A Robust BKSVD Method for Blind Color Deconvolution and Blood Detection on H&E Histological Images

Fernando Pérez-Bueno[1][0000−0002−7404−794X], Kjersti Engan[2][0000−0002−8970−0067], and Rafael Molina[1][0000−0003−4694−8588]

[1] Dpto. de Ciencias de la Computación e I. A., Universidad de Granada, Spain.*
fpb@ugr.es, rms@decsai.ugr.es

[2] Department of Electrical Engineering and Computer Science, University of Stavanger, Norway. kjersti.engan@uis.no

Abstract. Hematoxylin and Eosin (H&E) color variation between histological images from different laboratories degrades the performance of Computer-Aided Diagnosis systems. Histology-specific models to solve color variation are designed taking into account the staining procedure, where most color variations are introduced. In particular, Blind Color Deconvolution (BCD) methods aim to identify the real underlying colors in the image and to separate the tissue structure from the color information. A commonly used assumption is that images are stained with and only with the pure staining colors (e.g., blue and pink for H&E). However, this assumption does not hold true in the presence of common artifacts such as blood, where the blood cells need a third color component to be represented. Blood usually hampers the ability of color standardization algorithms to correctly identify the stains in the image, producing unexpected outputs. In this work, we propose a robust Bayesian K-Singular Value Decomposition (BKSVD) model to simultaneously detect blood and separate color from structure in histological images. Our method was tested on synthetic and real images containing different amounts of blood pixels.

1 Introduction

The development of Computer-Aided Diagnosis (CAD) systems for the analysis of Whole Slide Images (WSIs) is not exempt from challenges [9]. Just using data from different hospitals can hamper their performance [11], mostly due to color variation and artifacts in the image [6]. Therefore, preprocessing is often a key step [6] for reliable CAD systems.

Blind Color Deconvolution (BCD) methods [12] estimate the image-specific stain colors and structure (concentration). The separation itself can reduce the impact of color variation [14] and it is often included as a step for other approaches such as color normalization [19] or color augmentation [18]. See [6]

* Supported by project B-TIC-324-UGR20 FEDER/Junta de Andalucía and Universidad de Granada.

for a complete survey. However, the assumption that the image contains only the expected stains (e.g. Blue for nuclei and pink for cytoplasm and connective tissue in H&E images), is not true when artifacts are present. Artifacts such as blood, degrade the quality and diagnosis value of a WSI and introduce additional sources of color variation [6]. Blood artifacts react differently to H&E staining and get stained in a completely different color [3] (usually red), which is often used to detect them [3] but also hampers the performance of BCD methods [2].

Despite the relationship between artifacts and color, artifact detection and color variation have hardly been explored together. The presence of blood and other artifacts is often ignored by BCD methods. Similarly, it is hard to find works that use BCD to detect artifacts. Our work brings together the fields of artifact detection and color variation by focusing on blood, and how its presence affects BCD methods on H&E images. We propose the use of BCD for blood detection using its difference in color with the H&E stains. For this goal, we use the recently proposed Bayesian K-Singular Value Decomposition (BKSVD) [12] for BCD, which is able to identify the colors of the stains in the image but is affected by blood. We extend it to acknowledge the presence of blood, making it possible to perform blood detection and then obtain an estimation of the H&E stains.

The paper is organized as follows. Section 1.1 describes related works on BCD and blood detection. Section 2 provides an overview of the BKSVD method in [12], discusses the limitations, and provides the necessary improvements for its application to robust blood detection and BCD. In section 3 experimentally evaluate the proposed method. Finally, section 4 concludes the paper.

