

Fully Automatic Blind Color Deconvolution of Histological Images Using Super Gaussians

Fernando Pérez-Bueno
Dpto. C.C.I.A.¹
Univ. de Granada
Granada, Spain
fpb@ugr.es

Miguel Vega
Dpto. L.S.I²
Univ. de Granada
Granada, Spain
mvega@ugr.es

Valery Naranjo
I.I.I en Bioingeniería³
Univ. Poli. de València
Valencia, Spain
vnaranjo@dcom.upv.es

Rafael Molina
Dpto. C.C.I.A.
Univ. de Granada
Granada, Spain
rms@decsai.ugr.es

Aggelos K. Katsaggelos
Dept. E.E.C.S⁴
Northwestern Univ.
Evanston, IL, USA
aggk@eecs.northwestern.edu

Abstract—In digital pathology blind color deconvolution techniques separate multi-stained images into single stained bands. These band images are then used for image analysis and classification purposes. This paper proposes the use of Super Gaussian priors for each stain band together with the similarity to a given reference matrix for the color vectors. Variational inference and an evidence lower bound are then utilized to automatically estimate the latent variables and model parameters. The proposed methodology is tested on real images and compared to classical and state-of-the-art methods for histopathological blind image color deconvolution. Its use as a preprocessing step in prostate cancer classification is also analysed.

Index Terms—Blind color deconvolution, histopathological images, variational Bayes, Super Gaussian

I. INTRODUCTION

Histopathological tissues are frequently stained with a combination of stains that binds to specific proteins on the tissue, being Hematoxylin and Eosin (H&E) the most commonly used combination. Stained slides are then scanned to obtain high resolution Whole-Slide Images (WSI). Since the analysis of these images is very time consuming and requires a lot of effort, computer-aided diagnosis (CAD) systems have become a valuable ally for pathologists. CADs frequently make use of the different stains separately [1]. The separation of the stains in a WSI is known as Color Deconvolution (CD) and estimates each stain concentration at each pixel location.

Usually, the color spectral properties of each stain are also unknown since they vary from image to image. Color variations have a wide range of origins: different scanners, stain manufactures, or staining procedures, among others. They create inter- and intra-laboratory differences. Blind Color Deconvolution (BCD) techniques estimate image specific stain color-vectors together with stain concentrations. In other fields, this problem is referred as blind unmixing [2].

Several CD methods have been proposed (see [3] for a review). Ruifrok *et al.* [4] proposed one of the pioneer works. Non-negative Matrix Factorization (NMF) [1], [5], Singular

Value Decomposition (SVD) [6], [7] and Independent Component Analysis (ICA) [8], [9] have been applied to CD. Deep learning methods have also been recently proposed [10]–[12].

This paper contributes to the field with a general probabilistic framework to BCD. In [13]–[15], a prior on the color-vectors, favouring similarity to some reference stain color-vectors, as well as a smoothness Simultaneous Autoregressive (SAR) prior model on each stain concentrations was used. As the SAR prior tends to oversmooth the edges of the image structures, the use of a Total Variation (TV) prior on each stain was proposed in [16], extended in [17]. Improving the sparsity options of the TV prior, here we propose the use of Super Gaussian (SG) distributions as priors for BCD. SGs include distributions like l_p or log distributions which have proven to be very powerful prior models to induce sparsity in Bayesian Blind Deconvolution problems [18], [19]. They are always associated to energy functions with very interesting sparse properties. Furthermore, their inference procedure is easily carried out, as shown in section III.

The rest of the paper is organized as follows: in section II the problem of BCD is mathematically formulated. Following the Bayesian modelling and inference, in section III we propose a fully Bayesian method for the estimation of the concentrations and the color-vector matrix. In section IV, the proposed method is evaluated and its performance is compared with classical and state-of-the-art CD methods. Finally, section V concludes the work.

II. PROBLEM FORMULATION

Digital brightfield microscopes usually store a stained histological specimen's slide as an RGB color image of size $M \times N$, represented by the $MN \times 3$ matrix, \mathbf{I} . Each color plane is stacked into a $MN \times 1$ column vector $\mathbf{i}_c = (i_{1c}, \dots, i_{MNc})^T$, $c \in \{R, G, B\}$. Each value i_{ic} represents the transmitted light on color band $c \in \{R, G, B\}$ for the pixel i of the slide. However, for stain deconvolution is usual to work in the *Optical Density* (OD) space, where the Beer-Lambert law, for a slide stained with n_s stains, establishes

$$\mathbf{Y}^T = \mathbf{M}\mathbf{C}^T + \mathbf{N}^T, \quad (1)$$

where $\mathbf{Y} \in \mathbb{R}^{MN \times 3}$ is the observed OD image with three OD channels, i.e., $\mathbf{Y} = [\mathbf{y}_R \ \mathbf{y}_G \ \mathbf{y}_B]$ and each channel $\mathbf{y}_c \in$

