Computers in Biology and Medicine 111 (2019) 103344

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Color normalization of faded H&E-stained histological images using spectral matching

Thaína A. Azevedo Tosta[a,*], Paulo Rogério de Faria[b], Leandro Alves Neves[c], Marcelo Zanchetta do Nascimento[a,d]

[a] Center of Mathematics, Computing and Cognition, Federal University of ABC, Av. dos Estados, 5001, 09210-580, Santo André, São Paulo, Brazil
[b] Department of Histology and Morphology, Institute of Biomedical Science, Federal University of Uberlândia, Av. Amazonas, S/N, 38405-320, Uberlândia, Minas Gerais, Brazil
[c] Department of Computer Science and Statistics, São Paulo State University, R. Cristóvão Colombo, 2265, 15054-000, São José do Rio Preto, São Paulo, Brazil
[d] Faculty of Computer Science, Federal University of Uberlândia, Av. João Naves de Ávila, 2121, 38400-902, Uberlândia, Minas Gerais, Brazil

ARTICLE INFO

Keywords:
Color normalization
Histological images
Faded histological samples
Spectral matching

ABSTRACT

Histological samples stained with hematoxylin-eosin (H&E) are commonly used by pathologists in cancer diagnoses. However, the preparation, digitization, and storage of tissue samples can lead to color variations that produce poor performance when using histological image processing techniques. Thus, normalization methods have been proposed to adjust the color of the image. This can be achieved through the use of a spectral matching technique, where it is first necessary to estimate the H&E representation and the stain concentration in the image pixels by means of the RGB model. This study presents an estimation method for H&E stain representation for the normalization of faded histological samples. This application has been explored only to a limited extent in the literature, but has the capacity to expand the use of faded samples. To achieve this, the normalized images must have a coherent color representation of the H&E stain with no introduction of noise, which was realized by applying the methodology described in this proposal. The estimation method presented here aims to normalize histological samples with different degrees of fading using a combination of fuzzy theory and the Cuckoo search algorithm, and dictionary learning with an initialization method for optimization. In visual and quantitative comparisons of estimates of H&E stain representation from the literature, our proposed method achieved very good results, with a high feature similarity between the original and normalized images.

1. Introduction

The diagnosis of cancer is traditionally made based on a microscopic analysis of histological samples for identification of the morphological features seen in this particular type of disease [1]. Through the digitization of these samples, histological image processing techniques have been developed to aid pathologists in making a diagnosis. These techniques contribute toward diagnosis decisions that are less subjective and more agile, which makes these tools useful in medical applications [2].

Histological images are obtained through the preparation and digitization of tissue samples [3]. The identification of histological structures is possible due to the color defined by the hematoxylin-eosin (H&E) stain commonly used in histological image analysis [4,5]. However, the tasks of preparation, digitization, and storage make the histological sample susceptible to color variations.

The fixation step, which prevents tissue degeneration, can intensify the staining according to the fixative used and its fixation time [6,7]. When tissue dehydration is not performed efficiently, the stain may be opaque due to water droplets, and may eventually lose microscopic detail [8]. The tissue cutting and sample mounting steps can also lead to color variations, due to the tissue thickness and the introduction of artifacts such as dust and air bubbles, respectively [8,9]. The staining step may be influenced by the concentration and pH of the staining solution, chemical oxidation, and the staining time [4]. Different digitization systems also lead to color variations [10]. Finally, the storage conditions of the samples can alter their interaction with the stain, which after several years can cause tissue fading [3,11].

In addition to the challenges facing the development of computer-aided diagnosis (CAD) systems, variations in color can significantly

* Corresponding author.addresess:
E-mail addresses: tosta.thaina@gmail.com (T.A.A. Tosta), paulo.faria@ufu.br (P.R. de Faria), neves.leandro@gmail.com (L.A. Neves), marcelo.zanchetta@gmail.com (M.Z. do Nascimento).

https://doi.org/10.1016/j.compbiomed.2019.103344
Received 6 February 2019; Received in revised form 24 June 2019; Accepted 24 June 2019

reduce the performance of segmentation and classification methods due to the high sensitivity of color and texture features [3,12]. Normalization techniques are therefore useful in adjusting the color of H&E-stained histological images [12]. These techniques can be classified into three types: histogram matching, color transfer, and spectral matching [3]. Spectral matching has advantages over the other techniques due to its potential tissue preservation, an essential criterion for evaluating the normalization together with the introduction of noise [3,13].

Spectral matching methods are based on estimates of stain representation by color channel and the concentration of the stain in the image pixels. These estimates therefore need to robust and accurate, so that the histological structures can be preserved and noise is not introduced into the normalized image [3]. Through the use of these estimates, our method adjusts the stain representation of an original image, replacing it with a representation of a reference image chosen based on a visual analysis of its colors.

To ensure that our method is robust for different histological images, we used a new approach to estimate an important parameter. This estimation employs fuzzy theory to associate the intensities of the pixels with the hematoxylin and eosin dyes, with parameter estimation carried out using the Cuckoo search (CS). We were then able to obtain a global optimum result with the stain representation initialization by applying the technique reported in Ref. [3]. Finally, we investigated the estimation of the stain representation using dictionary learning [14].

One of the contributions of this work is a new estimation method for the normalization of faded H&E histological images. The color variations in histological images have negative consequences in terms of both diagnosis definition and the computational analysis of images [15]. For this reason, the proposal and use of color normalization are encouraged. This application can be extended to faded histology images, opening up the possibility of their use in more computational applications, for example in composing classifier training sets and in manual analyses by pathologists in their specific areas of interest.

Further contributions of this work are summarized below:

- A new normalization method was investigated based on biological properties and stain characteristics, bringing biomedical concepts closer to computational analyses;
- The proposed application allows for a broader use of faded histological images in manual and computational analyses;
- In addition to proposing a normalization method for faded images, this study also evaluated the performance of other normalization techniques in this specific application, which is a poorly investigated research topic.

This study is divided into five sections. The next section describes previous spectral matching methods for histological image normalization that have been published in the literature. Section 3 presents an overview of the proposed normalization, the image dataset, the method for estimating the stain representation and concentration, and the evaluation metrics of the results. In Section 4, the results of the proposed technique are discussed and compared to the spectral matching of Khan et al. [16], Macenko et al. [17] and Vahadane et al. [18], and the color transferences of Reinhard et al. [19] and Nguyen et al. [20]. Finally, Section 5 presents the conclusion and suggestions for future studies based on the limitations of this proposal.

2. Related studies

Between 2008 and 2017, the majority of normalization studies used color transfer techniques, except in 2015 [21]. Several studies published after 2017 proposed new H&E image normalization methods using spectral matching. In these methods, the estimates relied on non-negative matrix factorization (NMF) and some variations of this approach were widely used due to its non-negativity restriction, which is consistent with the biological properties of H&E [18]. Bilgin et al. [4], Li and Plataniotis [3], Saraswat and Arya [22], and Maji and Mahapatra [23] used the NMF to normalize histological images. In addition to the non-negativity restriction, Vahadane et al. [13] and Vahadane et al. [18] used the concept of sparsity, interpreting the pixels of the histological image as an H&E sparse mixture, which defined the sparse non-negative matrix factorization (SNMF). A second variation of NMF, known as graph regularized sparse non-negative matrix factorization (GSNMF), was used by Sha et al. [24] in an attempt to preserve texture information.

Bilgin et al. [4] performed NMF estimates using the non-negative least squares method with a stain concentration initialization to improve the representation of the eosin dye. After the stain concentration and representation were estimated, residuals were calculated to represent structures that cannot be fully represented by either dye, such as pigments and red blood cells. These residual channels were obtained by subtracting the reconstructed channels defined by the stain concentration and representation from the original channels. However, this study used gamma correction on the stain concentration estimation, leading to poor structure preservation through the altering of the intrinsic characteristics of the pixel.

The authors Li and Plataniotis [3] also used the NMF in conjunction with an initialization to estimate the stain representation, in order to reduce the influence of achromatic pixels. In addition to spectral normalization, the authors proposed an illumination normalization to correct inconsistencies in the tissue digitization step. The limitations of this method were not identified, since the authors did not specify the NMF algorithm used for the normalization.

As in the study by Bilgin et al. [4], Saraswat and Arya [22] also proposed a stain concentration initialization but using the channels a and b from the Lab color model. An empirical stain representation initialization using the RGB model was also proposed, and different NMF algorithms were evaluated. Color correction was performed by applying the method of Reinhard et al. [19]. The limitations of [22] were observed in terms of the definition of a maximum number of iterations for obtaining the stain representation. In this approach, there is no guarantee that the best estimation will be obtained for different images. In addition, this particular study incorporated an empirical stain representation initialization that was different from the unsupervised method used in our proposal.

Maji and Mahapatra [23] used NMF combined with a rough-fuzzy circular clustering and a saturation-weighted hue histogram. This algorithm clustered the histogram into three classes: achromatic, hematoxylin, and eosin, using the rough sets method. This assignment of classes to hues was weighted by the fuzzy membership values with the aim of representing the stain mixture. This was then optimized by a Lagrangian function with a cosine distance, and the stain representation was computed from the cluster centroids. The stain concentration was initially estimated and then refined with the stain representation using NMF to incorporate its non-negativity property. As in Ref. [3], the authors did not specify their NMF algorithm, presenting a detailed description of only the initialization technique. In this study, a limitation was observed in terms of the empirical definition of the parameter that controls the clustering, with an elevated influence on the centroid computation.

In addition to the non-negativity restriction, Vahadane et al. [13] and Vahadane et al. [18] used the sparsity concept through the SNMF technique. This concept was represented by the sparsity parameter (λ) in estimates performed using least angle regression and parameter-free block-coordinate descent with warm restart. One restriction in those studies is the empirical definition of λ, which makes it less robust.

Sha et al. [24] normalized histological images with GSNMF, based on its potential to improve data representation in comparison to NMF. After the conversion of the RGB images to the Lab model, each pixel was used as a vertex on a graph, and its n nearest neighbors (with a 5×5 kernel) were employed to form edges. These edges were weighted using the heat kernel method, but only the highest 50% of the weight edges

were used in the SNMF objective function to estimate the stain representation and concentration. However, no clear definition of its parameters was given, with only visual analyses of the results. Despite the motivation of the study, which was to preserve texture information, it was not quantitatively evaluated based on this criterion.

Unlike in the aforementioned works, Macenko et al. [17] defined the stain representation using the 1st and 99th percentiles of the optical density angles in relation to the first singular value decomposition (SVD) direction. One limitation of this method is the possibility that negative coefficients are obtained in its estimates, which constitutes an invalid biological condition [18]. Moreover, the method can provide inefficient results if the areas of hematoxylin and eosin regions are disproportionate [25]. The study also did not present a quantitative evaluation of its results, which makes it harder to verify its performance.