1.1 Related Work

Blind Color Deconvolution Most BCD methods use the Beer-Lambert law [15], which establishes a linear combination between the stains in the *optical density* (OD) space. Let I be a RGB image $\mathbf{I} \in \mathbb{R}^{3\times Q}$, where each value $i_{cq} \in \mathbf{I}$ correspond to pixel q and channel $c \in$ RGB. Then, the OD is defined as $y_{cq} = -\log(i_{cq}/i^0_{cq})$, where $i^0_{cq} = 255$ denotes the incident light. Then, a WSI $\mathbf{Y}$ stained with S stains follows the equation

$$\mathbf{Y} = \mathbf{MC} + \mathbf{N}\,, \tag{1}$$

where $\mathbf{M} = [\mathbf{m}_1, \cdots, \mathbf{m}_S] \in \mathbb{R}^{3\times S}$ is the normalized stains' specific color-vector matrix; $\mathbf{C} \in \mathbb{R}^{S\times Q}$ is the stain concentration matrix, its q-th column, $\mathbf{c}_q = [c_{1,q}, \ldots, c_{S,q}]^{\mathrm{T}}$, represents the contribution of each stain to the q-th pixel value in $\mathbf{Y}$; and, finally, $\mathbf{N} \in \mathbb{R}^{3\times Q}$ is a Gaussian noise matrix with independent components of variance β^{-1}.

The goal of BCD is to estimate $\mathbf{C}$ and $\mathbf{M}$ from $\mathbf{Y}$. Here we summarize the most relevant approaches in the literature. See [6] for a complete survey. Ruifrok *et al.* [15] use a given color-vector matrix to separate the stains, which is widely used as a standard. However, the actual color is usually considered to be unknown due to color variation. In the work by Macenko *et al.* [8] the H&E channels are

estimated using Singular Value Decomposition (SVD). Vahadane *et al.* [19] estimate the color-vector matrix using Non-Negative Matrix Factorization (NMF) and the assumption that most pixels in the image are stained by a single stain. Alsubaie *et al.* [1] apply Independent Component Analysis (ICA) in the wavelet domain, under the assumption that stains might not be independent. Few works consider the effect of noise and artifacts. The use of the method by Macenko *et al.* [8] is widely extended, although it is known to be affected by noise and artifacts [2]. To speed up the method, Vahadane *et al.* [19] perform a patch sampling of the WSI. They take advantage of their patch-wise stain estimation to calculate the median color-vector and provide a more robust estimation against artifacts. Alsubaie *et al.* [1] included a linear filtering to reduce the noise contamination when estimating stain matrix, but did not consider large artifacts. The presence of artifacts and their effect on color estimation is acknowledged in [2]. The estimation of the color-vector and the robust maximum (99th percentile) of the H&E concentrations were used to identify low-quality images, substituting the color-vector matrix with average estimates from other images when poor quality is detected.

A Bayesian approach is followed in [5] by defining a smoothness prior on the concentrations and a similarity prior on the color-vectors. To improve the quality of the concentration obtained, this work was extended in [13] by using a Total Variation (TV) prior and in [14] with general super-Gaussian priors. These Bayesian methods share a need for a reference color-vector matrix for the similarity prior. The use of a prior on the color [13,14] can reduce the effect of noise and artifacts, but limits the adaptability to different color distributions. To solve this issue, the work in [12] proposes the use of Bayesian K-SVD (BKSVD) to find the color-vector matrix as a dictionary learning problem.

Blood Detection Detection of blood is frequently formulated as a color-related problem. In [3] the authors classify blood segmentation techniques into (i) RGB segmentation, (ii) segmentation using other color space (such as HSV, Lab and LUV) and additional techniques, (iii) segmentation using one or two channels of a non-RGB color space. In [7] the detection of blood is approached with a combination of staining protocol and image processing, using mathematical morphology and thresholding the RGB channels of the image. Sertel *et al.* [16] use the color to distinguish five major components in the H&E images (i.e. nuclei, cytoplasm, background, blood, and extracellular material). First they threshold the RGB channels to remove blood and background, and then use k-means in the L*a*b color space. A Maximum Likelihood Estimation is implemented in [10] to classify the pixels into four classes (blood, cytoplasm, nuclei, and background) in the RGB color space. In [17] the magenta channel of the CMYK space is used to detect blood areas that are later classified into hemorrhages or vessels using mathematical morphology and a decision tree. Hue and saturation are used in [4] to detect intracerebral hemorrhage.

More complex approaches combine the choice of color space with clustering, mathematical morphology, classification, or DL [20]. See [6] for details.