This work was sponsored in part by Ministerio de Ciencia e Innovación under Contract BES-2017-081584 and project DPI2016-77869-C2-2-R. ¹Ciencias de la Computación e I.A. ²Lenguajes y Sistemas Informáticos. ³Instituto de Investigación e Innovación en Bioingeniería ⁴Electrical Engineering and Computer Science

$\mathbb{R}^{MN \times 1}$ is defined as $\mathbf{y}_c = -\log_{10}(\mathbf{i}_c/\mathbf{i}_c^0)$, where \mathbf{i}_c^0 denotes the incident light, and the division operation and $\log_{10}(\cdot)$ function are computed element-wise. $\mathbf{M} \in \mathbb{R}^{3 \times n_s}$ is the color-vector matrix, $\mathbf{C} \in \mathbb{R}^{MN \times n_s}$ is the stain concentration matrix and $\mathbf{N} \in \mathbb{R}^{MN \times 3}$ is a random matrix with i.i.d. zero mean Gaussian components with variance β^{-1} .

BCD techniques seek to estimate the stain concentration matrix \mathbf{C} where the i -th row $\mathbf{c}_{i,:}^T = (c_{i1}, \dots, c_{in_s})$, $i = 1, \dots, MN$, the contribution of each stain to the i -th \mathbf{Y} pixel value, $\mathbf{y}_{i,:}$, and the s -th column $\mathbf{c}_s = (c_{1s}, \dots, c_{MN s})^T$, $s \in \{1, \dots, n_s\}$, the concentrations of the s -th stain. BCD techniques also estimate the specific color of each stain, represented by the color-vector matrix $\mathbf{M} \in \mathbb{R}^{3 \times n_s}$ where each column \mathbf{m}_s in matrix \mathbf{M} is a unit ℓ_2 -norm stain color-vector containing the relative RGB color composition of the corresponding stain in the OD space.

In the following section we use Bayesian modeling and inference to estimate both \mathbf{C} and \mathbf{M} .

III. BAYESIAN MODELLING AND INFERENCE

Following the degradation model in (1), we have

$$p(\mathbf{Y}|\mathbf{C}, \mathbf{M}, \beta) = \prod_{i=1}^{MN} \mathcal{N}(\mathbf{y}_{i,:} | \mathbf{M}\mathbf{c}_{i,:}, \beta^{-1}\mathbf{I}_{3 \times 3}). \quad (2)$$

Bayesian methods start with a prior distribution on the unknowns. In this paper we adopt as priors SG distributions for the stain concentration

$$p(\mathbf{C}|\alpha) = \prod_{\nu=1}^J \prod_{s=1}^{n_s} \prod_{i=1}^{MN} Z(\alpha_{\nu s}) \exp[-\alpha_{\nu s} \rho(c_{\nu s}(i))], \quad (3)$$

with $\alpha_{\nu s} > 0$. In (3) $\mathbf{c}_{\nu s} = \mathbf{D}_{\nu} \mathbf{c}_s$, where $\{\mathbf{D}_{\nu}\}_{\nu=1}^J$ is a set of J high-pass and $Z(\alpha_{\nu s})$ is a partition function. For $p(\mathbf{c}_{\nu s}|\alpha_{\nu s})$ in (3) to be SG $\rho(\cdot)$ has to be symmetric around zero and the function $\rho(\sqrt{s})$ increasing and concave for $s \in (0, \infty)$. This condition is equivalent to $\rho'(s)/s$ being decreasing on $(0, \infty)$, and allows ρ to be represented as

$$\rho(c_{\nu s}(i)) = \inf_{\eta_{\nu s}(i) > 0} L(c_{\nu s}(i), \eta_{\nu s}(i)) \quad (4)$$

where $L(c_{\nu s}(i), \eta_{\nu s}(i)) = \frac{1}{2} \eta_{\nu s}(i) c_{\nu s}^2(i) - \rho^*(\frac{1}{2} \eta_{\nu s}(i))$, \inf denotes infimum and $\rho^*(\cdot)$ is the concave conjugate of $\rho(\cdot)$ and $\eta_{\nu s} = \{\eta_{\nu s}(i)\}_{i=1}^{MN}$ are positive parameters. The relationship dual to (4) is given by [20]

$$\rho^*\left(\frac{1}{2} \eta_{\nu s}(i)\right) = \inf_{c_{\nu s}(i)} \frac{1}{2} \eta_{\nu s}(i) c_{\nu s}^2(i) - \rho(c_{\nu s}(i)). \quad (5)$$

Table I shows some penalty functions, corresponding to SG distributions (see [18] for additional energy functions associated to SG distributions).