Since the method of Macenko et al. [17] can produce poor results if there are more hematoxylin than eosin regions or vice versa, Niethammer et al. [26] and Vicory et al. [11] proposed improvements based on this limitation. These studies considered that the image optical densities were distributed on a plane estimated by plane fitting. The authors also proposed partitioning the data points using the Otsu technique, so that the plane could not be biased to the most abundant dye. In this approach, the contribution of each dye cluster was the same as in the stain representation estimation through the use of weights defined by the number of points in each cluster. However, the primary limitation of these studies was reported by Khan et al. [16] as being convergence to local minima.

Another proposed method for improving the normalization of Macenko et al. [17] was published by Stanisavljevic et al. [27]. To reach this goal, the authors presented a multi-core implementation of the SVD, an iterative multi-batch implementation, and single-node and distributed multi-node versions. These enabled faster image processing that supported high-magnification images. Despite its computational optimization, this work did not propose improvements to the limitations of [17].

A study by Celis and Romero [28] defined the stain representation based on the minimum and maximum values of the $R - B$ means along the G-axis via an $R - B$, G plane. In contrast, Bejnordi et al. [25] proposed that microscopic images were better represented by the HSD color model, in comparison to the RGB and HSI models. This condition highlights the main limitation of the Celis and Romero method [28], since they used the RGB model. As an alternative to RGB, it is common in spectral matching techniques to use the optical density of pixels, which represents stain biological concepts [29].

Khan et al. [16] classified the pixels into hematoxylin, eosin, and background using the relevance vector machine method. As features, this classification used the RGB values and the mean and eigenvectors of covariance of the histograms quantized images. The mean color values of the hematoxylin, eosin, and background were used to define the stain representation. Color adjustment was performed using the B-spline method between statistical metrics of these classes from the original and reference images. One of the problems associated with this study was the use of a training set, which limited its robustness to the color variations in this set [18,25,28].

Another spectral matching method was published by Zheng et al. [30], in which stain representation estimation was carried out using the gradient descent algorithm and an objective function optimization with the sum of three weighted functions. The first function was based on prior knowledge that enforces the assignment of each pixel to one dye. The second function was used to control the proportion of the dyes represented in the image, so that the result was not biased when the image had more hematoxylin or eosin regions. Finally, the last function controlled the energy of the stain, with influence in the estimated stain intensity. The main limitation of this method was due to the first function, which assumed that the pixels represented only one dye and disregarded pixel modeling as a stain mixture in relative proportions [18].

According to Vahadane et al. [18], errors in the stain representation estimation are propagated to the estimation of pixels stain concentrations. For this reason, the stain representation must be efficiently estimated in order to achieve relevant results from normalization. Of the studies described above, we note that investigations into the normalization of faded H&E histological images are recent, and are limited in the literature. To the best of our knowledge, only the study of Vicory et al. [11] presented an evaluation for dealing with this color variation. Based on the abovementioned spectral matching studies, can be seen that this topic still faces great challenges ahead, and that new research is necessary for further progress to be made with this approach.

3. Materials and methods

3.1. Method overview

The proposed normalization is based on three steps: estimation of a sparsity parameter, initialization of the stain color appearance matrix, and its estimation. These steps are performed in this order, and are fully described in Subsection 3.3.

The algorithm was developed using the MatLab® language and the experiments were performed on a 2.40 GHz processor notebook (Acer E5-574-592S). Its processor was an Intel Core 6 i5 with 8 GB RAM.

3.2. Image dataset

The proposed spectral matching was evaluated on H&E histological samples of cutaneous melanoma scanned over different periods of time, with differences between two and seven years[1] [11]. The second sample scan therefore presents fading color variations. The evaluated dataset is composed of 23 pairs of whole slide images (WSIs) represented using the RGB color model (24 bits of quantization in tif files) with minimum dimensions of 1867×1616 pixels and maximum 3709×3552 pixels. Evaluation by a pathologist allowed us to divide these images into groups of 11, eight and four pairs with low, medium, and high degrees of fading, respectively.

Since the first and second scans were performed at different times, the histological samples were at different positions in the images, which also had different dimensions. This is noted in Fig. 1 (a), which presents the first scan of a histological sample with 1818×1802 pixels. Fig. 1(b) shows the second scan of the same histological sample, with 1728×1908 pixels. In the evaluation of our results, we had to compare the normalized version of the second scan with the corresponding first scan. These images were therefore registered so that the represented histological sample could be positioned in the exact same space in the images, with the same dimensions.

The registration of the corresponding WSIs was performed with the four control points used for the application of an affine geometric transform in the MatLab® language. Fig. 1(c) presents the result of registration of Fig. 1(a) according to Fig. 1(b), giving it the dimensions of 1728×1908 pixels.

3.3. Proposed method

Following the work by Vahadane et al. [18], let V be the relative optical density, W the stain color appearance matrix with dimensions 3×2 and H the stain density map represented by $2 \times n$, where n corresponds to the total number of pixels in the image. The Beer-Lambert law allows the definition of:

$$V = W \cdot H. \tag{1}$$

The V matrix is defined by Equation (2), in which I is the RGB image with matrix dimensions of $3 \times n$ and I_0 is the light intensity in the

[1] Available at https://midas3.kitware.com/midas/folder/11138.

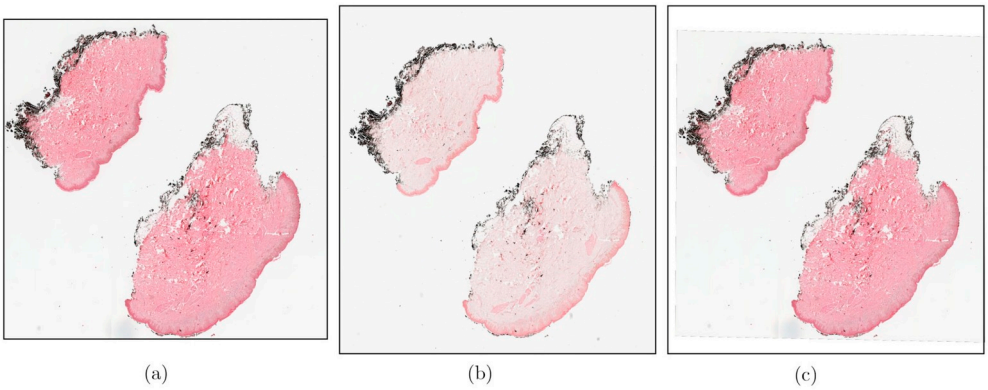


Fig. 1. WSIs registration of the case *BLGM0008* necessary to evaluate the normalization: (a) first sample scan in its original version, (b) the second sample scan, characterized by fading, and (c) the registered version of the first scan.

histological sample, which is equivalent to 255 in the proposed normalization [18].

$$V = log\frac{I_0}{I}. \tag{2}$$

The W matrix corresponds to the representation of the H&E stain (distributed into its columns) in terms of the RGB channels (represented by its rows) [3]. The H map corresponds to the estimation of the H&E stain concentration in each image pixel, which gives it the dimensions $2 \times n$ [4].

The aim of normalization is to estimate the W matrix and the H map of the original and reference images. Color correction is performed by the replacement of the W matrix of the original image by the W matrix of the reference image. Using this approach, the structures represented by the original image are preserved by retaining the H map, and only the H&E stain color of the original image is adjusted according to the reference image.

Fig. 2 illustrates the methodology proposed by Vahadane et al. [18] for the processing of WSIs, as employed in this normalization. The WSIs were divided into grids, and W_i matrices were estimated on the i regions centered in the intersection of this division, as indicated by the green dots in Fig. 2. To prevent the background regions affecting the estimation of W, pixels with L values from the Lab model higher than 0.9 were disregarded, as defined by Vahadane et al. [18]. The W matrix was then obtained based on the median of the W_i matrices of the regions considered.

To estimate the W matrix and the H map, SNMF was used to represent the biological properties of the H&E histological samples. Based on these properties, these estimates were obtained using Equation (3) [14]:

$$\min_{W,H} \tfrac{1}{2}||V - WH||_1^1 + \lambda \sum_{j=1}^{2} ||H(j,:)||_1, \text{ such that } W, H \geq 0, \quad ||W(:,j)||_2^2 = 1. \tag{3}$$

For this purpose, it was necessary to define the parameter λ, the matrix $W_{initial}$ corresponding to the W matrix initialization, and the parameter h that defines the number of pixels processed at each iteration of the algorithm, called the mini-batch.

The parameter λ was estimated using the fuzzy theory with three-parameter membership to model the image as an H&E stain mixture [31]. This method allowed us to assign membership degrees of hematoxylin and eosin to each image intensity level, which quantify the H&E representation by the image intensities. To achieve this, it was necessary to estimate the parameters (u, v and w), which was performed using a CS. Fig. 3 illustrates the different distributions of membership degrees using fuzzy theory with the parameters u, v and w.

As shown in Fig. 3, the membership degrees represent different distributions according to the values of u, v and w used in the M functions, which quantify the membership degrees in Equations (4) and (5):

$$M_{Hematoxylin} = S(k, u, v, w), \tag{4}$$

$$M_{Eosin} = Z(k, u, v, w), \tag{5}$$

where k represents the image intensity levels, and the S and Z functions are defined by Ref. [31]:

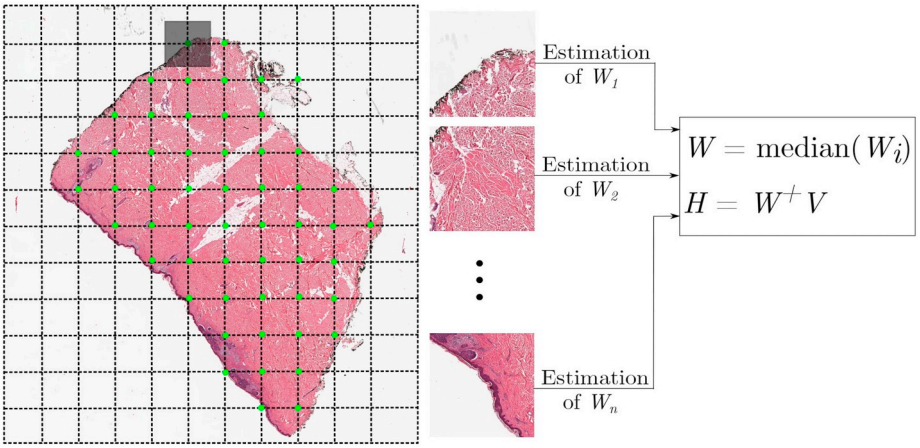
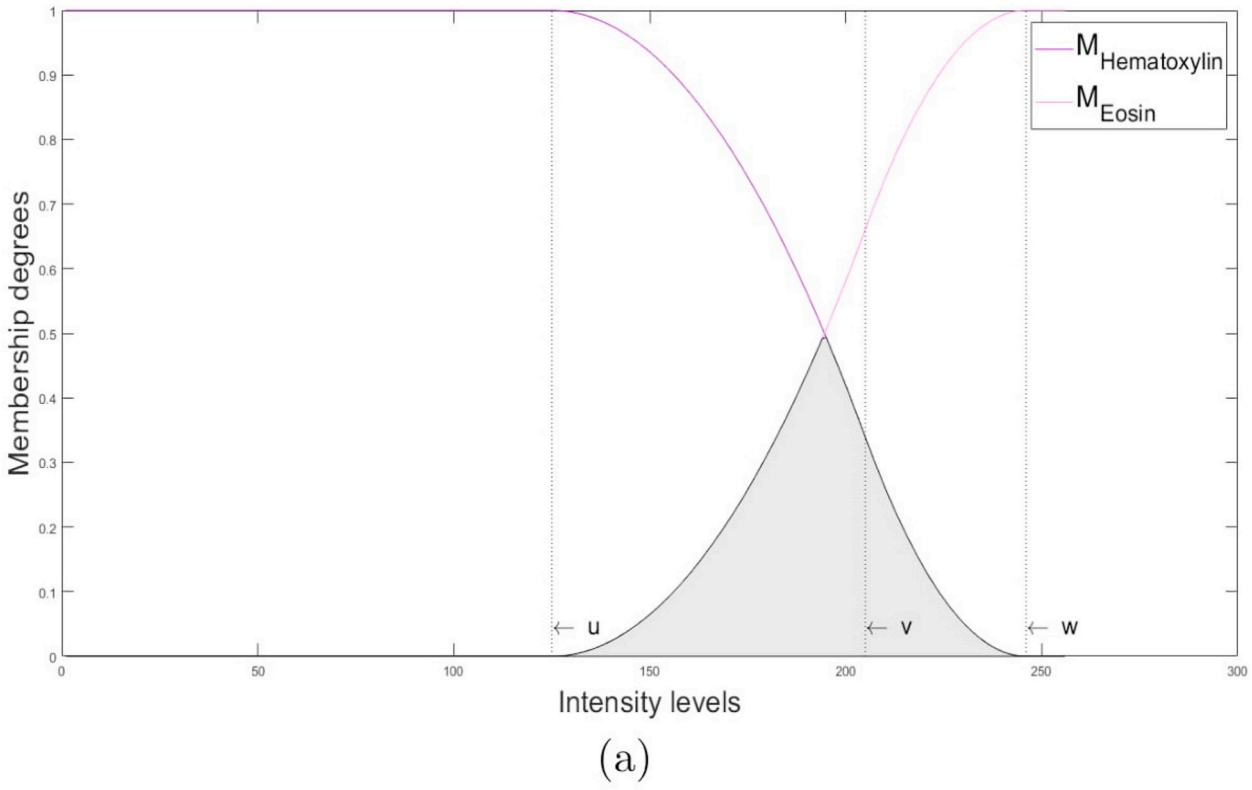


Fig. 2. Flowchart of the method used to estimate the W matrix and the H map of the WSIs (Figure adapted from Ref. [18]© 2016 IEEE).



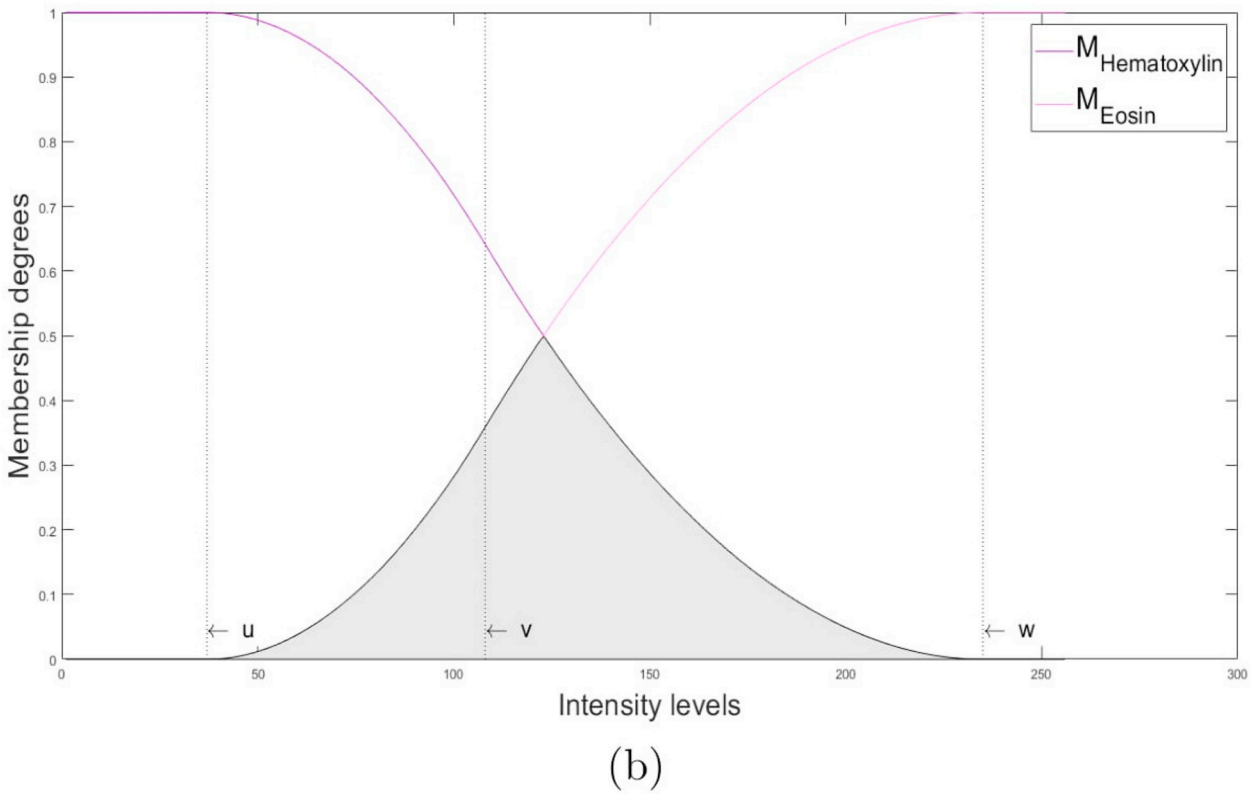


Fig. 3. Distributions of membership degrees (Y-axis) of the intensity levels (X-axis) defined by the $M_{Hematoxylin}$ and M_{Eosin} curves through the assignments of (a) (125, 205, 246) and (b) (37, 108, 235) to the (u, v, w) tuple.

$$S(k, u, v, w) = \begin{cases} 1, & k \leq u \\ 1 - \frac{(k-u)2}{(w-u)\cdot(v-u)}, & u < k \leq v \\ \frac{(k-w)2}{(w-u)\cdot(w-v)}, & v < k \leq w \\ 0, & k > w \end{cases} \quad (6)$$

$$Z(k, u, v, w) = 1 - S(k, u, v, w). \quad (7)$$

Assignment to the (u, v, w) tuple was performed using a CS [32]. This method was chosen based on a comparison of the different optimization algorithms, as described in Subsection 4.1. The CS is based on the reproduction strategy of the cuckoo. This involves parasitism, as cuckoos lay their eggs in the nests of other birds called hosts. In this way, a cuckoo chick can share the food of the host chicks, and can therefore grow [33].

Initially, the CS used a random generation with 100 tuples, as in Ref. [34], sorted by $0 \leq u \leq v \leq w \leq 255$. These tuples represent the host nests, and the u, v and w parameters correspond to the host eggs. To compare the assignments to the tuples, the Shannon entropy (E) was used, as shown in Equation (8) [31,35]. This metric is a quantitative measure of the information represented by the image, and reaches higher values as more information is represented [36]. This metric was calculated only in the intersection region between the $M_{Hematoxylin}$ and M_{Eosin} curves, as indicated by the gray color in Fig. 3, in an attempt to represent the H&E sparse mixture. In this approach, a high value of E indicates the presence of H&E stain regions.

$$E(u, v, w) = -P_{Hematoxylin}\cdot log(P_{Hematoxylin}) - P_{Eosin}\cdot log(P_{Eosin}), \quad (8)$$

where the probabilities (P) of the H&E stain are determined by:

$$P_{Hematoxylin} = \sum_{k=0}^{255} p(k)\cdot M_{Hematoxylin}(k), \quad P_{Eosin} = \sum_{k=0}^{255} p(k)\cdot M_{Eosin}(k), \quad (9)$$

where $p(k)$ represents the probability of occurrence of the k intensity.

To explore different assignments given to the fuzzy parameters (u, v and w), the CS is based on three principles: (1) one cuckoo egg is laid in a random nest; (2) the best nest, in our case tuple, is always kept for the next iteration; and (3) the cuckoo egg can be discovered by the host bird with a probability of p_a, assigned to 0.25 [32].

After the random generation of 100 host nests, the best one is identified by the highest entropy value of all the tuples. This nest cannot receive a cuckoo egg, and it is left untouched for the next generation. Excluding this nest, the laying of the cuckoo eggs is modeled in the CS by Lévy flight among the 99 remaining nests, implemented as described in Ref. [37]. This step obtains a new nest population at each iteration of the algorithm, always retaining the best identified nest.

In our method, if the host bird discovers the cuckoo egg, it throws it away [33]. In order to achieve this, a random probability is generated for each cuckoo egg, which represents the fuzzy parameters. If this probability is higher than p_a, a new value is obtained for the fuzzy parameter using a random approach. Then, all the nests are evaluated using the entropy function (Equation (8)), and the best nest is identified in each iteration of the algorithm.

The method of Hammouche et al. [38] was employed as the CS stop criteria. With this technique, the CS was executed until the mean of the H&E stain intensity levels, as defined by the $M_{Hematoxylin}$ and M_{Eosin} curves, was the same over two consecutive iterations.

After identification of the best assignment to the (u, v, w) tuple, the membership degree distribution (c) of the intersection region of the $M_{Hematoxylin}$ and M_{Eosin} curves allowed us to calculate λ using Equation (10) [39]:

$$\lambda = \frac{\sqrt{\sum_{k=0}^{255} c_k^2}}{\sum_{k=0}^{255} c_k}. \quad (10)$$

The initialization of W ($W_{initial}$) proposed by Li and Plataniotis [3] was used to obtain stable results. This method disregards achromatic pixels corresponding to noise or regions that are poorly stained, which can affect the estimation of W.