2 Material and Methods

2.1 Bayesian k-SVD for Blind Color Deconvolution

Although the color in the slide is a combination of both H&E stains, each biological structure will present structure-specific color properties [19,12]. This allows pathologists to distinguish structures based on their color. In [12] a Bayesian framework is used to approach the problem as a dictionary learning problem, with sparse concentrations [19], which encourages the framework to find the color-vector matrix $\mathbf{M}$ that better represents the differential staining of the structures in the image.

Following (1), the OD observed image $\mathbf{Y}$ is modeled with a Gaussian distribution $\mathrm{p}(\mathbf{Y}|\mathbf{C},\mathbf{M},\beta)$, where β controls the noise precision. That is, $\mathrm{p}(\mathbf{Y}|\mathbf{C},\mathbf{M},\beta) = \prod_{q=1}^{Q}\mathcal{N}(\mathbf{y}_q|\mathbf{M}\mathbf{c}_q,\beta^{-1}\mathbf{I}_{3\times 3})$. The sparsity of the solution is promoted by a zero-mean Laplace prior on the concentrations, that is $\mathrm{p}(\mathbf{c}_q) \propto \exp\big(-\sqrt{\lambda_q}\|\mathbf{c}_q\|_1\big)$, with $\lambda_q > 0$ the scale parameter. A flat prior $\mathrm{p}(\mathbf{M})$ on the color-vector matrix is used, and unit norm for each column $\mathbf{m}_s$ is assumed on their posterior estimation.

The true posterior $\mathrm{p}(\boldsymbol{\Theta}|\mathbf{Y}) = \mathrm{p}(\mathbf{Y},\boldsymbol{\Theta})/\mathrm{p}(\mathbf{Y})$, where $\boldsymbol{\Theta} = \{\beta,\mathbf{M},\mathbf{C}\}$ is the set of unknowns, is approximated with variational inference [12]. WSIs are usually very large, rendering their processing computationally expensive. The number of pixels necessary to estimate the color-vector matrix $\hat{\mathbf{M}}$ is reduced by using a uniform random sampling of the WSI. The estimations $\hat{\mathbf{M}}$ and $\hat{\mathbf{C}}$ are then initialized to $\underline{\mathbf{M}}$ and $\hat{\mathbf{C}} = \underline{\mathbf{M}}^{+}\mathbf{Y}$, where $\underline{\mathbf{M}}$ is the Ruifrok's standard matrix [15] and $\underline{\mathbf{M}}^{+}$ the Moore-Penrose pseudo-inverse of $\underline{\mathbf{M}}$. After iteratively optimizing the model parameters, estimations for $\hat{\mathbf{M}}$ and $\hat{\mathbf{C}}$ are provided. Notice that the method in [12] did not consider the presence of blood. When the image contains blood artifacts $\mathbf{M}$ will be wrongly estimated because it will be forced to have two color vectors (H&E). Here we extend the method and propose its use for robust blood detection and BCD estimation.

2.2 Robust Blind Color Deconvolution and Blood Detection

Algorithm 1 Robust BKSVD for blood detection

Require: Observed image $\mathbf{I}$, initial normalized $\underline{\mathbf{M}}$, batch size B, threshold thr.
Ensure: Estimated stain color-vector matrix, $\hat{\mathbf{M}}$, concentrations, $\hat{\mathbf{C}}$, blood mask.
1: Obtain the OD image $\mathbf{Y}$ from $\mathbf{I}$
2: First stage: set $S = 3$. Estimate $\hat{\mathbf{M}}$ and $\hat{\mathbf{C}}$ using BKSVD [12]
3: Sort $\hat{\mathbf{M}}$ and $\hat{\mathbf{C}}$ using the correlation of the columns of $\hat{\mathbf{M}}$ and $\underline{\mathbf{M}}$
4: Create blood mask using $\hat{\mathbf{C}}_{:,3} > thr$
5: Remove blood-positive pixels.
6: Second stage: set $S = 2$. Re-estimate $\hat{\mathbf{M}}$ and $\hat{\mathbf{C}}$ on remaining pixels using BKSVD
7: **return** $\hat{\mathbf{M}}$, $\hat{\mathbf{C}} = \hat{\mathbf{M}}^{+}\mathbf{Y}$ and the blood mask.

