{E}	σ_{H}	σ_{E}
ytma10_010704	Red	**2.37**	3.69	2.51	3.79	2.39	**3.65**
	Green	**1.25**	1.04	1.35	1.09	1.39	**1.03**
	Blue	1.26	2.52	**1.21**	**2.49**	1.59	2.55
ytma12_010804	Red	2.00	5.78	**1.89**	5.85	**1.89**	**5.73**
	Green	1.24	2.06	1.18	2.08	**1.17**	**2.03**
	Blue	0.95	1.20	**0.91**	1.21	0.94	**1.19**
ytma23_022103	Red	1.70	**5.30**	**1.67**	5.42	**1.67**	5.31
	Green	1.26	1.81	1.27	1.83	**1.24**	**1.80**
	Blue	**0.63**	**1.36**	0.66	1.39	0.64	1.38
ytma49_042003	Red	3.29	**2.36**	3.45	2.37	**3.28**	**2.36**
	Green	**1.99**	0.55	2.11	**0.54**	**1.99**	**0.54**
	Blue	**0.66**	**0.94**	0.94	0.95	0.69	**0.94**
ytma49_042203	Red	2.25	3.09	2.39	3.07	**2.24**	**3.07**
	Green	**1.44**	0.90	1.65	**0.87**	1.49	**0.87**
	Blue	0.72	**1.80**	**0.64**	1.90	**0.64**	1.89
ytma49_042403	Red	1.73	3.93	1.73	4.08	**1.72**	**3.80**
	Green	0.96	1.13	**0.94**	1.17	**0.94**	**1.10**
	Blue	0.76	0.77	**0.70**	0.87	0.76	**0.72**
ytma49_072303	Red	**1.41**	**2.83**	**1.41**	2.89	**1.41**	2.86
	Green	**1.13**	**1.08**	1.15	1.11	1.14	1.09
	Blue	0.31	0.69	0.32	0.73	**0.30**	**0.67**
ytma49_111003	Red	2.81	**5.26**	**2.67**	5.53	2.79	5.34
	Green	2.37	**2.64**	**2.32**	2.82	2.39	2.69
	Blue	1.13	**1.92**	1.10	1.97	**1.08**	1.93
ytma49_111303	Red	2.28	**5.96**	**2.25**	6.17	2.29	6.09
	Green	**2.04**	**2.50**	2.05	2.59	2.06	2.54
	Blue	**0.78**	**1.79**	0.84	1.82	0.80	1.80
ytma55_030603	Red	1.94	3.86	**1.85**	4.08	**1.91**	**3.84**
	Green	1.54	**1.62**	**1.51**	1.72	1.52	**1.62**
	Blue	0.78	1.71	0.78	1.73	**0.77**	**1.67**

Note: σ_{X} represents the standard deviation of the estimated stain vectors of X-stain, corresponding to all images of the same biopsy set; $\mathrm{X} \in \{\mathrm{H}, \mathrm{E}\}$

stained histological image set. The proposed method deals with the overlapping nature of stain classes as well as addresses the non-negativity constraint corresponding to stain color appearance map and corresponding stain density map. From the reported quantitative and qualitative results, it is clear that the proposed method outperforms existing stain separation as well as color normalization approaches, in terms of satisfying the physical constraints and maintaining the within-image color consistency after color normalization. Moreover, the proposed method does not need any prior information other than the information about the number of stains involved in the staining treatment of histological images under consideration.