For the unknown color-vector matrix $\mathbf{M} = [\mathbf{m}_1, \dots, \mathbf{m}_{n_s}]$ we incorporate similarity to a reference color-vector matrix [2] $\underline{\mathbf{M}} = [\underline{\mathbf{m}}_1, \dots, \underline{\mathbf{m}}_{n_s}]$ into its prior model as

$$p(\mathbf{M}|\gamma) \propto \prod_{s=1}^{n_s} \gamma_s^{\frac{3}{2}} \exp\left(-\frac{1}{2} \gamma_s \|\mathbf{m}_s - \underline{\mathbf{m}}_s\|^2\right), \quad (6)$$

TABLE I
SOME CHOICES FOR THE PENALTY FUNCTION

Label	$\rho(s)$	$\rho'(s)/ s $
$\ell_p, 0 < p \leq 1$	$\frac{1}{p} s ^p$	$ s ^{p-2}$
log	$\log(\epsilon + s)$	$(\epsilon + s)^{-1} s ^{-1}$

where γ_s , $s = 1, \dots, n_s$, controls our confidence on the accuracy of \mathbf{m}_s .

Then, we define the joint probability distribution as

$$p(\mathbf{Y}, \mathbf{C}, \mathbf{M}, \beta, \alpha, \gamma) = p(\mathbf{M}|\gamma)p(\gamma)p(\mathbf{Y}|\mathbf{C}, \mathbf{M}, \beta)p(\beta) \times \prod_{\nu=1}^J \prod_{s=1}^{n_s} p(\mathbf{c}_{\nu s}|\alpha_{\nu s})p(\alpha_{\nu s}), \quad (7)$$

where constant hyperpriors $p(\gamma)$, $p(\beta)$ and $p(\alpha_{\nu s})$ on the model hyperparameters have been included.

Following the Bayesian paradigm, inference will be based on the posterior distribution $p(\Theta|\mathbf{Y})$ with $\Theta = \{\mathbf{Y}, \mathbf{C}, \mathbf{M}, \beta, \alpha, \gamma\}$ the set of all unknowns. In this paper we use the mean-field variational Bayesian model [21] to approximate $p(\Theta|\mathbf{Y})$ by the distribution $q(\Theta)$ of the form $q(\Theta) = \prod_{s=1}^{n_s} q(\mathbf{m}_s) \prod_{\nu=1}^J q(\mathbf{c}_{\nu s})$ that minimizes the Kullback-Leibler (KL) divergence [22] defined as

$$\mathbf{KL}(q(\Theta) || p(\Theta|\mathbf{Y})) = \int q(\Theta) \log \frac{q(\Theta)}{p(\Theta, \mathbf{Y})} d\Theta + \log p(\mathbf{Y}). \quad (8)$$

Even with this factorization, the SG prior for \mathbf{C}_{ν} hampers the evaluation of this divergence, but the quadratic bound for ρ allows us to bound the prior in (3) with a Gaussian form

$$p(c_{\nu s}(i)|\alpha_{\nu s}) \geq Z(\alpha_{\nu s}) \exp[-\alpha_{\nu s} L(c_{\nu s}(i), \eta_{\nu s}(i))], \quad (9)$$

$\forall \eta_{\nu s}(i) > 0$. We then define

$$\mathcal{M}_{\nu}(\mathbf{C}, \eta_{\nu}|\alpha_{\nu}) = \prod_{s=1}^{n_s} \prod_{i=1}^{MN} Z(\alpha_{\nu s}) \exp[-\alpha_{\nu s} L(c_{\nu s}(i), \eta_{\nu s}(i))] \quad (10)$$

$$F(\Theta, \mathbf{Y}) = p(\mathbf{M}, \gamma)p(\gamma)p(\mathbf{Y}|\mathbf{C}, \mathbf{M}, \beta)p(\beta) \times \prod_{\nu} \mathcal{M}_{\nu}(\mathbf{C}, \eta_{\nu}|\alpha_{\nu})p(\alpha_{\nu}), \quad (11)$$

to obtain the inequality $\log p(\Theta, \mathbf{Y}) \geq \log F(\Theta, \mathbf{Y})$.

Utilizing the lower bound $F(\Theta, \mathbf{Y})$ for the joint probability distribution in (8) we minimize $\mathbf{KL}(q(\Theta) || F(\Theta, \mathbf{Y}))$ instead of $\mathbf{KL}(q(\Theta) || p(\Theta|\mathbf{Y}))$.

As shown in [21], for each unknown $\theta \in \Theta$, $q(\theta)$ will have the form

$$q(\theta) \propto \exp\langle \log F(\Theta, \mathbf{Y}) \rangle_{q(\Theta \setminus \theta)}, \quad (12)$$

where $\Theta \setminus \theta$ represents all the variables in Θ except θ and $\langle \cdot \rangle_{q(\Theta \setminus \theta)}$ denotes the expected value calculated using the distribution $q(\Theta \setminus \theta)$. When point estimates are required $\hat{\theta} = \langle \theta \rangle_{q(\theta)}$ is used.