Blood and other artifacts hamper the estimation of the color-vector matrix and the separation of the stains. Obtaining a robust method to detect blood and perform BCD estimation is closely related to correctly identifying these elements.

BKSVD [12] uses two channels to separate H&E images, but it can be extended to include more channels. A third channel is often used by BCD [15,8,13], which content is considered to be residual when there are only two stains in the image and is often referred to as "background" channel. As blood gets stained in a different color, often used for its detection [3], it can be seen as an additional effective stain, that can be detected using the third channel. However, including a third channel when not necessary, hampers the quality of the H&E estimated color-vectors $\hat{\mathbf{m}}_h$, $\hat{\mathbf{m}}_e$, and the corresponding concentrations [13]. Assuming blood or other elements might appear in the image, we split the BCD process into two stages. The first stage uses three channels to represent all elements in the image. The second stage uses two and focuses on the H&E channels only.

The first stage starts from the H&E vectors from $\underline{\mathbf{M}}$ [15] and a third color-vector $\underline{\mathbf{m}}_b$ which is orthogonal to $\underline{\mathbf{m}}_h$ and $\underline{\mathbf{m}}_e$ [13]. After estimating $\hat{\mathbf{M}}$ and $\hat{\mathbf{C}}$, it is critical to keep the right order of the stain channels. Thus, we calculate the correlation between the columns of $\hat{\mathbf{M}}$ and the initial $\underline{\mathbf{m}}_h$, $\underline{\mathbf{m}}_e$, and choose $\hat{\mathbf{m}}_h$, $\hat{\mathbf{m}}_e$ as those with maximum correlation to their respective references. The remaining channel is selected as the third (blood) channel, and the order of $\hat{\mathbf{C}}$ is modified accordingly.

We utilize the values in the third concentration channel $\hat{\mathbf{C}}_{:,3}$ to detect anomalies. Since, as explained in [12], the dictionary learning approach of BKSVD aims at finding a representation of the image where each pixel is assigned to only one channel whenever possible, pixels assigned or having a higher value in the third channel are not correctly represented by the H&E channels and can be considered an anomaly. Pixels with a blood channel component above a threshold are marked and discarded. The value for the threshold is experimentally determined in section 3.

Finally, utilizing only the blood-free pixels and starting from $\underline{\mathbf{M}}$, the second stage re-estimates the color-vector matrix and concentrations for the remaining pixels. The procedure is summarized in Algorithm 1.

2.3 Databases

– **Synthetic Blood Dataset (SBD)** We use the Warwick Stain Separation Benchmark (WSSB) [1], which contains 24 H&E images of different tissues from different laboratories for which their ground truth color-vector matrix $\mathbf{M}_{GT}$ is known. WSSB images are free from artifacts, therefore we synthetically combine them with blood pixels obtained in the Stavanger University Hospital, incrementally added as columns to each image. This allows us to control the amount of blood in the image, which makes it possible to measure the effect of blood on the estimation of the color-vector matrix. The amounts of blood considered are $\{0, 0.1, 0.2, \ldots, 0.9\}$ times the size of the original images, creating a dataset of 240 images.

Table 1. Jaccard index and F1-score for the blood mask obtained with the proposed method and different threshold values on SBD images.

	Jaccard index						F1-score					
Amount	Threshold value						Threshold value					
of blood	0.2	0.3	0.4	0.5	0.6	0.7	0.2	0.3	0.4	0.5	0.6	0.7
0.1	0.615	0.715	0.797	0.814	0.798	**0.861**	0.722	0.791	0.860	0.871	0.850	**0.913**
0.2	0.729	0.794	**0.874**	0.789	0.778	0.839	0.823	0.859	**0.923**	0.845	0.834	0.896
0.3	0.759	**0.804**	0.786	0.797	0.749	0.720	0.849	**0.868**	0.845	0.865	0.818	0.794
0.4	0.741	0.808	0.802	**0.858**	0.786	0.808	0.828	0.872	0.867	**0.909**	0.852	0.870
0.5	0.791	0.818	0.825	**0.830**	0.731	0.748	0.871	0.885	0.881	**0.888**	0.801	0.825
0.6	0.780	**0.865**	0.774	0.803	0.759	0.764	0.863	**0.920**	0.843	0.866	0.828	0.845
0.7	0.795	0.853	0.794	**0.863**	0.727	0.721	0.873	0.912	0.856	**0.919**	0.808	0.798
0.8	0.794	**0.839**	0.834	0.757	0.759	0.784	0.870	**0.902**	0.894	0.829	0.833	0.864
0.9	**0.833**	0.808	0.809	0.714	0.728	0.655	**0.901**	0.878	0.876	0.797	0.803	0.738
Mean	0.760	**0.811**	0.810	0.803	0.757	0.767	0.844	**0.876**	0.871	0.865	0.825	0.838

