A Nonlinear Mapping Approach to Stain Normalization in Digital Histopathology Images Using Image-Specific Color Deconvolution

Adnan Mujahid Khan*, Nasir Rajpoot, *Senior Member, IEEE*, Darren Treanor, and Derek Magee

Abstract—Histopathology diagnosis is based on visual examination of the morphology of histological sections under a microscope. With the increasing popularity of digital slide scanners, decision support systems based on the analysis of digital pathology images are in high demand. However, computerized decision support systems are fraught with problems that stem from color variations in tissue appearance due to variation in tissue preparation, variation in stain reactivity from different manufacturers/batches, user or protocol variation, and the use of scanners from different manufacturers. In this paper, we present a novel approach to stain normalization in histopathology images. The method is based on nonlinear mapping of a source image to a target image using a representation derived from color deconvolution. Color deconvolution is a method to obtain stain concentration values when the stain matrix, describing how the color is affected by the stain concentration, is given. Rather than relying on standard stain matrices, which may be inappropriate for a given image, we propose the use of a color-based classifier that incorporates a novel stain color descriptor to calculate image-specific stain matrix. In order to demonstrate the efficacy of the proposed stain matrix estimation and stain normalization methods, they are applied to the problem of tumor segmentation in breast histopathology images. The experimental results suggest that the paradigm of color normalization, as a preprocessing step, can significantly help histological image analysis algorithms to demonstrate stable performance which is insensitive to imaging conditions in general and scanner variations in particular.

Index Terms—Histopathology images analysis, nonlinear mapping, principal color histograms (PCH), stain color descriptor (SCD), stain estimation, stain normalization.

I. INTRODUCTION

HISTOPATHOLOGY is the diagnosis of disease by visual examination of tissue under the microscope. In order to examine tissue sections (which are virtually transparent), tissue sections are prepared using colored histochemical stains that bind selectively to cellular components. Color variation is a problem in histopathology based on light microscopy due to a range of factors such as the use of different scanners, variable chemical coloring/reactivity from different manufacturers/batches of stains, coloring being dependent on staining procedure (timing, concentrations, etc.), and light transmission being a function of section thickness. Lyon *et al.* [1] outline the need for standardization of reagents and procedures in histological practice. However, because of issues like manual sectioning variability and stains fading over time, complete standardization is not possible to achieve with the current technology. Current practice is limited to physical and procedural quality-control methods, including subjective assessment of stain quality and interlaboratory comparisons of staining, in order to minimize the visible variability in staining and its impact on diagnostic quality.

With the advent of digital imaging and automatic image analysis, color variation in histopathology has become more of an issue. For example, many commercial image analysis algorithms require parameters defining the expected color of anatomy of interest and fail if these parameters are incorrect. Although methods have been proposed for improving color constancy in images formed via Lambertion (reflective) model of image formation (see [2] for a good overview[1]), these methods are not applicable to color images formed via light transmission through a tissue specimen, and thus are inappropriate for histopathology image analysis.

Consequently, a large number of methods presented in the area of automatic image analysis of color histopathology images bypass the problem of color constancy by transforming the images to grayscale. For example, texture analysis for tissue type classification has been performed on grayscale images using features based on grayscale cooccurrence matrices [3], local binary patterns [4], or the wavelet packet transform [5]. This can be successful in cases where grayscale intensity is the primary cue. For example, Basavanhally *et al.* [6] use the fact that cell nuclei are much darker under certain stains than surrounding anatomy. Luminance is used to classify different types of nuclei in their work. However, conversion to grayscale ignores the