Table 4. Comparative Performance Analysis of Different Algorithms on UCSB Data

Different Biopsy Sets	Different Channels	PF		HTN		SPCN		EM		Proposed	
		σ_{H}	σ_{E}	σ_{H}	σ_{E}	σ_{H}	σ_{E}	σ_{H}	σ_{E}	σ_{H}	σ_{E}
ytma10_010704	Red	3.04	**2.72**	2.87	3.07	2.42	3.93	2.45	4.10	**2.39**	3.65
	Green	2.23	2.01	1.77	0.72	1.66	0.73	1.61	**0.69**	**1.39**	1.03
	Blue	2.67	5.79	1.07	**2.34**	1.01	2.55	**0.92**	2.51	1.59	2.55
ytma12_010804	Red	1.99	**3.79**	**1.36**	5.77	2.20	6.26	2.08	7.73	1.89	5.73
	Green	2.16	2.82	**1.08**	**1.81**	1.80	2.14	1.66	2.54	1.17	2.03
	Blue	1.72	6.81	**0.52**	1.31	0.79	1.36	0.83	1.56	0.94	**1.19**
ytma23_022103	Red	1.77	**3.43**	2.65	5.41	**1.07**	5.75	1.18	7.27	1.67	5.31
	Green	1.95	**0.46**	2.49	1.61	**1.10**	1.81	1.16	2.21	1.24	1.80
	Blue	1.31	3.34	0.72	1.59	0.86	1.71	0.89	1.86	**0.64**	**1.38**
ytma49_042003	Red	**3.11**	2.11	3.23	**2.01**	3.14	2.98	4.56	4.53	3.28	2.36
	Green	2.98	2.02	2.42	**0.49**	2.44	0.74	3.40	0.96	**1.99**	0.54
	Blue	1.09	3.67	**0.51**	1.00	0.79	1.18	1.21	1.42	0.69	**0.94**
ytma49_042203	Red	2.29	**2.18**	4.72	3.09	1.99	3.36	**1.93**	3.88	2.24	3.07
	Green	2.95	1.40	4.52	0.85	2.02	**0.84**	2.06	1.18	**1.49**	0.87
	Blue	1.88	3.72	1.24	2.05	**0.60**	2.03	0.69	2.02	0.64	**1.89**
ytma49_042403	Red	3.08	**3.13**	3.64	4.00	2.99	4.89	3.39	6.66	**1.72**	3.80
	Green	3.16	3.18	2.75	**1.03**	2.38	1.38	2.72	1.67	**0.94**	1.10
	Blue	1.04	5.38	**0.58**	**0.43**	0.93	0.64	0.92	0.80	0.76	0.72
ytma49_072303	Red	2.09	2.63	2.57	2.44	1.99	**1.57**	2.52	1.99	**1.41**	2.86
	Green	2.48	1.22	2.55	0.97	2.15	**0.66**	2.67	0.95	**1.14**	1.09
	Blue	0.57	1.43	0.43	0.82	0.46	0.79	0.36	0.98	**0.30**	**0.67**
ytma49_111003	Red	1.90	13.17	1.77	7.03	4.00	9.65	**1.62**	8.12	2.79	**5.34**
	Green	2.60	11.29	1.78	3.59	1.69	5.79	**1.45**	4.63	2.39	**2.69**
	Blue	3.62	10.73	1.27	2.81	9.77	4.17	1.29	2.62	**1.08**	**1.93**
ytma49_111303	Red	2.15	6.12	**1.22**	6.45	1.44	6.85	1.63	8.87	2.29	**6.09**
	Green	2.52	**1.75**	**1.32**	2.39	1.56	2.85	1.75	3.27	2.06	2.54
	Blue	1.50	4.30	**0.22**	2.28	0.34	2.31	0.55	2.70	0.80	**1.80**
ytma55_030603	Red	2.18	7.85	2.30	16.90	1.35	5.22	**1.25**	7.60	1.91	**3.84**
	Green	3.39	3.90	2.24	2.28	1.48	2.36	**1.37**	3.19	1.52	**1.62**
	Blue	2.73	8.20	0.79	8.35	0.70	**1.62**	**0.67**	2.28	0.77	1.67

Note: σ_{X} represents the standard deviation of the estimated stain vectors of X-stain, corresponding to all images of the same biopsy set; X $\in$ {H, E}

Acknowledgement

This publication is an outcome of the R&D work undertaken in the project under the Visvesvaraya PhD Scheme of Ministry of Electronics and Information Technology, Government of India, being implemented by Digital India Corporation (formerly Media Lab Asia). The second author, S. Mahapatra, acknowledges financial support of Council of Scientific & Industrial Research, Government of India.