– **TCGA Blood dataset** In order to test the method on real images containing blood, we used 8 breast biopsies from The Cancer Genome Atlas (TCGA). Breast biopsies often contain blood due to the biopsy procedure [6] and it is also possible to find blood vessels in the tissue. We selected 16 2000×2000 H&E image patches where blood was manually labeled.

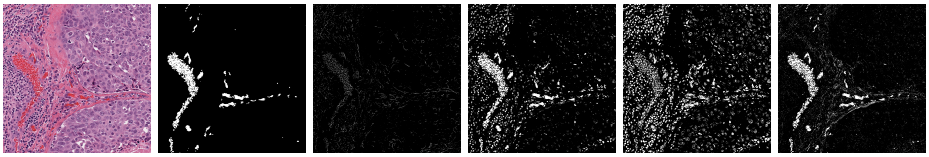
3 Experiments

We have designed a set of experiments to assess our method. First, we examine its blood detection capability and then, its robustness using BCD.

3.1 Blood Detection

Table 2. Jaccard index and F1-score for different approaches to blood detection on the SBD images. The proposed method uses $thr = 0.3$.

	SBD images							
Amount	Jaccard index				F1-score			
of blood	R-RGB	S-HSV	M-CMYK	Proposed	R-RGB	S-HSV	M-CMYK	Proposed
0.1	0.0078	0.6984	0.5369	**0.7152**	0.0154	**0.8010**	0.6814	0.7911
0.2	0.0125	0.7484	0.5964	**0.7936**	0.0247	0.8447	0.7381	**0.8588**
0.3	0.0157	0.7605	0.6156	**0.8037**	0.0309	0.8569	0.7566	**0.8677**
0.4	0.0165	0.7819	0.6385	**0.8078**	0.0324	**0.8727**	0.7756	0.8720
0.5	0.0173	0.7914	0.6498	**0.8181**	0.0340	0.8800	0.7850	**0.8850**
0.6	0.0179	0.7978	0.6577	**0.8646**	0.0352	0.8848	0.7914	**0.9204**
0.7	0.0184	0.8041	0.6650	**0.8531**	0.0360	0.8893	0.7972	**0.9121**
0.8	0.0186	0.8087	0.6700	**0.8391**	0.0365	0.8925	0.8011	**0.9023**
0.9	0.0187	**0.8104**	0.6730	0.8076	0.0368	**0.8938**	0.8034	0.8777
Mean	0.0159	0.7780	0.6337	**0.8114**	0.0313	0.8684	0.7700	**0.8763**
	TCGA images							
	Jaccard index				F1-score			
Mean	0.0402	0.3299	0.1245	**0.4342**	0.0723	0.4667	0.2065	**0.5648**



a) Observed b) GT mask c) R-RGB d) S-HSV e) M-CMYK f) Proposed

Fig. 1. Blood detection comparison on a TCGA image. a) Observed image. b) Ground truth manual mask. c-f) Blood mask obtained by thresholding different channels and with the proposed method.

In this experiment, we assess the performance of the proposed method to identify blood in the image. We evaluate the most appropriate value for thresholding the third channel [13] and compare the results with other approaches for blood detection.

The concentration channels take values $\in [0, -log(1/255)]$. We have evaluated thresholds for the blood component thr in the whole range with a 0.1 step. Pixels with a value above the thr are marked as blood, and the mask is then compared with the ground truth label in the SBD and TCGA datasets. We use the Jaccard index (intersection over union) and the F1-score.