1) *Concentration Update:* Defining

$$\begin{aligned} \mathbf{e}_{i,:}^{-s} &= \mathbf{y}_{i,:} - \sum_{k \neq s} \langle c_{ik} \rangle \langle \mathbf{m}_k \rangle \\ z_i^{-s} &= \langle \mathbf{m}_s \rangle^T \mathbf{e}_{i,:}^{-s}, \quad i = 1, \dots, MN, \end{aligned} \quad (13)$$

and using eq. (12) we can easily show that $q(\mathbf{c}_s) = \mathcal{N}(\mathbf{c}_s | \langle \mathbf{c}_s \rangle, \Sigma_{\mathbf{c}_s})$, where

$$\begin{aligned} \Sigma_{\mathbf{c}_s}^{-1} &= \beta \langle \|\mathbf{m}_s\|^2 \rangle \mathbf{I}_{MN \times MN} + \sum_{\nu} \alpha_{\nu s} \mathbf{D}_{\nu}^T \text{diag}(\boldsymbol{\eta}_{\nu s}) \mathbf{D}_{\nu} \\ \langle \mathbf{c}_s \rangle &= \beta \Sigma_{\mathbf{c}_s} \mathbf{z}^{-s}. \end{aligned} \quad (14)$$

2) *Color-Vector Update:* In a similar way, using (13), we can show that $q(\mathbf{m}_s) = \mathcal{N}(\mathbf{m}_s | \langle \mathbf{m}_s \rangle, \Sigma_{\mathbf{m}_s})$, where

$$\begin{aligned} \Sigma_{\mathbf{m}_s}^{-1} &= \left(\sum_{\nu=1}^H \beta_{\nu} \sum_{i=1}^{MN} \langle c_{\nu i s}^2 \rangle + \gamma_s \right) \mathbf{I}_{3 \times 3}, \\ \langle \mathbf{m}_s \rangle &= \Sigma_{\mathbf{m}_s} \left(\sum_{\nu=1}^H \beta_{\nu} \sum_{i=1}^{MN} \langle c_{\nu i s} \rangle \mathbf{e}_{\nu i,:}^{-s} + \gamma_s \underline{\mathbf{m}}_s \right). \end{aligned} \quad (15)$$

Notice that $\langle \mathbf{m}_s \rangle$ may not be a unitary vector even if $\underline{\mathbf{m}}_s$ is. We can always replace $\langle \mathbf{m}_s \rangle$ by $\langle \mathbf{m}_s \rangle / \|\langle \mathbf{m}_s \rangle\|$ and $\Sigma_{\mathbf{m}_s}$ by $\Sigma_{\mathbf{m}_s} / \|\langle \mathbf{m}_s \rangle\|^2$.

3) *Variational Parameter Update:* To estimate the $\boldsymbol{\eta}$ matrix, we need to solve, for each $s \in \{1, \dots, n_s\}$, $\nu \in \{1, \dots, L\}$ and $i \in \{1, \dots, MN\}$

$$\begin{aligned} \hat{\eta}_{\nu s}(i) &= \arg \min_{\eta_{\nu s}(i)} \langle J(c_{\nu s}(i), \eta_{\nu s}(i)) \rangle_{q(\mathbf{c}_s)} \\ &= \arg \min_{\eta_{\nu s}(i)} \frac{1}{2} \eta_{\nu s}(i) u_{\nu s}^2(i) - \rho^* \left(\frac{1}{2} \eta_{\nu s}(i) \right) \end{aligned} \quad (16)$$

where $u_{\nu s}(i) = \sqrt{\langle c_{\nu s}^2(i) \rangle}$. Since

$$\rho^* \left(\frac{\hat{\eta}_{\nu s}(i)}{2} \right) = \min_x \frac{1}{2} \hat{\eta}_{\nu s}(i) x^2 - \rho(x) \quad (17)$$

whose minimum is achieved at $x = u_{\nu s}(i)$, we have, differentiating the right hand side of (17) with respect to x ,

$$\hat{\eta}_{\nu s}(i) = \rho'(u_{\nu s}(i)) / u_{\nu s}(i). \quad (18)$$

4) *Parameter Update:* The estimates of the noise and color-vectors parameters are obtained as

$$\hat{\beta}^{-1} = \frac{\text{tr} \langle (\mathbf{Y}^T - \mathbf{M}\mathbf{C}^T)(\mathbf{Y}^T - \mathbf{M}\mathbf{C}^T)^T \rangle_{q(\Theta)}}{3MN}, \quad (19)$$

$$\hat{\gamma}_s^{-1} = \frac{\text{tr} \langle (\mathbf{m}_s - \underline{\mathbf{m}}_s)(\mathbf{m}_s - \underline{\mathbf{m}}_s)^T \rangle}{3}. \quad (20)$$

From (12) we obtain the following distribution for $\alpha_{\nu s}$

$$q(\alpha_{\nu s}) = \text{const} + \sum_{i=1}^{MN} \log Z(\alpha_{\nu s}) \exp[-\alpha_{\nu s} \rho(c_{\nu s}(i))]. \quad (21)$$

Using the mode of this distribution as the $\alpha_{\nu s}$ estimate results in the following equation for $\hat{\alpha}_{\nu s}$

$$\frac{\partial Z(\hat{\alpha}_{\nu s})}{\partial \hat{\alpha}_{\nu s}} = \frac{\text{tr} \langle (\mathbf{D}_{\nu}^T \mathbf{D}_{\nu}) \langle \mathbf{c}_s \mathbf{c}_s^T \rangle \rangle}{MN}. \quad (22)$$

Algorithm 1 Variational Bayesian SG Blind CD

Require: Observed image \mathbf{I} and reference (prior) color-vector matrix $\underline{\mathbf{M}}$.