To iteratively estimate the W matrix using mini-batches of size h, iterative estimation of H was carried out using the weighted least squares method [40]:

$$H = \underset{H}{argmin} \left|\left| \begin{pmatrix} V \\ 0 \end{pmatrix} - \begin{pmatrix} W \\ \lambda I_{id} \end{pmatrix} H \right|\right|, \quad (11)$$

where I_{id} is an identity matrix. In the first iteration, the W matrix was assigned to $W_{initial}$, and in the following iterations, the coefficients of W were iteratively updated in each mini-batch using a linear system based on the conjugate gradient method [14]:

$$C_t^j = W(j, :)\cdot M_t^j, \quad (12)$$

where j represents each channel of the RGB color model, $M_t^j = M_{t-1}^j + \sum_{i=1}^{h} w_i^j H_i H_i^T$, $C_t^j = C_{t-1}^j \sum_{i=1}^{h} w_i^j V_i^j H_i^T$, and t represents the index of the different mini-batches of the image. In the first iteration, the variables M and C were assigned to null values. The w_i^j variable was defined by:

$$w_i^j = \frac{1}{\sqrt{(V_i^j - W(j)H_i)^2 + \delta}}, \quad (13)$$

where δ was assigned as 10^{-6}.

Following this, the estimation of W was defined by the median of the W_i matrices of the regions resulting from the WSI division (Fig. 2), and the H map was defined by Equation (14) [29]:

$$H = W^+ V, \quad (14)$$

where $H \geq 0$ and $W^+ = (W^T W)^{-1} W^T$.

3.4. Evaluation metrics

The estimation of λ by the CS method was evaluated based on the values assigned to this variable, the number of iterations to convergence and its execution time. The proposed normalization was also evaluated based on its running time, according to the different sizes of each image.

The feature similarity index metric (FSIM) was used for the quantitative evaluation of the WSI normalization. The FSIM quantifies the similarity of features from the original and normalized images [32,41]; the higher the values of this metric, the better the results.

In addition to these metrics, the Canberra distance [42] and the earth mover's distance (EMD) [11] were applied to quantify the difference in the histograms between the normalized WSIs and the corresponding first scan, which was used as the reference image. These metrics were employed for the RGB color channels, giving three values that indicated good results by their minimization.

The color correction of the faded samples was also analyzed using the histograms of the density channel from the HSD color model. The HSD was designed for stain representation in microscopy images [43], and its density channel is linearly related to the amount of stain in the histological sample [25,44]. It was therefore expected that the normalization of the second scan would have a density histogram distribution that was close to that of the first scan image (the reference image).

4. Results and discussion

The evaluation of the proposed normalization scheme is presented in this section, and the CS method is compared to other optimization methods in Subsection 4.1. The results of this normalization were evaluated by applying the first scan of each the histological samples as the reference image, giving 23 reference images in total. The color of

the second scan of the 23 samples, which is characterized by fading, was adjusted according to its respective first scan. Estimation of the W matrix was also evaluated using a reference image extracted from a colorectal cancer WSI, as illustrated in Fig. 5(a). Using this methodology, the 46 images in the dataset were evaluated as original images, and these are presented in Subsection 4.2.

Based on these different reference images, we compared our method to the spectral matching and color transfer techniques, as described in Subsection 4.3. We also analyzed the influence of artifacts in the histological samples on the results of normalization, as presented in Subsection 4.4. Finally, we evaluated the running time of the proposed method for W estimation in comparison to other estimates from the literature, and this is presented in Subsection 4.5.

4.1. Evaluation of optimization algorithms for estimation of λ

In order to evaluate the estimation of λ, we performed a quantitative comparison of the CS with the evolutionary algorithms of the artificial bee colony (ABC) [45], differential evolution (DE) [46], genetic algorithm (GA) [47] and particle swarm optimization (PSO) [48]. These algorithms were evaluated based on the obtained values for λ, the number of iterations to convergence and their execution time. This analysis was performed using 100 executions on the 46 dataset images, and the mean results are presented in Table 1.

We investigated the results for λ, the number of iterations and the execution times using the Kruskal-Wallis test with a 5% significance level. This test was applied because none of the optimization methods had a normal distribution. All of the algorithms showed no statistically significant differences ($p < 0.05$) over 100 executions for each criterion evaluated.

Table 1 shows the performance of the optimization algorithms, with a maximum number of iterations of around six and a convergence time of around 0.4 s. Experiments performed by Vahadane et al. [18] indicated that the larger the value of λ, the more local structures are erased. In comparison to a ground truth, assignments of around 0.1 reach a flat normalization performance. In this regard, the techniques of CS, DE, and GA stand out, since they obtained results closest to 0.1. However, the number of iterations to convergence and the execution time showed that CS was the best optimization algorithm for our application.

4.2. Visual evaluation of the proposed method

Fig. 4 illustrates the normalization results of the original images, which are characterized by high (Fig. 4(b)), medium (Fig. 4(f)) and low (Fig. 4(j)) fading, with application of their first scan as the reference images, as shown in Figs. 4(a), 4(e) and 4(i), respectively. The results of the proposed method are shown in Figs. 4(c), 4(g) and 4(k).

Normalization of the sample with high fading (Fig. 4(b)) gave good results for tissue representation, as can be seen in the magnified region in Fig. 4(c). For the representation of background artifacts, as indicated by the green arrow in Fig. 4(c), the proposed technique gave good results. Our method gives a similar representation of the artifacts to that seen in the original image, as shown in shades of green and blue.

Table 1
Mean results for the values of λ, the number of iterations to convergence and the execution time (in seconds) of ABC, CS, DE, GA and PSO for λ estimation using 100 executions on the 46 dataset images.

Optimization algorithms	λ	Number of iterations	Execution time
ABC	0.1967	3.9493	0.3383
CS	0.1366	1.7263	0.0915
DE	0.1449	3.5376	0.1591
GA	0.1400	2.3237	0.1045
PSO	0.1709	5.5952	0.3897

The normalization of the sample with medium fading (Fig. 4(f)) achieved significant results using the proposed method (Fig. 4(g)), since the image background was preserved and the coherent H&E color representation was clearly visible. In addition, the proposed estimation method was able to preserve the color of the artifact, as indicated by the green arrow in Figs. 4(f) and 4(g). Our normalization represented the tissue region in less vibrant shades of pink than in the original image, which was more similar to the reference image (Fig. 4(e)).

The normalization of the sample with low fading (Fig. 4(j)) achieved significant results using our technique (Fig. 4(k)), when considering the criteria of the preservation of structures and introduction of noise. This can be seen in the amplified regions, where the structures represented in the original image are also present in the normalized images. In addition, there is no alteration in the background coloring or color of the small artifacts, as indicated by the green arrow. This condition shows that high performance is achieved for color correction when the original and reference images are similar, which is a desirable condition in spectral matching techniques with blind estimates.

In order to analyze only the tissue color, we performed a segmentation of the reference images (Figs. 4(a), 4(e) and 4(i)) by the Otsu method, which uses the sum of the density and the saturation channels from the HSD model. The masks obtained in this way were applied to the reference, original and normalized images in Fig. 4. The density values of the segmented histological regions were plotted on histograms with intervals of [0–255], as illustrated in Figs. 4(d), 4(h) and 4(l).

Fig. 4(d) shows that the proposed method gave a density distribution with greater similarity to the reference image for normalization of the sample with high fading than the original image. Our method was able to improve the density distribution over the entire interval of [0–255]. The high fading sample is indicated in this representation by the peak of the original image curve at [0–50], representing a low stain density. This was corrected in our method by the stretching of this peak to around [0–80].

The distributions in Fig. 4(h) demonstrate that our method obtained good results, with a close distribution to the reference image. The original image also has a similar density distribution to the reference image; this is as expected, since we are dealing with the case of medium fading. However, the original image curve has a high peak at [0–50], which was smoothed by our technique. Between 100 and 255, the original curve shows a clear difference from the reference curve. In this interval, the proposed method obtained a similar distribution to the reference curve, thus correcting the original image density.

The curves in Fig. 4(l) demonstrate the good performance of our technique in the normalization of the sample with low fading. Between 0 and around 90, the original image has a close distribution to the reference image. However, above this value, these curves show a marked difference. At around [0–80], our technique increased the density of the original image to above the reference curve, but the limitation of the original curve around [90–120] was corrected by our method, with high similarity to the reference image. Additionally, the proposed method also gave a distribution with greater similarity to the reference image than the original curve.

We also evaluated our normalization using a colorectal cancer image as a reference image to correct the color of the 46 WSIs. Fig. 5 exemplifies the results of the proposed method (Fig. 5(c)) for normalization of the original image (Fig. 5(b)) using the reference image in Fig. 5(a).

This figure again demonstrates that the proposed method (Fig. 5(c)) gave significant results. Considering the criterion of the preservation of structures, we note that the histological regions of the amplified area in the original image (Fig. 5(b)) were also represented in the results from our method. We obtained a coherent stain color, with pink shades for the eosin and purple shades for the hematoxylin. These results are also relevant in terms of the criterion of the introduction of noise, since noise was not added to the normalized image and the background color is well represented.