Table 1 shows the performance of the method on the SBD images for different values of the threshold. Low values of the threshold are able to correctly separate the blood in the image without including an excessive number of false positives. The best results in mean are obtained using $thr = 0.3$. Values of $th > 1$ were experimentally found to be irrelevant for blood detection.

Using a value of $thr = 0.3$ we compare the proposed method with the blood detection approaches in [16,4,17] by thresholding the red RGB, the saturation HSV and the magenta CMYK channels, respectively. Results are summarized in table 2. RGB and CMYK do not obtain a close mask, while HSV obtains a fair estimation. Our approach obtains the best Jaccard index and F1-score in most cases and also has the best mean value. The proposed method obtains the best mean value in both SBD and TCGA datasets. Figure 1 depicts an example using a TCGA image.

3.2 Effect of Blood in the Estimation of the Color-Vector Matrix

All the images in SBD were deconvolved and the obtained color-vector matrix $\hat{\mathbf{M}}$ for H&E was compared with the expected $\mathbf{M}_{GT}$ using the euclidean distance. Table 3 summarizes the results obtained according to the amount of blood in the image for different values of thr. When the amount of blood in the image is none or small, removing excessive pixels from the image is not beneficial, and therefore the optimal thr is higher. However, in concordance with the previous experiment, a value of $thr = 0.3$ obtains the best result for a realistic range of the amount of blood.

A visual example is depicted in figure 2.a. We represent the color-vector matrix $\mathbf{M}$ as a tuple where each square shows the normalized color vector for each stain, H and E, respectively. The effect of blood is most appreciated in the

Table 3. Mean Euclidean distance between ground truth $\mathbf{M}_{GT}$ and obtained $\mathbf{M}$ color-vector matrix using the proposed method and different threshold values.

Amount of blood	Threshold value 0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.1386	0.1248	0.0927	0.0826	0.0726	0.0775	0.0748	**0.0653**
0.1	0.0790	0.0852	0.0971	**0.0699**	0.0704	0.0699	0.0920	0.0893
0.2	0.0910	0.0768	0.0683	**0.0661**	0.1038	0.0994	0.1391	0.1500
0.3	0.0984	**0.0760**	0.0942	0.0826	0.1020	0.1498	0.1607	0.1720
0.4	**0.0843**	0.0966	0.0934	0.0992	0.1296	0.1489	0.1775	0.1881
0.5	0.1019	**0.0985**	0.1214	0.1264	0.1427	0.1663	0.1814	0.1903
0.6	0.1145	0.1210	**0.1001**	0.1342	0.1558	0.1724	0.1817	0.1953
0.7	0.1228	**0.1016**	0.1229	0.1294	0.1505	0.1842	0.1908	0.2034
0.8	0.1201	**0.1165**	0.1365	0.1391	0.1676	0.1862	0.1949	0.2009
0.9	0.1338	**0.1287**	0.1294	0.1588	0.1710	0.1844	0.1939	0.2110
Mean	0.1084	**0.1026**	0.1056	0.1088	0.1266	0.1439	0.1587	0.1666

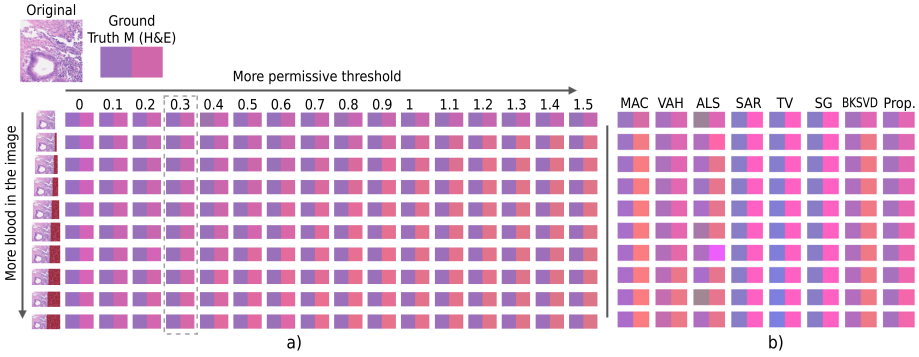


Fig. 2. Qualitative comparison of color-matrices on an SBD image. a) Obtained color-vector matrix $\mathbf{M}$ for different thresholds and amounts of blood in the image. The first row corresponds to the clean image and the amount of blood increases with the row number. The columns fix the threshold value, with zero being the most restrictive. The grey dashed box indicates the column with a closer euclidean distance to the ground truth. b) Obtained $\mathbf{M}$ for different state-of-the-art methods.