Obtain the OD image \mathbf{Y} from \mathbf{I} and set $\langle \mathbf{m}_s \rangle^{(0)} = \underline{\mathbf{m}}_s$, $\Sigma_{\mathbf{m}_s}^{(0)} = \mathbf{0}$, $\Sigma_{\mathbf{c}_s}^{(0)} = \mathbf{0}$, $\langle \mathbf{c}_s \rangle^{(0)}$, $\forall s = 1, \dots, n_s$, from the matrix \mathbf{C} obtained as $\mathbf{C}^T = \underline{\mathbf{M}}^+ \mathbf{Y}^T$, with $\underline{\mathbf{M}}^+$ the Moore-Penrose pseudo-inverse of $\underline{\mathbf{M}}$, and $n = 0$.

while convergence criterion is not met **do**

1. Set $n = n + 1$.

2. Obtain $\beta^{(n)}$, $\gamma_s^{(n)}$ and $\alpha_{\nu s}^{(n)}$ from (19), (20) and (22).

3. Using $\langle \mathbf{c}_s \rangle^{(n-1)}$ and $\Sigma_{\mathbf{c}_s}^{(n-1)} \forall s$, update the new variational parameters $\hat{\boldsymbol{\eta}}_{\nu s}^{(n)}$ from (18) $\forall \nu$.

4. Using $\langle \mathbf{c}_s \rangle^{(n-1)}$, $\Sigma_{\mathbf{c}_s}^{(n-1)}$ and $\langle \mathbf{m}_s \rangle^{(n-1)}$ update the color-vectors $\Sigma_{\mathbf{m}_s}^{(n)}$ and $\langle \mathbf{m}_s \rangle^{(n)}$ from (15), $\forall s$.

5. Using $\langle \mathbf{m}_s \rangle^{(n)}$, $\Sigma_{\mathbf{m}_s}^{(n)}$ and $\hat{\boldsymbol{\eta}}_{\nu s}^{(n)} \forall \nu$ update the concentrations $\Sigma_{\mathbf{c}_s}^{(n)}$ and $\langle \mathbf{c}_s \rangle^{(n)}$ from (14), $\forall s$.

end while

Output color-vector $\hat{\mathbf{m}}_s = \langle \mathbf{m}_s \rangle^{(n)}$ and $\hat{\mathbf{c}}_s = \langle \mathbf{c}_s \rangle^{(n)}$.

The ℓ_p penalty function shown in Table I produces proper priors, for which the partition function can be evaluated, but the log penalty function produces an improper prior. We tackle this problem examining, for $\alpha_{\nu s}$, the behavior of

$$Z(\alpha_{\nu s}, K)^{-1} = \int_{-K}^K \exp[-\alpha_{\nu s} \rho(t)] dt \quad (23)$$

and keeping in $\partial Z(\alpha_{\nu s}) / \partial \alpha_{\nu s}$ the term that depends on $\alpha_{\nu s}$. This produces for the log prior

$$\frac{\partial Z(\hat{\alpha}_{\nu s})}{\partial \hat{\alpha}_{\nu s}} = (\hat{\alpha}_{\nu s} - 1)^{-1}. \quad (24)$$

5) *Calculating the concentration covariance matrices:* The matrix $\Sigma_{\mathbf{c}_s}$ must be explicitly calculated to find its trace and also $\hat{\eta}_{\nu s}(i)$. However, since its calculation is very intense, we propose the following approximation of the covariance matrix. We first approximate $\text{diag}(\boldsymbol{\eta}_{\nu s})$ using

$$\text{diag}(\boldsymbol{\eta}_{\nu s}) \approx z(\boldsymbol{\eta}_{\nu s}) \mathbf{I}, \quad (25)$$

where $z(\boldsymbol{\eta}_{\nu s})$ is calculated as the mean of the values in the diagonal. We then use the approximation

$$\Sigma_{\mathbf{c}_s}^{-1} \approx \beta \langle \|\mathbf{m}_s\|^2 \rangle \mathbf{I}_{MN \times MN} + \sum_{\nu} \alpha_{\nu s} z(\boldsymbol{\eta}_{\nu s}) \mathbf{D}_{\nu}^T \mathbf{D}_{\nu} = \mathbf{B}.$$

Finally we have $\langle c_{\nu s}^2(i) \rangle \approx (\langle c_{\nu s}(i) \rangle)^2 + \frac{1}{MN} \text{tr}[\mathbf{B}^{-1} \mathbf{D}_{\nu}^T \mathbf{D}_{\nu}]$.

6) *Proposed Algorithm:* Based on the previous derivations, we propose the Variational Bayesian SG BCD in Algorithm 1. The linear equations problem in (14), used in step 4 of Alg. 1, has been solved using the Conjugate Gradient approach.