Table 5. Comparison of Different Color Normalization Algorithms Using NMI

Different Methods	Normalized Median Intensity (NMI)				
	Mean	Median	StdDev	Wilcoxon Test	Paired-*t* Test
CT	0.661974	0.680989	0.076786	4.83E-06	3.40E-06
PF	0.689482	0.689037	0.050312	8.24E-04	2.05E-04
SCD	0.602906	0.586130	0.109718	2.47E-10	3.81E-14
HTN	0.706873	0.701627	0.057786	2.58E-04	1.65E-03
SPCN	0.680182	0.679416	**0.044284**	4.49E-06	5.92E-07
Proposed	**0.710847**	**0.707624**	0.057697		

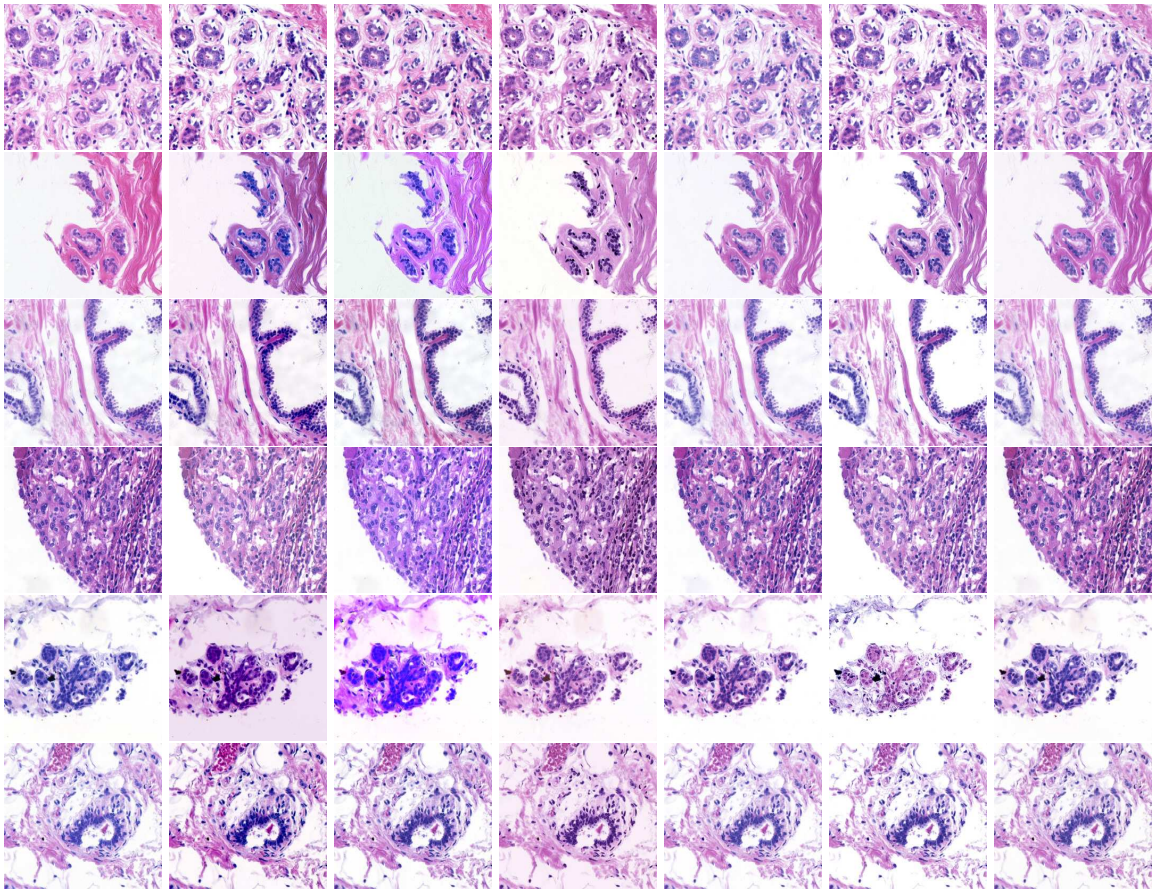

Figure 3. Original and color normalized images of UCSB data set obtained using different methods : (a) Original; (b) CT; (c) PF; (d) SCD; (e) HTN; (f) SPCN; (g) Proposed

References

[1] Ghaznavi F, Evans A, Madabhushi A, Feldman M. Digital Imaging in Pathology: Whole-Slide Imaging and Beyond. Annual Review of Pathology: Mechanisms of Disease. 2013;8(1):331–359. doi:10.1146/annurev-

pathol-011811-120902.