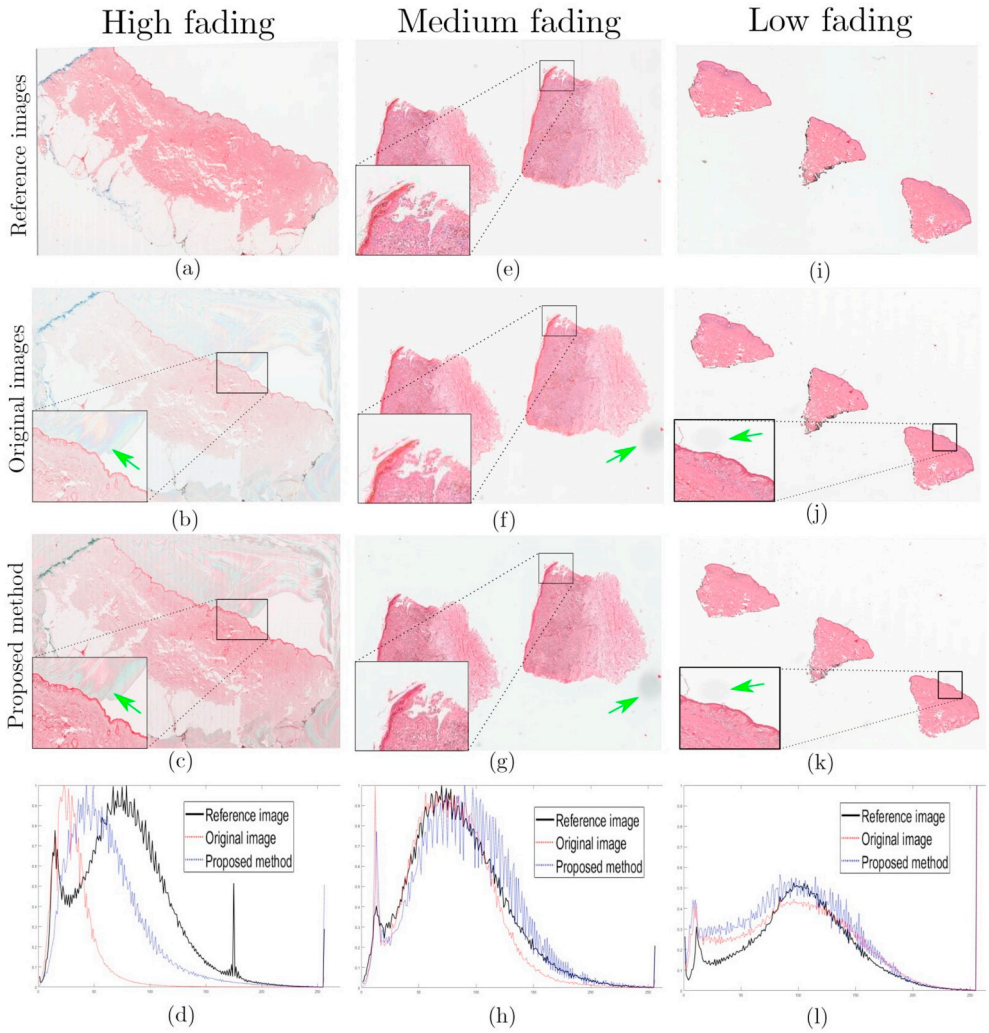


Fig. 4. Original images with high (b), medium (f) and low (j) fading with their first scan used as reference images ((a) case *BLGM0001*, (e) case *blgm6082* scaled for illustration, and (i) case *blgm6129*, respectively), normalized by the proposed *W* estimation ((c), (g) and (k)) with their density histograms, respectively in (d), (h) and (l).

4.3. Performance comparison with other studies

The estimation of *W* via the proposed method was compared to estimates using the methods of Khan et al. [16], Macenko et al. [17] and Vahadane et al. [18]. In this evaluation, the *H* map estimates for the proposed method and for these techniques were performed using Equation (14). In this way, we were able to compare the performance of the estimation of *W* for each spectral matching task. We also compared our proposed normalization method to the color transferences of Reinhard et al. [19] and Nguyen et al. [20].

Tables 2 and 3 present the quantitative normalization results using the first scan as the reference image. These tables present the mean and standard deviation of the FSIM and EMD metrics of the proposed normalization and those of Khan et al. [16], Macenko et al. [17] and Vahadane et al. [18] for all spectral matching techniques. In samples with high fading, the proposed method achieved the highest FSIM result of the spectral matching methods, with a value of 0.9608. The same is noted for the normalizations of the medium and low fading samples, with results of 0.9786 and 0.9981, respectively. Based on the EMD, our normalization achieved the best results for all RGB channels and fading degrees except for the *R* channel of the high fading sample.

Both color transferences obtained very good quantitative results based on the FSIM and EMD metrics. However, it is important to note that the methodology of Reinhard et al. [19] assumes that the reference

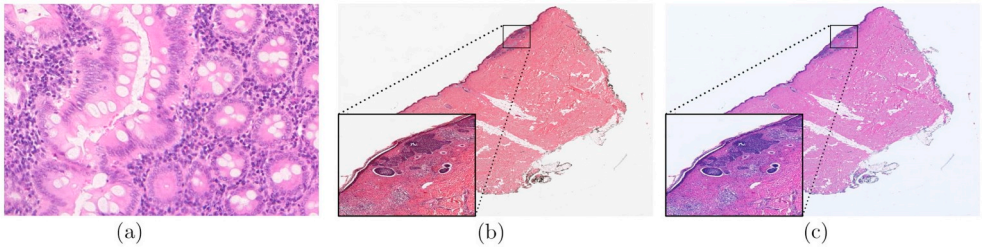


Fig. 5. Normalization results of an original WSI (case *blgm6003A*) (b) with a reference image of colorectal cancer (a) by the proposed *W* estimation (c).

Table 2
Mean and standard deviation of the FSIM metric of the proposed normalization and the spectral matching methods of Khan et al. [16], Macenko et al. [17] and Vahadane et al. [18], and the color transferences of Reinhard et al. [19] and Nguyen et al. [20] with the histological samples first scan as reference images.

Class	Normalization	High fading	Medium fading	Low fading
Spectral matching	Proposed	0.9608 (0.0312)	0.9786 (0.0376)	0.9981 (0.0016)
	Khan et al. [16]	0.8963 (0.0756)	0.9272 (0.0151)	0.9258 (0.0713)
	Macenko et al. [17]	0.9313 (0.0627)	0.9655 (0.0169)	0.9732 (0.0493)
	Vahadane et al. [18]	0.8841 (0.0638)	0.9279 (0.0479)	0.9566 (0.0587)
Color transfer	Reinhard et al. [19]	0.9344 (0.0529)	0.9768 (0.0240)	0.9983 (0.0041)
	Nguyen et al. [20]	0.9765 (0.0141)	0.9751 (0.0217)	0.9975 (0.0049)

Table 3
Mean and standard deviation of the EMD in the *R*, *G* and *B* channels between the normalized images and its respective first scan used as reference images.

Class	Normalization	High fading			Medium fading			Low fading		
		R	*G*	*B*	*R*	*G*	*B*	*R*	*G*	*B*
Spectral matching	Proposed	6.71 (3.59)	11.93 (2.56)	9.93 (2.56)	5.55 (4.11)	10.07 (5.61)	8.39 (4.48)	2.91 (1.23)	4.88 (2.94)	4.48 (2.49)
	Khan et al. [16]	9.88 (4.62)	18.40 (4.77)	15.62 (3.65)	8.13 (6.07)	16.88 (10.32)	12.58 (6.03)	12.06 (9.52)	24.67 (16.61)	18.15 (12.51)
	Macenko et al. [17]	6.74 (3.70)	14.11 (4.67)	12.29 (4.10)	6.24 (2.99)	11.02 (3.33)	9.45 (2.61)	3.82 (3.50)	8.24 (7.40)	6.07 (6.07)
	Vahadane et al. [18]	5.53 (1.98)	17.30 (4.38)	12.38 (3.60)	7.12 (3.02)	13.73 (6.01)	10.13 (3.73)	5.99 (6.41)	13.96 (11.02)	9.83 (8.86)
Color transfer	Reinhard et al. [19]	4.75 (2.01)	8.97 (2.27)	7.11 (2.09)	3.91 (2.11)	6.62 (3.94)	5.56 (3.07)	1.66 (0.69)	2.91 (2.88)	2.75 (2.06)
	Nguyen et al. [20]	4.22 (1.22)	12.40 (4.12)	9.89 (3.15)	3.63 (2.00)	6.28 (4.22)	5.39 (3.28)	2.23 (1.10)	3.17 (2.38)	2.96 (1.90)

and original images have the same content; this is not a correct assumption but is satisfactory in this application [18]. The method of Nguyen et al. [20] also achieved good results, but the color correction was extremely subtle, with minimal alteration of the background in the normalized images. Since this region is a huge part of the WSIs, these quantitative results are highly augmented, even with poor tissue color correction.

In the normalization of the samples with high fading, the method of Khan et al. [16] obtained good FSIM and EMD results, although the artifacts had an incoherent color, with dark shades of pink rather than green and blue. Based on this criterion, the results from the method of Macenko et al. [17] were similar to those of Khan et al. [16]. However, the former method achieved a better tissue color result, as demonstrated by its higher FSIM value and lower EMD result. We also noted the low performance of the method of Vahadane et al. [18] samples with high fading, in terms of the most incoherent representation of the artifacts and the tissue, as indicated by its FSIM result. Even with bad EMD results in the *G* and *B* channels, this technique achieved the best *R* channel result. The *R* channel histogram from the reference image represents the pink shades from the tissue, but the histogram for the channel from Vahadane et al. [18]'s result represents the background in shades of pink, which makes the EMD of their *R* channel low.

In the normalization of the sample with high fading in Fig. 4(b), the method of Reinhard et al. [19] obtained a good stain color representation in the tissue, as shown magnified in Fig. 6(a). This method was also able to correctly represent the artifacts indicated by a green arrow, preserving their green and blue shades, as in the proposed method. However, the method of Nguyen et al. [20] had a much more subtle color correction, with greater similarity to the original image (Fig. 4(b)), as shown in Fig. 6(b). Based on a consideration of the artifact from the original image, this method obtained significant results. However, it was not able to correct the tissue color of this degree of fading in the same way as the proposed normalization method. This is demonstrated in Fig. 6(c), which shows that the curve for the method of Nguyen et al. [20] is very similar to the original image density distribution. The method proposed by Reinhard et al. [19] obtained better performance, but its curve has a lower degree of similarity to the reference image than the proposed method.

A statistical analysis of the FSIM and EMD results of the normalization of the high fading samples, given by Tables 2 and 3, was performed using the Kruskal-Wallis test with a 5% significance level. Considering the spectral matching methods and color transferences, the FSIM and EMD results did not show a statistically significant difference ($p < 0.05$).

In the normalization of the medium fading sample, the method of Khan et al. [16] gave an alteration to the original WSI background, with a marked color change in its artifact, as indicated by the FSIM and EMD results. This is also shown by the green arrow in Fig. 7(a), which represents the normalization result of Fig. 4(f) by Khan et al. [16]. In addition, the histological sample boundary contained a high amount of noise; this can be seen in the amplified region, which has a brighter color representation than the background, as indicated by a red arrow. This limitation does not apply to the proposed normalization method, as shown in Fig. 4(g). Using the quality of the histological sample representation as an evaluation criterion, the method of Khan et al. [16] gave a blurry representation with loss of some histological detail, as can

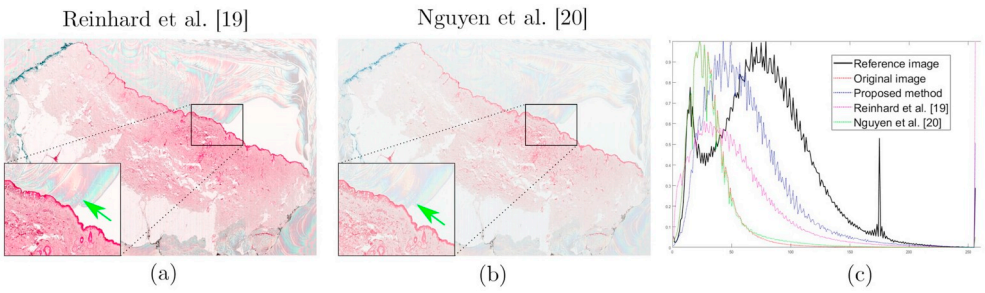


Fig. 6. Results of the color transferences of Reinhard et al. [19] and Nguyen et al. [20] using the high fading reference and original images in Fig. 4. Their stain density histograms using only the tissue regions segmented from Fig. 4(a) is presented in (c).