IV. EXPERIMENTAL RESULTS

We compare the proposed approach with classical and state-of-the-art CD methods on two different goals that blind color deconvolution methods should achieve: image reconstruction and image classification. For comparison, two SG priors have been utilized, log and ℓ_1 . First order horizontal and vertical

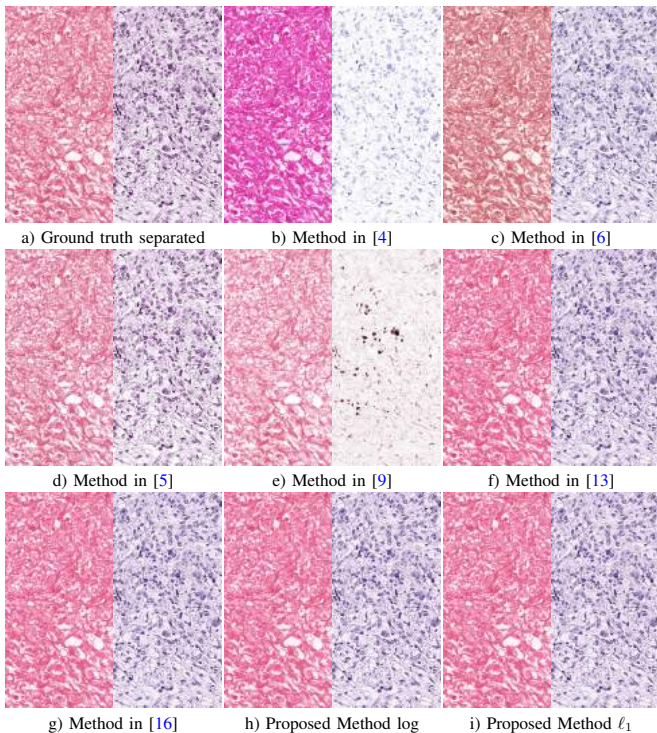


Fig. 1. Breast RGB ground truth separated E-only and H-only images from the WSSB dataset in [9], and results for the competing and proposed methods. Eosin separation is presented in the left hand side of each image and hematoxylin separation in the right hand side.

differences have been used as filters in (3), so $J = 2$. We compare our models to the methods in [4]–[6], [9], [13], [16].

Further details of each experiment are provided below.

1) *Color Deconvolution Experiments*: The first experiment was carried out on the Warwick Stain Separation Benchmark (WSSB) in [9]. WSSB includes 24 H&E stained images of breast, colon and lung tissues 2000×2000 whose ground truth stain color-vector matrices, \mathbf{M}_{GT} , were manually selected based on biological structures. The ground truth concentrations were obtained using $\mathbf{C}_{GT}^T = \mathbf{M}_{GT}^+ \mathbf{Y}^T$. A breast ground truth RGB separation is depicted in Fig. 1a. The method was run until the criterion $\|\langle \mathbf{c}_s \rangle^{(n)} - \langle \mathbf{c}_s \rangle^{(n-1)}\|^2 / \|\langle \mathbf{c}_s \rangle^{(n)}\|^2 < 10^{-3}$ was met by both stains, that is, $s = 1, 2$. Since different tissues may have different color characteristics, the reference color-vector matrix \mathbf{M} was obtained by selecting, by non-medical experts, a single pixel containing mainly hematoxylin and a pixel containing mainly eosin from each type of tissue.

The resulting H-only and E-only images were compared both visually and numerically by means of the Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) metrics. Numerical results, presented in Table II, show that the proposed ℓ_1 method produces higher PSNR and SSIM values than the competing models. The proposed log method produces similar results to the method in [16]. The separated images from the observed image in Fig. 1a are shown in Fig. 1c–j. The proposed method and the methods in [5], [13], [16] produce colors very similar to the ground truth separation in Fig. 1b although the new method and the one in [16]

produce sharper images than the method in [13] and richer details than the method in [5].

The obtained results clearly show an advantage when using the ℓ_1 method, meaning that this prior captures the real behaviour of the dyes on the tissue. As we have already indicated, each stain fixes itself only and completely to certain proteins on the tissue, making the stain concentration differences at neighbouring pixels sparse [5]. However, the experiments show that the differences are not as sparse as expected. The ℓ_1 prior, with a lower kurtosis than log prior, allows to keep more non-zero values. This makes the ℓ_1 prior a good choice, as it induces sparsity in a softer way than the log prior.