[2] Khan AM, Rajpoot N, Treanor D, Magee D. A Nonlinear Mapping Approach to Stain Normalization in Digital Histopathology Images Using Image-Specific Color Deconvolution. IEEE Transactions on Biomedical Engineering. 2014;61(6):1729–1738. doi:10.1109/TBME.2014.2303294.

[3] Peter L, Mateus D, Chatelain P, Schworm N, Stangl S, Multhoff G, et al. Leveraging Random Forests for Interactive Exploration of Large Histological Images. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention; 2014. p. 1–8. doi:10.1007/978-3-319-10404-1_1.

[4] Tabesh A, Teverovskiy M, Pang HY, Kumar VP, Verbel D, Kotsianti A, et al. Multifeature Prostate Cancer Diagnosis and Gleason Grading of Histological Images. IEEE Transactions on Medical Imaging. 2007;26(10):1366–1378. doi:10.1109/TMI.2007.898536.

[5] Kothari S, Phan JH, Moffitt RA, Stokes TH, Hassberger SE, Chaudry Q, et al. Automatic Batch-invariant Color Segmentation of Histological Cancer Images. In: Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 2011. p. 657–660. doi:10.1109/ISBI.2011.5872492.

[6] Ruderman DL, Cronin TW, Chiao CC. Statistics of Cone Responses to Natural Images: Implications for Visual Coding. Journal of the Optical Society of America A. 1998;15(8):2036–2045. doi:10.1364/JOSAA.15.002036.

[7] Reinhard E, Adhikhmin M, Gooch B, Shirley P. Color Transfer Between Images. IEEE Computer Graphics and Applications. 2001;21(5):34–41. doi:10.1109/38.946629.

[8] Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, et al. A Method for Normalizing Histology Slides for Quantitative Analysis. In: Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 2009. p. 1107–1110. doi:10.1109/ISBI.2009.5193250.

[9] Li X, Plataniotis KN. A Complete Color Normalization Approach to Histopathology Images Using Color Cues Computed from Saturation-Weighted Statistics. IEEE Transactions on Biomedical Engineering. 2015;62(7):1862–1873. doi:10.1109/TBME.2015.2405791.

[10] Vahadane A, Peng T, Sethi A, Albarqouni S, Wang L, Baust M, et al. Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. IEEE Transactions on Medical Imaging. 2016;35(8):1962–1971. doi:10.1109/TMI.2016.2529665.

[11] Li X, Plataniotis K. Circular Mixture Modeling of Color Distribution for Blind Stain Separation in Pathology Images. IEEE Journal of Biomedical and Health Informatics. 2017;21(1):150–161. doi:10.1109/JBHI.2015.2503720.

[12] Hanbury A. Circular Statistics Applied to Colour Images. In: Proceedings of 8th Computer Vision Winter Workshop. vol. 91; 2003. p. 53–71.

[13] Lai YK, Rosin PL. Efficient Circular Thresholding. IEEE Transactions on Image Processing. 2014;23(3):992–1001. doi:10.1109/TIP.2013.2297014.

[14] Gelasca ED, Byun J, Obara B, Manjunath BS. Evaluation and Benchmark for Biological Image Segmentation. In: Proceedings of IEEE International Conference on Image Processing; 2008. p. 1816–1819. doi:10.1109/ICIP.2008.4712130.

[15] Plataniotis KN, Venetsanopoulos AN. Color Image Processing and Applications. New York, USA: Springer-Verlag Berlin Heidelberg; 2000. doi:10.1007/978-3-662-04186-4.

[16] Nyul LG, Udupa JK, Zhang X. New Variants of A Method of MRI Scale Standardization. IEEE Transactions on Medical Imaging. 2000;19(2):143–150. doi:10.1109/42.836373.