---

[1]Please refer to the electronic version of the manuscript for color figures.
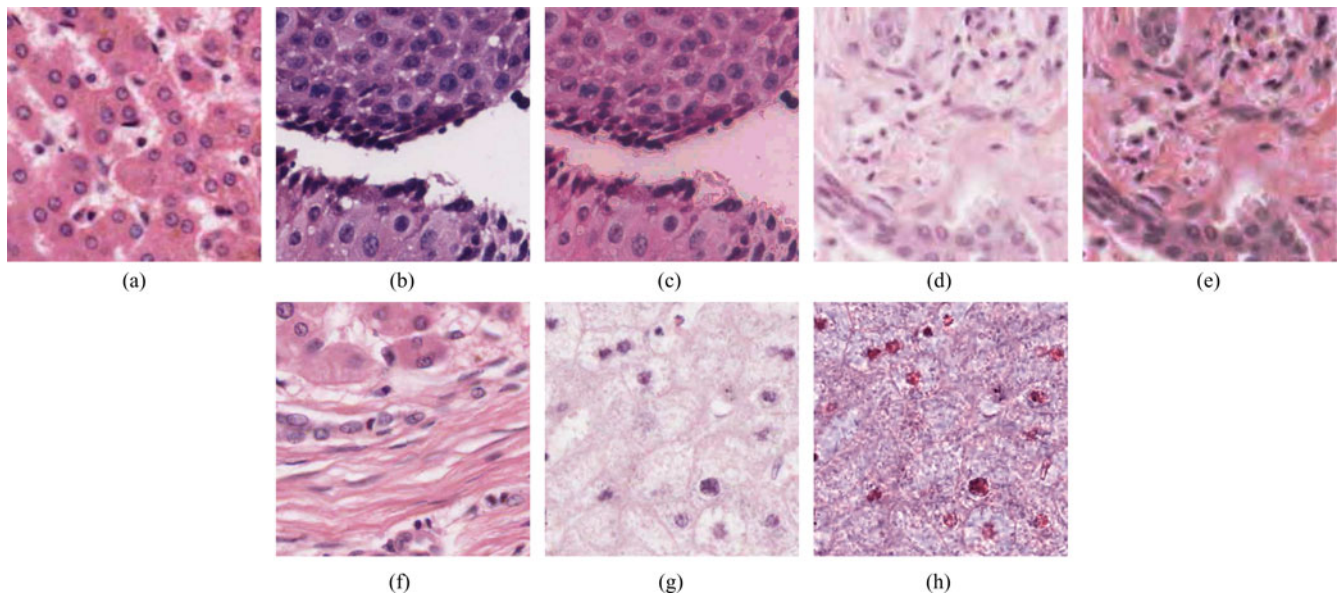
Fig. 1.   Color normalization artifacts when using Reinhard *et al.* [7], histogram specification [8], and Macenko *et al.* [9] color normalization methods. (a) Target image, (b) source Image-1, (c) Reinhard *et al.* normalization output (artifacts at pixel class boundaries, bgd appearing stained), (d) source Image-2, (e) histogram specification normalization output [color mismatch due to different proportions of stains in normalized and target image, (f) target image, and (g) source Image-1, (h) Macenko *et al.* normalization output (noise/artifacts)] .

wealth of information in the color representation used routinely by pathologists. Typically, two or three different colored stains are used to highlight cellular and subcellular target components. The intensity of each color is related to the concentration of the corresponding component. Additionally, more than one target component protein may be present in a given area, resulting in a mix of colors. Converting images to grayscale results in an image representing the total concentration of all tissue components, rather than the relative amounts of each.

Some authors have included color information within texture-based image classification in digital histopathology image analysis [10]. Kong *et al.* [11] use cooccurrence matrices in individual channels of the Lab colorspace as texture descriptor, and evaluate a range of different classifiers for grading neuroblastic differentiation. Sertel *et al.* [12] cluster color vectors in the Lab colorspace using *k*-means clustering and use a cooccurrence representation based on color prototypes as a texture feature. Considering the variation in colors within/across histopathology sections, color texture features may be highly sensitive to staining/scanner variations and thus may significantly affect the performance of an automated system.

In order to overcome these limitations, Wang *et al.* [13] take a different approach and *normalize* color distributions of *source image* to those of a *target image* by using [7] before performing color-based segmentation. In the remainder of this paper, we use the term "*stain or color normalization*" to refer to the process of adjusting the color values of an image on a pixel-by-pixel basis so as to match the color distribution of the source image to that of a target image.

In the literature, a few existing stain normalization methods can be found [7]–[9], [14]–[16]. Histogram specification [8] is a method closely related to histogram equalization previously used for color normalization in oral histopathology images [17].

A major drawback of histogram-based approaches is that they introduce considerable visual artifacts in images. This is due to the implicit assumption that the proportion of pixels of each stain type is same in the target and source images. This is clearly not always correct (see Fig. 1). Kothari *et al.* [14] proposed a variation on histogram normalization, where the presence of a color, rather than frequency is used for color normalization. This has the disadvantage that rare (potentially noise), and common, pixel values are treated as equally important.