2) *Prostate Cancer Classification Experiments*: Finally, the main objective of BCD methods for histopathology is to improve the performance of classification methods. In this section we study how the use of different stain deconvolution methods affects the performance of classifiers. We use the SICAPv1 database, a prostate cancer histopathological database recently presented in [23], that contains 60 WSI (17 benign and 43 pathological). In order to capture complete glands, the images were downsampled to 10x scale and those in the training set were divided into patches of size 1024×1024 pixels with a 50% overlap. The dataset was color deconvolved using the proposed and competing methods. The H&E concentration image in the OD space was used to extract features to be utilized as input to the classifiers. Following [23], we used the concatenation of Local Binary Patterns Variance [24] and Geodesic granulometries features [23]. A set of shallow and deep classifiers were trained with those descriptors and their results were compared. We used Random Forest (RF) [25], Extreme Gradient Boosting (XgBoost) [26], Gaussian Processes (GP) [27] and Deep Gaussian Processes (DGP) [28]. For each classifier, a five-fold cross-validation was applied to compare its performance with each deconvolution method. Further details regarding feature extraction and classifier configuration can be found in [23]. The proposed method, is configured using $n_s = 3$ stains in order to remove the background from the hematoxylin and eosin channels.

The results in III show that the proposed method outperformed almost all the other methods. Only [4], the most extended method for histopathology color deconvolution obtain similar results. However, performance of [4] in reconstruction was the lowest. The SG approach allows us to lead in classification performance without losing fidelity to the image.

V. CONCLUSIONS

This work presented the use of Super Gaussian priors for Blind Color Deconvolution (BCD) automatically estimating the model parameters and examining its use in classification problems. The use of the SG family induces sparsity on the differences of stain concentrations at neighbouring pixels. This is a desired quality during the staining process and a theoretical feature of the stained tissue. The variety of SG distributions available allowed us to explore a range of possible sparse solutions within a common inference, using an easy procedure. This work explored two penalty functions, ℓ_1 and log. On

TABLE II
PSNR AND SSIM FOR THE DIFFERENT METHODS ON THE WSSB DATASET [9].

Image	Stain	Ruifrok's Method [4]		Macenko's Method [6]		Vahadane's Method [5]		Alsubaie's Method [9]		Hidalgo-Gavira's Method [13]		Vega's Method [16]		Proposed log prior		Proposed ℓ_1 prior	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Colon	H	22.27	0.8141	23.91	0.8095	25.83	0.8851	21.11	0.7241	28.57	0.9542	28.62	0.9544	28.66	0.9531	29.01	0.9638
	E	20.70	0.7456	21.55	0.6365	26.29	0.8904	21.94	0.8540	27.58	0.9139	27.60	0.9161	27.74	0.9212	28.38	0.9414
Breast	H	15.27	0.6215	26.24	0.9552	25.46	0.9239	24.60	0.8068	28.81	0.9528	29.14	0.9560	29.23	0.9464	30.50	0.9751
	E	17.66	0.7644	23.62	0.9336	27.68	0.9550	25.92	0.9380	26.60	0.9464	26.76	0.9492	26.74	0.9444	27.71	0.9645
Lung	H	22.47	0.7987	19.52	0.7389	25.87	0.8912	20.62	0.5551	32.91	0.9763	33.10	0.9757	31.21	0.9415	35.21	0.9898
	E	22.05	0.7734	18.09	0.5088	25.53	0.8195	23.95	0.8939	30.77	0.9306	31.02	0.9353	29.99	0.9338	33.07	0.9654
Mean	H	20.00	0.7448	23.22	0.8345	25.72	0.9100	22.11	0.6953	30.10	0.9611	30.29	0.9621	29.70	0.9470	31.57	0.9762
	E	20.14	0.7611	21.08	0.6930	26.50	0.8883	23.94	0.8953	28.32	0.9303	28.46	0.9336	28.16	0.9331	29.72	0.9571

TABLE III
PERFORMANCE OF THE PROPOSED AND COMPETING DECONVOLUTION METHODS WITH DIFFERENT CLASSIFIERS.

Method	RF	GP	XgBoost	DGP
Ruifrok [4]	0.9789	0.9855	0.9764	0.9737
Macenko [6]	0.9315	0.9535	0.9425	0.8802
Vahadane [5]	0.9222	0.9479	0.9295	0.9420
Alsubaie [9]	0.9262	0.9442	0.9246	0.9344
Hidalgo-Gavira [13]	0.9157	0.9542	0.9228	0.8997
Vega [16]	0.9242	0.9498	0.9294	0.9249
Prop. ℓ_1 prior	0.9796	0.9842	0.9796	0.9729
Prop. log prior	0.9796	0.9842	0.9798	0.9723

image classification, both SG priors obtained a high accuracy with all the classifiers. Our method far exceed most of the competing method and even outperformed the most extended method for histopathology color deconvolution [4]. Regarding image reconstruction, we found that the ℓ_1 prior, which has a more moderate peak, captured better the real distribution of the stains in the images. The sharper edges than the obtained in [13] and the softer sparsity induced than [16] takes us to a BCD method that far outperformed all the methods compared with.