E channel, which becomes reddish as the blood increases, and a higher *thr* is chosen. The E channel changes to represent blood rather than eosin, and the H channel shifts to represent the mix of H&E which are closer in color between them than with blood.

Table 4 compares the results of the proposed method with $thr = 0.3$ with state-of-the-art methods in terms of euclidean distance to the ground truth color-vector matrix. Specifically, we compare our method with the methods by Macenko *et al.*[8] (MAC), the robust method by Vahadane *et al.*[19] (VAH), Alsubaie *et al.*[1] (ALS), the reference-prior based Bayesian methods using Simultaneous AutoRegressive (SAR) [5], Total Variation (TV) [13], and Super Gaussian (SG) [14], and the previous non-robust BKSVD [12]. The proposed method outperforms previous methods, showing that the current state-of-the-art for stain separation is not well adapted when there is blood on the image.

Table 4. Mean Euclidean distance between ground truth $\mathbf{M}_{GT}$ and obtained $\mathbf{M}$ color-vector matrix for different BCD methods. The proposed method uses $thr = 0.3$.

Amount of blood	Method MAC	VAH	ALS	SAR	TV	SG	BKSVD	Proposed
0	0.2428	0.1019	0.3059	0.3118	0.3328	0.3119	**0.0838**	0.1248
0.1	0.3004	0.0972	0.3291	0.3118	0.3783	0.3119	0.2427	**0.0852**
0.2	0.3074	0.1116	0.3983	0.3118	0.3889	0.3119	0.2484	**0.0768**
0.3	0.3090	0.1227	0.3824	0.3118	0.3971	0.3119	0.2487	**0.0760**
0.4	0.3099	0.1340	0.4116	0.3118	0.4061	0.3119	0.2515	**0.0966**
0.5	0.3101	0.1398	0.4305	0.3118	0.4141	0.3119	0.2510	**0.0985**
0.6	0.3096	0.1893	0.4272	0.3118	0.4244	0.3119	0.2525	**0.1210**
0.7	0.3091	0.2494	0.4012	0.3118	0.4321	0.3119	0.2489	**0.1016**
0.8	0.3089	0.2710	0.5008	0.3118	0.4417	0.3119	0.2481	**0.1165**
0.9	0.3090	0.3002	0.5228	0.3118	0.4503	0.3119	0.2492	**0.1287**
Mean	0.3016	0.1717	0.4110	0.3118	0.4066	0.3119	0.2325	**0.1026**

In figure 2.b, we present a qualitative comparison of the color-vector matrix obtained by the different methods. Small amounts of blood affect most of the methods. VAH works relatively well when there are small amounts of blood. The Bayesian methods SAR, TV, and SG that use a reference prior are not strongly affected by blood, but the estimated color-vector matrix remains close to the prior. The proposed method solves the limitations of the base BKSVD and is able to correctly estimate the stains in the image even with the presence of a large amount of blood.

4 Conclusions

In this work, we have extended the BKSVD method for stain separation to acknowledge the presence of blood on the images, obtaining a robust estimation of the stains and making it possible to detect blood. This work proposed, for the first time, the use of BCD techniques for blood detection, connecting the fields of color-preprocessing and artifact detection, which benefits both fields. On the one hand, the presence of blood hampers the estimation of the stains in the images and the correct separation of the stains. On the other hand, considering the appearance of the stains provides a framework that facilitates the detection of blood.

The proposed approach has been tested on synthetic and real images containing blood, showing promising performance. The method is able to accurately estimate the color-vector matrix and separate the stains when there is blood in the image, and correctly identify blood pixels.