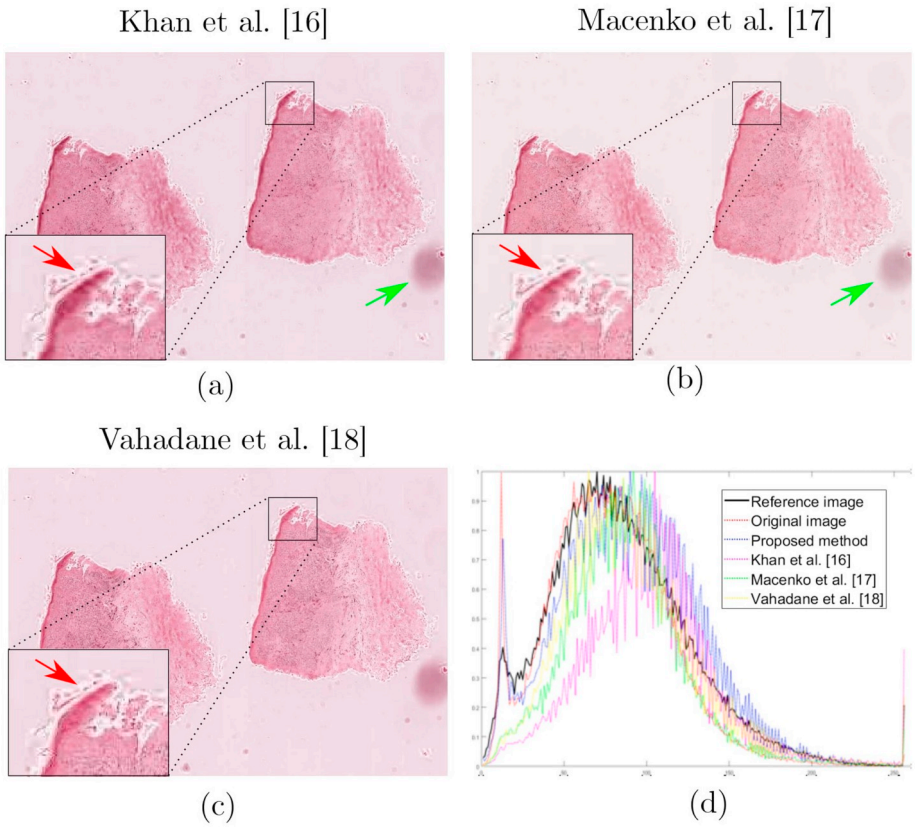


Fig. 7. Results of the spectral matching methods of Khan et al. [16], Macenko et al. [17] and Vahadane et al. [18] using the medium fading reference and original images in Fig. 4. Their stain density histograms using only the tissue regions segmented from Fig. 4(f) is presented in (d).

be observed in the magnified region of Fig. 7(a).

As in the study by Khan et al. [16], Macenko et al. [17] also obtained background coloring, noise representation in the histological sample boundary and a blurry tissue representation, as indicated in Fig. 7(b). The same is noted in the result obtained by Vahadane et al. [18] (Fig. 7(c)), which led to FSIM and EMD results of poorer quality than the proposed method. However, the background color from the method of Macenko et al. [17] is better represented than the results of Khan et al. [16] and Vahadane et al. [18], as expressed by their higher FSIM value and lower EMD result.

The common limitations of the methods of Khan et al. [16], Macenko et al. [17] and Vahadane et al. [18] can affect analyses by both pathologists and CAD systems. In an analysis carried out by pathologists, the loss of information due to the blurring of a histological sample means that certain details that are important for diagnoses may disappear. In CAD systems, these limitations can impose challenges on the segmentation and feature extraction steps. Although the introduction of noise and the background coloring do not alter the tissue regions, they may influence these steps by altering the features of the sample boundary and the background, respectively.

The density distributions of the sample with medium fading in Fig. 7(d) demonstrate that the results from the methods of Khan et al. [16], Macenko et al. [17] and Vahadane et al. [18] are very different from that of the reference image, especially between 0 and around 80 on the X-axis. Of the normalization methods investigated, our proposed technique obtained relevant results, with a distribution close to that of the reference image even when the other normalization algorithms did not achieve this.

As in our normalization, both color transferences did not suffer from the limitations of the methods of Khan et al. [16], Macenko et al. [17] and Vahadane et al. [18] for the normalization of medium fading. This condition represents a similar aspect between these methods and the proposed normalization and a positive difference in comparison to the other spectral matching methods. This condition is expressed by a statistical analysis of the FSIM and EMD results for the medium fading samples using the Kolmogorov-Smirnov test with a 5% significance level. The proposed normalization method was not statistically different ($p > 0.05$) from the methods of Reinhard et al. [19] or Nguyen et al. [20]. However, the spectral matching methods showed a statistically significant difference ($p < 0.05$).

The normalization of a sample with low fading achieved good results with any of the spectral matching methods, with the lowest value for FSIM of 0.9258. In the EMD results, poor performance was achieved by the method of Khan et al. [16] due to the background coloring obtained by this method. Among the spectral matching techniques, we obtained the best results across all RGB channels. With regard to the color transferences, these techniques also obtained very good FSIM results of close to 1, and low EMD values. This evaluation gave particularly good results for Reinhard et al.'s method [19]. The poor tissue color correction of Nguyen et al. [20] is not expressed by these metrics.

The Kolmogorov-Smirnov test with a 5% significance level for the FSIM results indicated that the spectral matching techniques showed statistically significant differences ($p < 0.05$). However, the proposed method showed a significant difference only in comparison to the method of Reinhard et al. [19]. With regard to the EMD results, the proposed method showed statistically significant differences ($p < 0.05$) in all channels only in comparison to the spectral matching techniques.

We also evaluated the normalization techniques using the colorectal cancer image in Fig. 5(a) as a reference for correcting the color of the 46 WSIs. Fig. 8 illustrates the results from the methods of Khan et al. [16] (Fig. 8(a)), Macenko et al. [17] (Fig. 8(b)), Vahadane et al. [18] (Fig. 8(c)), Reinhard et al. [19] (Fig. 8(d)) and Nguyen et al. [20] (Fig. 8(e)) for normalization of the original image (Fig. 5(b)).

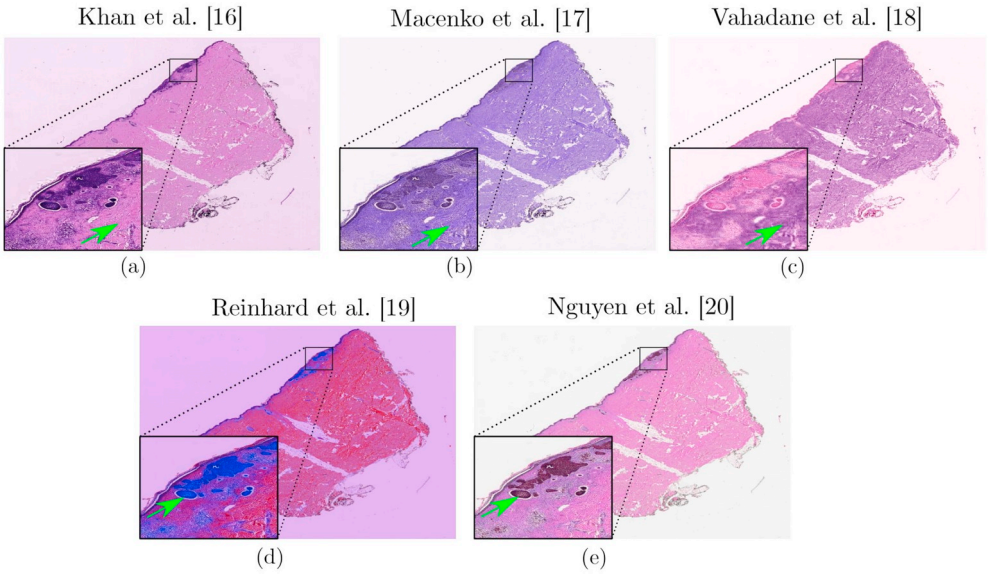


Fig. 8. Normalization results of the original WSI (case *blgm6003A*) from Fig. 5(b) with the reference image from Fig. 5(a) by the methods of Khan et al. [16] (a), Macenko et al. [17] (b), Vahadane et al. [18] (c), Reinhard et al. [19] (d) and Nguyen et al. [20] (e).

The method of Khan et al. [16] (Fig. 8(a)) gave purple shades for the eosin representation, as indicated by a green arrow, and showed small color alterations in the WSI background. Poor visual results were also obtained by the techniques of Macenko et al. [17] (Fig. 8(b)) and Vahadane et al. [18] (Fig. 8(c)), which gave a purple color for the eosin dye. The method of Vahadane et al. [18] also represented the hematoxylin stain in pink shades, indicating that this method reversed the color representation of the dyes.

As shown in Fig. 8(d), the method of Reinhard et al. [19] represented the hematoxylin in blue shades, as indicated by the green arrow, unlike the purple representation of this dye in the reference image. This example shows the poor performance achieved by this method when the original and the reference images have different contents. Moreover, this method obtained a colored background. The normalization of Nguyen et al. [20] was limited in terms of the visual quality of the stain representation. This can be seen in the region indicated by a green arrow in Fig. 8(e), which is represented in purple unlike the expected color, and evidently provided in the result of the proposed method (Fig. 5(c)).

Table 4 presents the FSIM results from the evaluated spectral matching techniques using the reference image in Fig. 5(a). In this evaluation, the proposed method achieved the best results of the spectral matching techniques. The second best result was obtained by the method of Macenko et al. [17], which achieved at least 0.9041 (0.9469–0.0428). Despite the small difference, these values are smaller than the minimum result for FSIM for the proposed technique, which was 0.9336 (0.9706–0.0370).

An evaluation of the color transferences was also quantitatively performed using the FSIM. The techniques of Reinhard et al. [19] and Nguyen et al. [20] achieved good results with small differences compared to our method. Despite the bad visual results, the method of Reinhard et al. [19] obtained a high feature similarity between the normalized and original images. Since the method of Nguyen et al. [20] performs a subtle background color alteration, its FSIM is once again high.

Table 4
Mean and standard deviation of FSIM by the WSIs normalization using the reference image of Fig. 5(a) by the proposed *W* estimation and the techniques of Khan et al. [16], Macenko et al. [17], Vahadane et al. [18], Reinhard et al. [19] and Nguyen et al. [20].