REFERENCES

- [1] A. Rabinovich, S. Agarwal, *et al.*, "Unsupervised color decomposition of histologically stained tissue samples," in *NIPS*, pp. 667–674, 2004.
- [2] M. Arngren, M. Schmidt, and J. Larsen, "Unmixing of hyperspectral images using bayesian non-negative matrix factorization with volume prior," *Signal Processing Systems*, vol. 65, pp. 479–496, 12 2011.
- [3] S. Roy, A. K. Jain, *et al.*, "A study about color normalization methods for histopathology images," *Micron*, vol. 114, pp. 42–61, 2018.
- [4] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Analytical and quantitative cytology and histology*, vol. 23, pp. 291–299, 2001.
- [5] A. Vahadane, T. Peng, *et al.*, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Trans. on Medical Imaging*, vol. 35, pp. 1962–1971, 2016.
- [6] M. Macenko, M. Niethammer, *et al.*, "A method for normalizing histology slides for quantitative analysis," in *ISBI*, pp. 1107–1110, 2009.
- [7] M. T. McCann, J. Majumdar, *et al.*, "Algorithm and benchmark dataset for stain separation in histology images," in *ICIP*, pp. 3953–3957, 2014.
- [8] N. Alsubaie, S. E. A. Raza, and N. Rajpoot, "Stain deconvolution of histology images via independent component analysis in the wavelet domain," in *ISBI*, pp. 803–806, 2016.
- [9] N. Alsubaie, N. Trahearn, *et al.*, "Stain deconvolution using statistical analysis of multi-resolution stain colour representation," *PLOS ONE*, vol. 12, p. e0169875, 2017.
- [10] F. G. Zanjani, S. Zinger, B. E. Bejnordi, J. A. W. M. v. d. Laak, and P. H. N. d. With, "Stain normalization of histopathology images using generative adversarial networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 573–577, Apr. 2018.
- [11] M. T. Shaban, C. Baur, N. Navab, and S. Albarqouni, "Staining: Stain style transfer for digital histological images," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 953–956, IEEE, 2019.
- [12] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, J. Shi, and C. Xue, "Adaptive color deconvolution for histological WSI normalization," *Computer Methods and Programs in Biomedicine*, vol. 170, pp. 107–120, Mar. 2019.
- [13] N. Hidalgo-Gavira, J. Mateos, M. Vega, R. Molina, and A. K. Katsaggelos, "Variational bayesian blind color deconvolution of histopathological images," *IEEE Transactions on Image Processing*, vol. 29, no. 1, p. 2026–2036, 2020.
- [14] N. Hidalgo-Gavira, J. Mateos, *et al.*, "Blind color deconvolution of histopathological images using a variational Bayesian approach," in *ICIP*, pp. 983–987, 2018.
- [15] N. Hidalgo-Gavira, J. Mateos, *et al.*, "Fully automated blind color deconvolution of histopathological images," in *MICCAI*, pp. 183–191, 2018.
- [16] M. Vega, J. Mateos, R. Molina, and A. K. Katsaggelos, "Variational bayes color deconvolution with a total variation prior," in *2019 27th European Signal Processing Conference (EUSIPCO)*, pp. 1–5, 2019.
- [17] F. Pérez-Bueno, M. López-Pérez, M. Vega, J. Mateos, V. Naranjo, R. Molina, and A. K. Katsaggelos, "A tv-based image processing framework for blind color deconvolution and classification of histological images," *Digital Signal Processing*, vol. 101, p. 102727, 2020.
- [18] S. D. Babacan, R. Molina, M. N. Do, and A. K. Katsaggelos, "Blind deconvolution with general sparse image priors," in *ECCV*, 2012.
- [19] X. Zhou, M. Vega, F. Zhou, R. Molina, and A. K. Katsaggelos, "Fast bayesian blind deconvolution with huber super gaussian priors," *Digital Signal Processing*, vol. 60, pp. 122–133, 2017.
- [20] R. Rockafellar, *Convex analysis*. Princeton University Press, 1996.
- [21] C. Bishop, *Pattern Recognition and Machine Learning*, pp. 454–455. Springer, 2006.
- [22] S. Kullback, *Information Theory and Statistics*. Dover Pub., 1959.
- [23] A. E. Esteban, M. Lopez-Perez, A. Colomer, M. A. Sales, R. Molina, and V. Naranjo, "A new optical density granulometry-based descriptor for the classification of prostate histological images using shallow and deep Gaussian processes," *Computer Methods and Programs in Biomedicine*, vol. 178, pp. 303–317, 2019.
- [24] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern Recognition*, vol. 43, no. 3, pp. 706–719, 2010.
- [25] M. Valkonen, K. Kartasalo, K. Liimatainen, M. Nykter, L. Latonen, and P. Ruusuvoori, "Metastasis detection from whole slide images using local features and random forests," *Cytometry Part A*, vol. 91, no. 6, pp. 555–565, 2017.
- [26] A. Pimkin, G. Makarchuk, V. Kondratenko, M. Pisov, E. Krivov, and M. Belyaev, "Ensembling neural networks for digital pathology images classification and segmentation," *LNCS*, vol. 10882, pp. 877–886, 2018.
- [27] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2006.
- [28] A. Damianou and N. Lawrence, "Deep Gaussian processes," *Journal of Machine Learning Research*, vol. 31, pp. 207–215, 2013.