Reinhard *et al.* [7] proposed a method of color normalization where the mean and standard deviation of each channel of the image are matched to that of the target by means of a set of linear transforms in Lab colorspace. However, the assumption of unimodal distribution of pixels in each channel of Lab colorspace does not hold if multiple colored stains are used. As a result, this can result in background (bgd) areas being mapped as colored regions, and foreground being incorrectly mapped, as shown in Fig. 1. Magee *et al.* [15] proposed an automatic segmentation extension to [7]. First, Gaussian-mixture model-based color segmentation is used to automatically identify multiple pixel classes, then linear normalization is applied separately to each pixel class, where class membership is defined by a pixel being colored by a particular chemical stain, or bgd. A major limitation of this approach is that it introduces artifacts near pixels that lie on the class boundary.

Color deconvolution (CD) [7] is used extensively in histopathology image analysis for separating an RGB image into (up to) three channels, each corresponding to the actual colors of the stains used (see Section II for details). Magee *et al.* [15] and Macenko *et al.* [9] simultaneously proposed methods for stain normalization based on a CD-derived representation. Both methods automatically derive image-specific CD matrices. Magee *et al.* use a supervised pixel classification-based

approach to estimate stain colors, whereas Macenko *et al.* use an singular value decomposition (SVD)-based approach to directly estimate the matrices. Niethammer *et al.* [16] extend the stain matrix estimation method in [9] using priors to estimate stain matrices to improve stability in cases where images contain uneven proportions of each stain, at the cost of abandoning the closed-form solution in the original work—thus introducing an additional local optima failure mode. Macenko *et al.* use linear per-channel normalization based on a pseudomaximum (the 99th percentile) to map source image values to match the target image, whereas Magee *et al.* use a nonlinear mapping based on pixel classifications. Either method can fail if the stain matrix estimation process fails, the mapping function is inappropriate, or the channel statistics calculations are inaccurate due to excessive noise (e.g., saturated pixels). It can be argued that linear normalization is always inappropriate as it treats optically and chemically saturated pixels identically to other pixels, modifying their values (see Fig. 1). Additionally, Macenko *et al.* modifies the color distribution of both source and target images, which is sometimes not desirable if we have a reference image with stain characteristics suitable for an automated system.

The method presented in this study is an evolution of [15] and overcomes limitations of previous work by estimating stable stain matrices using an image-specific color descriptor and a robust color classification framework based on a variety of training data for a particular stain. Moreover, we propose a regularized nonlinear mapping of stain channels which ensures smooth color transformation without introducing visual artifacts. The following list accounts for our novel contributions:

1) We introduce a novel whole-image stain color descriptor (SCD) that grossly quantifies the concentration of stains in an image. We demonstrate that pixel classification performance is robust for a wide variety of images if our SCD descriptor is used along with R,G,B pixel information.
2) We propose the use of a color-based classifier to calculate image-specific stain matrices.
3) We perform nonlinear mapping of source image channels to the target image channels using regularized spline-based functions estimated from image statistics.
4) We demonstrate that a tumor segmentation algorithm for breast histopathology images [18] demonstrates stable performance if preceded by a stain normalization step especially if data are coming from different scanners.

The remainder of this paper is organized as follows. Section II outlines the details of CD model which is essential to our proposed stain normalization framework. In Section III, we propose a novel method for automatic derivation of stain matrices by incorporating global image-specific stain information with local RGB pixel information in a supervised classification framework [19]. We use the estimated stain matrix to develop a novel stain normalization method that automatically adjusts the RGB color distribution of a source image to that of a target image. Finally, we demonstrate in Section IV that stain normalization can play a critical role in stability of automatic histopathology image analysis algorithms especially when there is variation in the staining protocol or tissue, or the data come from different scanners.

II. CD MODEL

In 2001, Ruifrok and Johnston [20] proposed a CD framework with potential application in histopathology image analysis. This method has been used in a variety of applications: quantification of immunohistochemical (IHC) stains [21] and nuclei detection [22] to name but a few. The CD framework transforms the RGB colorspace Ψ to a new colorspace $\hat{\Psi}$ defined by the stains used for staining the tissue section. If image $\mathcal{I} = (\mathcal{C}, \Psi)$ is defined as a two-dimensional (2D) set of pixels $\mathcal{C}$ with associated colorspace function Ψ assigning red, green, and blue intensities to each pixel, the relationship between colorspaces Ψ and $\hat{\Psi}$ is defined by Lambert–Beers law as follows:

$$\Psi = \exp(-\mathbf{S}\hat{\Psi}) \tag{1}$$

where $\mathbf{S}$ is the stain matrix that defines the stain vectors (absorption factors) associated with each stain used on the tissue,