We believe that our work remarks on an important issue of stain separation techniques that are the basis for the color-preprocessing of histopathological images. In future research, we plan to extend these results and assess how blood impacts the color normalization and augmentation of histological images for cancer classification.

References

1. Alsubaie, N., et al.: Stain deconvolution using statistical analysis of multi-resolution stain colour representation. PLOS ONE **12**, e0169875 (2017)
2. Anghel, A., et al.: A High-Performance System for Robust Stain Normalization of Whole-Slide Images in Histopathology. Front. Med. **6** (sep 2019)
3. Bukenya, F., et al.: An automated method for segmentation and quantification of blood vessels in histology images. Microvas. Res. **128**, 103928 (2020)
4. Chen, Z., et al.: Histological quantitation of brain injury using whole slide imaging: A pilot validation study in mice. PLOS ONE **9**(3), 1–10 (03 2014)
5. Hidalgo-Gavira, N., Mateos, J., Vega, M., Molina, R., Katsaggelos, A.K.: Variational Bayesian blind color deconvolution of histopathological images. IEEE Trans. Image Process. **29**(1), 2026–2036 (2020)
6. Kanwal, N., Pérez-Bueno, F., Schmidt, A., Molina, R., Engan, K.: The devil is in the details: Whole slide image acquisition and processing for artifacts detection, color variation, and data augmentation. a review. IEEE Access pp. 1–1 (2022)
7. Kim, N.T., et al.: An Original Approach for Quantification of Blood Vessels on the Whole Tumour Section. Anal. Cell. Pathol. **25**(2), 63–75 (2003)
8. Macenko, M., et al.: A method for normalizing histology slides for quantitative analysis. In: Int. Symp on biomed Imaging (ISBI). pp. 1107–1110 (2009)
9. Morales, S., Engan, K., Naranjo, V.: Artificial intelligence in computational pathology – challenges and future directions. Digit. Signal Process. p. 103196 (2021)
10. Mosaliganti, K., et al.: An imaging workflow for characterizing phenotypical change in large histological mouse model datasets. J. biomed Inform. **41**(6), 863–873 (2008)
11. Perry, T.S.: Andrew ng x-rays the ai hype. IEEE Spectrum, May 2021
12. Pérez-Bueno, F., Serra, J., Vega, M., Mateos, J., Molina, R., Katsaggelos, A.K.: Bayesian K-SVD for H&E blind color deconvolution. Applications to stain normalization, data augmentation, and cancer classification. Comput. Med. Imaging Graph. (2022)
13. Pérez-Bueno, F., López-Pérez, M., Vega, M., Mateos, J., Naranjo, V., Molina, R., et al.: A TV-based image processing framework for blind color deconvolution and classification of histological images. Digit. Signal Process. **101**, 102727 (2020)
14. Pérez-Bueno, F., Vega, M., Sales, M.A., Aneiros-Fernández, J., Naranjo, V., Molina, R., Katsaggelos, A.K.: Blind color deconvolution, normalization, and classification of histological images using general super gaussian priors and bayesian inference. Comput. Meth. Prog. Bio. **211**, 106453 (2021)
15. Ruifrok, A.C., Johnston, D.A.: Quantification of histochemical staining by color deconvolution. Anal. Quant. Cytol. Histol. **23**, 291–299 (2001)
16. Sertel, O., et al.: Texture classification using nonlinear color quantization: Application to histopathological image analysis. In: 2008 IEEE Int. Conf. on Acoust. Speech Signal Process. pp. 597–600 (2008)
17. Swiderska-Chadaj, Z., et al.: Automatic quantification of vessels in hemorrhoids whole slide images. In: IEEE Int. Conf. Comput. Probl. Elec. Eng. pp. 1–4 (2016)
18. Tellez, D., et al.: Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Med. Image Anal. **58**, 101544 (2019)
19. Vahadane, A., et al.: Structure-preserving color normalization and sparse stain separation for histological images. IEEE Trans. Med. Imag. **35**, 1962–1971 (2016)
20. Wetteland, R., Engan, K., et al.: A multiscale approach for whole-slide image segmentation of five tissue classes in urothelial carcinoma slides. Technol. Cancer Res. Treat. **19** (2020)