Class	Normalization	FSIM
Spectral matching	Proposed	0.9706 (0.0370)
	Khan et al. [16]	0.8500 (0.0877)
	Macenko et al. [17]	0.9469 (0.0428)
	Vahadane et al. [18]	0.9190 (0.0526)
Color transfer	Reinhard et al. [19]	0.9667 (0.0364)
	Nguyen et al. [20]	0.9816 (0.0213)

4.4. Analysis of the influence of artifacts on normalization results

Table 5 presents the mean and standard deviation of the Canberra distance between the *R*, *G* and *B* channels of the normalized images and their respective first scans. Of the spectral matching methods, our proposed normalization method achieved the best mean results for all degrees of fading. Unlike the color transferences, the best results varied between the proposed method and the techniques of Reinhard et al. [19] and Nguyen et al. [20] for the different degrees of fading and the RGB channels. The color transfer results are a further demonstration of the most appropriate conditions for the application of the Reinhard et al. [19] method and the subtle color correction of Nguyen et al. [20], as described in Subsection 4.3.

Based on the Canberra distances for samples with high fading, the proposed method did not show statistically significant differences ($p > 0.05$) from the spectral matching and color transfer techniques. The same can be seen for the medium fading results. However, for low fading, the Canberra distances of all RGB channels of our proposed method showed statistically significant differences ($p < 0.05$) from the spectral matching methods.

From Table 5, we can see the high standard deviation of the proposed normalization method for the medium fading samples, which guaranteed distances of up to 133.55 (86.15 + 47.40), 119.21 (87.05 + 32.16) and 130.94 (91.12 + 39.82) for the *R*, *G* and *B* channels, respectively. These results can be explained by the many artifacts in this set of samples.

In order to provide more detail on this limitation with regard to medium fading, Fig. 9 illustrates the results of the proposed normalization method (Fig. 9(c)) for the original image (Fig. 9(b)), through the color of the reference image in Fig. 9(a). In the original image (Fig. 9(b)), there are significant artifacts, as indicated by the red arrows in the magnified region, which led to an incoherent color being obtained in the *W* estimation result, as indicated by the red arrows in

Table 5
Mean and standard deviation of the Canberra distance in the *R*, *G* and *B* channels between the normalized images and its respective first scan used as reference images.

Class	Normalization	High fading			Medium fading			Low fading		
		R	*G*	*B*	*R*	*G*	*B*	*R*	*G*	*B*
Spectral matching	Proposed	92.13 (43.98)	104.90 (17.19)	98.39 (24.67)	86.15 (47.40)	87.05 (32.16)	91.12 (39.82)	50.41 (17.56)	34.86 (19.35)	46.23 (20.59)
	Khan et al. [16]	103.53 (35.81)	117.53 (16.22)	114.68 (23.96)	88.06 (27.59)	97.44 (16.41)	100.44 (22.57)	89.60 (46.08)	82.88 (38.50)	90.79 (39.42)
	Macenko et al. [17]	97.94 (46.73)	109.29 (16.68)	102.88 (19.83)	95.94 (38.64)	93.95 (23.30)	99.24 (35.08)	61.22 (31.30)	48.77 (29.21)	51.47 (32.15)
	Vahadane et al. [18]	110.47 (28.75)	126.56 (18.66)	124.15 (20.74)	98.84 (36.12)	96.69 (25.26)	99.58 (30.10)	66.35 (40.73)	57.97 (36.34)	59.71 (39.26)
Color transfer	Reinhard et al. [19]	96.87 (44.03)	108.94 (15.78)	101.15 (26.54)	89.74 (29.27)	84.44 (31.36)	88.21 (34.74)	50.04 (16.61)	37.43 (27.57)	42.39 (26.94)
	Nguyen et al. [20]	90.65 (45.88)	134.95 (18.25)	134.12 (25.34)	63.16 (21.45)	77.27 (35.71)	82.58 (32.87)	45.04 (16.75)	29.30 (12.14)	33.73 (10.77)

Fig. 9(c). The methods of Khan et al. [16], Macenko et al. [17], Vahadane et al. [18], Reinhard et al. [19] and Nguyen et al. [20] presented a more coherent artifact color than that obtained by the proposed method, as demonstrated by their standard deviation for samples with medium fading in Table 5. Based on this evaluation, our limitation is defined in dealing with artifacts, which are altered in the normalized versions of the histological images.

It is important to note that even with small alterations in the artifacts, such as those obtained by the methods of Khan et al. [16], Macenko et al. [17], Vahadane et al. [18], Reinhard et al. [19] and Nguyen et al. [20], the performance of CAD systems that use these images can be negatively influenced if the typical dark color of the artifacts is lost. In an analysis by pathologists, the alteration of these artifacts may not be very significant, but it is still necessary to verify its influence on the tissue representation.

4.5. Running time of *W* estimation in spectral matching methods

We evaluated the computational efficiency of the spectral matching methods based on the running time necessary to estimate the *W* matrix over the 46 images from the dataset used. This experiment was performed on the machine described in Subsection 3.1.

The WSIs have a number of pixels (n) that is between 800 thousand ($0.8 \cdot 10^6$) and 22 million ($22 \cdot 10^6$). We analyzed the mean execution time for *W* estimation according to n, as shown in Table 6.

Table 6 shows that all of the *W* estimation techniques take longer to execute with an increase in n. Regardless of the number of pixels, the method of Macenko et al. [17] performed very well based on this evaluation criterion. Although the normalization module does not represent a bottleneck in a CAD system, the results of this method are subject to the limitations discussed in Subsection 4.3, which can be undesirable for system results. The same conclusions can be drawn for the method of Vahadane et al. [18].

Using the method presented by Khan et al. [16], the estimation of *W* required the longest execution times of the techniques compared, as this uses a pixel-wise classifier [30]. In our method, the high execution times are a consequence of the use of mini-batches, which leads to a time-consuming iterative approach. This analysis highlights the necessity of improving this limitation, while preserving the quality of the results already obtained.

5. Conclusion

The color shown in H&E-stained histological images is susceptible to variations due to the preparation and digitization steps of the tissue samples. Furthermore, the storage of samples can alter the interaction between tissue and stain, leading to fading over time. The performance of the segmentation and classification methods applied to faded histological images can be affected by the high sensitivity of these techniques to color and texture features, which are influenced by color variations.

Normalization techniques are useful in medical research to adjust the color of histological images. The small number of studies dedicated to the color correction of faded histological images means that an investigation of the normalization of these images is necessary so that histological image processing systems can be made more robust to this condition. This paper therefore presents a proposal for the *W* matrix estimation of a spectral matching method, which has potential for the preservation of histological structures.

The use of SNMF allowed us to adopt biological concepts for the proposed normalization, meaning that it has a valid biological interpretation, and thus an advantage over the methods presented by Khan et al. [16] and Macenko et al. [17]. In an attempt to deal with the empirical definition of λ by Vahadane et al. [18], the unsupervised method of assignment makes our technique robust to images with great color variations. The method of Lu et al. [14] for the estimation of *W*

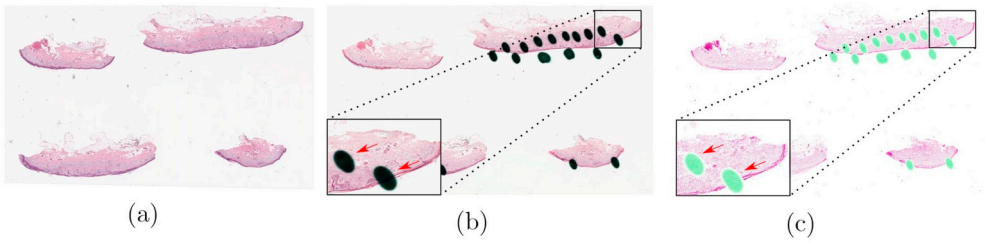


Fig. 9. Normalization of a medium fading sample (case *BLGM1022*) (b) with the color of its first scan (a) by the proposed technique (c).

Table 6
Mean execution times of the spectral matching methods of Khan et al. [16], Macenko et al. [17], Vahadane et al. [18] and the proposed estimation, in seconds, considering the number of pixels in the images (n) in the scale of 10^6.

W estimation	$n < 4$	$4 < n < 12$	$n > 12$
Proposed	17.35	48.20	90.51
Khan et al. [16]	21.67	55.10	183.28
Macenko et al. [17]	3.61	8.71	20.20
Vahadane et al. [18]	6.95	11.26	20.13

was also relevant in the representation of the H&E color, background regions and tissue boundary, which are limitations that have been observed in the results of Khan et al. [16], Macenko et al. [17] and Vahadane et al. [18] (Figs. 7(a)-(c)). The density histogram (Fig. 7(d)), the EMD metric (Table 3) and the Canberra distance (Table 5) demonstrated the satisfactory performance of this technique in correcting the color of faded images, giving color distributions of greater similarity to those of the reference image than the other spectral matching methods. Based on the FSIM metric (Tables 2 and 4), the good performance of this proposal was evident due to the high similarity of the resulting features in relation to the original images.

The proposed technique gives comparable quantitative results to the color transferences of Reinhard et al. [19] and Nguyen et al. [20] (Tables 2–5). However, evaluation of the density histograms (Fig. 6(c)) indicated that the method of Nguyen et al. [20] gave poor color correction of highly faded samples, which is the most important evaluation for this application. The advantage of the proposed technique in comparison to the method of Reinhard et al. [19] is demonstrated by the use of a colorectal sample as a reference image (Fig. 8), which results in a poor color representation of the tissue and the background.

The evaluation of the similarity between the normalized WSIs and their respective first scans, as quantified by the Canberra distance, identified the main limitation of this method. From the many artifacts in the second scan of the samples with medium fading, we can see the low performance of the proposed normalization method. This condition requires future studies of methods that are able to identify these artifacts and minimize their influence on the estimation of W. Using the Canberra distance with the density histograms (Figs. 4, 6 and 7), we show that although the tissue color correction by the proposed technique gives good performance, it is still possible to achieve a better representation of the background of the WSIs. In future studies, we therefore plan to improve the representation of these regions in a way that preserves the good tissue color correction already obtained.

The evaluation of the running time (Table 6) indicates that a possible improvement could be made in the proposed W estimation in future work, in order to make these results as good as those achieved for the visual and quantitative evaluations. Future work will also explore the definition of the parameters of the CS to verify their influence on the visual and quantitative results of the normalization. We also intend to evaluate different membership functions that can be employed in this evolutionary algorithm. We expect to use more images to examine the influence of the optimization algorithms on the normalization results, using a ground truth as in Ref. [18]. Finally, we plan to prepare a new faded WSIs dataset with larger images in order to analyze the performance of this algorithm on larger data dimensions. Since the purpose of this work is the normalization of faded histological samples, it is necessary to wait for the samples themselves to fade, which means this is an experiment for the future.

Conflicts of interest

None declared.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The authors gratefully acknowledge the financial support of National Council for Scientific and Technological Development – CNPq (Grant #304848/2018-2, Grant #430965/2018-4 and Grant #313365/2018-0), the State of Minas Gerais Research Foundation - FAPEMIG (Grant #APQ-00578-18) and CAPES (Grant #1575210).

References

[1] C. Demir, B. Yener, Automated Cancer Diagnosis Based on Histopathological Images: a Systematic Survey, Rensselaer Polytechnic Institute, Tech. Rep, 2005.
[2] M. Oger, P. Belhomme, M.N. Gurcan, A general framework for the segmentation of follicular lymphoma virtual slides, Comput. Med. Imag. Graph. 36 (6) (2012) 442–451.
[3] X. Li, K.N. Plataniotis, A complete color normalization approach to histopathology images using color cues computed from saturation-weighted statistics, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 62 (7) (2015) 1862–1873.
[4] C.C. Bilgin, J. Rittscher, R. Filkins, A. Can, Digitally adjusting chromogenic dye proportions in brightfield microscopy images, J. Microsc. 245 (3) (2012) 319–330.
[5] K. Kuru, Optimization and enhancement of h&e stained microscopical images by applying bilinear interpolation method on lab color mode, Theor. Biol. Med. Model. 11 (9) (2014).
[6] A. Babic, I.R. Loftin, S. Stanislaw, M. Wang, R. Miller, S.M. Warren, W. Zhang, A. Lau, M. Miller, P. Wu, et al., The impact of pre-analytical processing on staining quality for h&e, dual hapten, dual color in situ hybridization and fluorescent in situ hybridization assays, Methods 52 (4) (2010) 287–300.
[7] D.H. Cormack, A.W. Ham, Histologia, 8 edition, (1983).
[8] G.O. Rolls, N.J. Farmer, J.B. Hall, Artifacts in Histological and Cytological Preparations vol. 106, Leica Microsystems, Melbourne, 2008.
[9] E. Michail, E.N. Kornaropoulos, K. Dimitropoulos, N. Grammalidis, T. Koletsa, I. Kostopoulos, Detection of centroblasts in h&e stained images of follicular lymphoma, 22nd Signal Processing and Communications Applications Conference (SIU), IEEE, 2014, pp. 2319–2322.
[10] A. Janowczyk, A. Basavanhally, A. Madabhushi, Stain normalization using sparse autoencoders (stanosa): application to digital pathology, Comput. Med. Imag. Graph. 57 (2017) 50–61.
[11] J. Vicory, H.D. Couture, N.E. Thomas, D. Borland, J.S. Marron, J. Woosley, M. Niethammer, Appearance normalization of histology slides, Comput. Med. Imag. Graph. 43 (2015) 89–98.
[12] R.A. Hoffman, S. Kothari, M.D. Wang, Comparison of normalization algorithms for cross-batch color segmentation of histopathological images, Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2014, pp. 194–197.
[13] A. Vahadane, T. Peng, S. Albarqouni, M. Baust, K. Steiger, A.M. Schlitter, A. Sethi, I. Esposito, N. Navab, Structure-preserved color normalization for histological images, International Symposium on Biomedical Imaging, IEEE, 2015, pp. 1012–1015.
[14] C. Lu, J. Shi, J. Jia, Online robust dictionary learning, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 415–422.
[15] M. Yamaguchi, Computer-aided differentiation for pathology images, Image-Based

Computer-Assisted Radiation Therapy, Springer, 2017, pp. 67–84.
[16] A.M. Khan, N. Rajpoot, D. Treanor, D. Magee, A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 61 (6) (2014) 1729–1738.
[17] M. Macenko, M. Niethammer, J.S. Marron, D. Borland, J.T. Woosley, X. Guan, C. Schmitt, N.E. Thomas, A method for normalizing histology slides for quantitative analysis, International Symposium on Biomedical Imaging: from Nano to Macro, IEEE, 2009, pp. 1107–1110.
[18] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A.M. Schlitter, I. Esposito, N. Navab, Structure-preserving color normalization and sparse stain separation for histological images, IEEE Trans. Med. Imaging 35 (8) (2016) 1962–1971.
[19] E. Reinhard, M. Adhikhmin, B. Gooch, P. Shirley, Color transfer between images, IEEE Comput. Graph. Appl. Mag. 21 (5) (2001) 34–41.
[20] L. Nguyen, A.B. Tosun, J.L. Fine, A.V. Lee, D.L. Taylor, S.C. Chennubhotla, Spatial statistics for segmenting histological structures in h&e stained tissue images, IEEE Trans. Med. Imaging 36 (7) (2017) 1522–1532.
[21] T.A.A. Tosta, P.R. de Faria, L.A. Neves, M.Z. do Nascimento, Computational normalization of h&e-stained histological images: progress, challenges and future potential, Artif. Intell. Med. 95 (2018) 118–132.
[22] M. Saraswat, K.V. Arya, Colour normalisation of histopathological images, Comput. Methods Biomech. Biomed. Eng.: Imaging Vis. 1 (4) (2013) 185–197.
[23] P. Maji, S. Mahapatra, Rough-fuzzy circular clustering for color normalization of histological images, Fundam. Inf. 164 (1) (2019) 103–117.
[24] L. Sha, D. Schonfeld, A. Sethi, Color normalization of histology slides using graph regularized sparse nmf, Medical Imaging 2017: Digital Pathology, vol. 10140, International Society for Optics and Photonics, 20171014010–1 – 1014010–11.
[25] B.E. Bejnordi, G. Litjens, N. Timofeeva, I. Otte-Höller, A. Homeyer, N. Karssemeijer, J. AWM van der Laak, Stain specific standardization of whole-slide histopathological images, IEEE Trans. Med. Imaging 35 (2) (2016) 404–415.
[26] M. Niethammer, D. Borland, J.S. Marron, J.T. Woosley, N.E. Thomas, Appearance normalization of histology slides, Mach. Learn. Med. Imaging 6357 (2010) 58–66.
[27] M. Stanisavljevic, A. Anghel, N. Papandreou, S. Andani, P. Pati, J.H. Rüschoff, P. Wild, M. Gabrani, H. Pozidis, A fast and scalable pipeline for stain normalization of whole-slide images in histopathology, European Conference on Computer Vision, Springer, 2018, pp. 424–436.
[28] R. Celis, E. Romero, Unsupervised color normalisation for h and e stained histopathology image analysis, International Symposium on Medical Information Processing and Analysis, Page 968104, International Society for Optics and Photonics, 2015.
[29] A.C. Ruifrok, D.A. Johnston, et al., Quantification of histochemical staining by color deconvolution, AQCH (Anal. Quant. Cytol. Histol.) 23 (4) (2001) 291–299.
[30] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, J. Shi, C. Xue, Adaptive color deconvolution for histological wsi normalization, Comput. Methods Progr. Biomed. 170 (2019) 107–120.
[31] Y. Tang, W. Mu, Y. Zhang, X. Zhang, A fast recursive algorithm based on fuzzy 2-partition entropy approach for threshold selection, Neurocomputing 74 (17) (2011) 3072–3078.
[32] A.K. Bhandari, A. Kumar, G.K. Singh, Tsallis entropy based multilevel thresholding for colored satellite image segmentation using evolutionary algorithms, Expert Syst. Appl. 42 (22) (2015) 8707–8730.
[33] X. Yang, S. Deb, Cuckoo search via lévy flights, World Congress on Nature & Biologically Inspired Computing (NaBIC), IEEE, 2009, pp. 210–214 2009.
[34] M. Guerrero, O. Castillo, M. García, Cuckoo search via lévy flights and a comparison with genetic algorithms, Fuzzy Logic Augmentation of Nature-Inspired Optimization Metaheuristics, Springer, 2015, pp. 91–103.
[35] S. Walkowski, J. Szymas, Quality evaluation of virtual slides using methods based on comparing common image areas, Diagn. Pathol. 6 (1) (2011) S14.
[36] D. Tsai, Y. Lee, E. Matsuyama, Information entropy measure for evaluation of image quality, J. Digit. Imaging 21 (3) (2008) 338–347.
[37] X. Yang, S. Deb, Multiobjective cuckoo search for design optimization, Comput. Oper. Res. 40 (6) (2013) 1616–1624.
[38] K. Hammouche, M. Diaf, P. Siarry, A multilevel automatic thresholding method based on a genetic algorithm for a fast image segmentation, Comput. Vis. Image Understand. 109 (2) (2008) 163–175.
[39] N. Hurley, S. Rickard, Comparing measures of sparsity, IEEE Trans. Inf. Theory 55 (10) (2009) 4723–4741.
[40] C. Zhao, X. Wang, W. Cham, Background subtraction via robust dictionary learning, J. Image Video Process. (1) (2011) 972961 2011.
[41] L. Zhang, L. Zhang, X. Mou, D. Zhang, Fsim: a feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (8) (2011) 2378–2386.
[42] S. Mukhopadhaya, Exploring Measures of Similarity and Dissimilarity for Fuzzy Classifier: from Data Quality to Distance Quality, Master's thesis University of Twente, 2016.
[43] J.A.W.M. van der Laak, M.M.M. Pahlplatz, A.G.J.M. Hanselaar, P. de Wilde, Hue-saturation-density (hsd) model for stain recognition in digital images from transmitted light microscopy, Cytometry 39 (4) (2000) 275–284.
[44] B.E. Bejnordi, N. Timofeeva, I. Otte-Holler, N. Karssemeijer, J. van der Laak, Quantitative analysis of stain variability in histology slides and an algorithm for standardization, Proc. Med. Imag.: Digit. Pathol. (2014) 9041.
[45] Ankita Bose, Kalyani Mali, Fuzzy-based artificial bee colony optimization for gray image segmentation, Signal Image Video Process. 10 (6) (2016) 1089–1096.
[46] E. Cuevas, D. Zaldívar, M. Perez-Cisneros, Image segmentation based on differential evolution optimization, Applications of Evolutionary Computation in Image Processing and Pattern Recognition, Springer, 2016, pp. 9–22.
[47] M. Paulinas, A. Ušinskas, A survey of genetic algorithms applications for image enhancement and segmentation, Inf. Technol. Control 36 (3) (2007).
[48] K.P. Remamany, T. Chelliah, K. Chandrasekaran, K. Subramanian, Brain tumor segmentation in mri images using integrated modified pso-fuzzy approach, Int. Arab J. Inf. Technol. 12 (6A) (2015) 797–804.