$$\mathbf{S} = \begin{bmatrix} \bar{s}_{r,1} & \bar{s}_{g,1} & \bar{s}_{b,1} \\ \bar{s}_{r,2} & \bar{s}_{g,2} & \bar{s}_{b,2} \\ \bar{s}_{r,3} & \bar{s}_{g,3} & \bar{s}_{b,3} \end{bmatrix} \tag{2}$$

where $\bar{s}_{r,1}$, $\bar{s}_{g,1}$, and $\bar{s}_{b,1}$ denote the predefined, normalized red, green, and blue values for s_1 channel. Similarly, the second and third rows of $\mathbf{S}$ are defined for s_2 and s_3 channels, respectively. Ruifrok and Johnston [20] provided a closed-form solution to the inversion of (1), in which they demonstrated that the intensity of a pixel $c \in \mathcal{C}$ in the new colorspace $\hat{\Psi}$ is defined as

$$\hat{\Psi}(c) = \mathbf{D}\Phi(c) \tag{3}$$

$$\text{where } \mathbf{D} = \mathbf{S}^{-1} \tag{4}$$

$$\text{and } \Phi(c) = -\log(\Psi(c)). \tag{5}$$

Here, $\mathbf{D}$ is the CD matrix obtained by calculating the inverse of the stain matrix $\mathbf{S}$, Φ is the so-called optical density (OD) space where a linear combination of stains results in a linear combination of OD values, and $\hat{\Psi}(c)$ represents the amount of each stain (s_1, s_2, and s_3) corresponding to the pixel c.

CD requires accurate estimation of $\mathbf{S}$ (the stain matrix). Although Ruifrok and Johnston [20] provide standard stain matrices for a variety of stain combinations, image-specific stain matrix is usually more optimal for stain separation and is recommended to be estimated from control tissue stained with single stains by a laborious manual process.

III. STAIN NORMALIZATION ALGORITHM

The proposed algorithm (see Fig. 2) consists of four modules: stain matrix estimation, CD, nonlinear mapping of channel statistics, and reconstruction. Broadly, we first map both target $\Im(\mathcal{C}, \Psi)$ and source $\mathcal{X}(\mathcal{C}, \Psi)$ images to a representation $\hat{\Im}(\mathcal{C}, \hat{\Psi})$ and $\hat{\mathcal{X}}(\mathcal{C}, \hat{\Psi})$, where each channel relates to a separate chemical stain (see Section II). Next, we apply a nonlinear correction (*mapping*) to normalize each channel of $\hat{\mathcal{X}}$ separately (based on the statistics calculated from the corresponding channel of $\hat{\Im}$). Finally, we *reconstruct* the normalized source image $\mathcal{X}^{\mathrm{norm}}$ using the normalized stain channels of $\hat{\mathcal{X}}$. The following sections present the details of each of these modules.
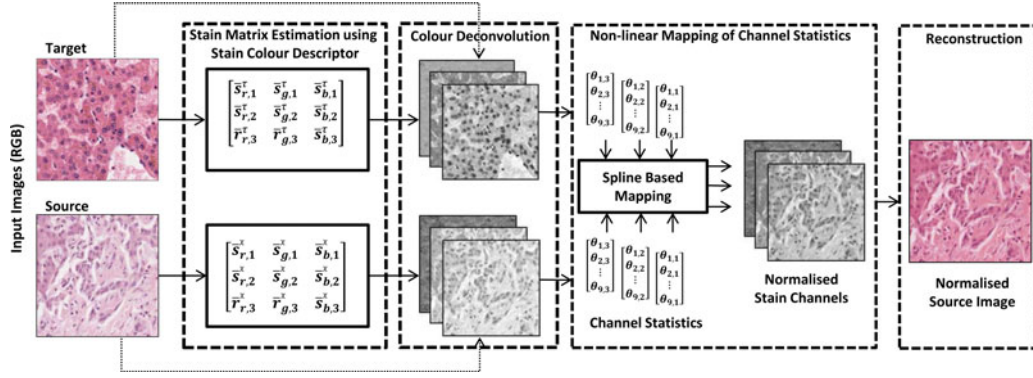
Fig. 2.   Overview of the proposed stain normalization algorithm: 1) Deconvolution of both target and source images to constituent stain channels by estimating image-specific stain matrix (see Fig. 3); 2) Nonlinear mapping of the statistics of each channel of deconvolved source image to those of the statistics of the corresponding channel in deconvolved target image (see Fig. 7); 3) Channel recombination to obtain the normalized source image. Note that $\theta_{m,n}$ represents the mth statistic associated with nth deconvolved channel of corresponding source or target image, where $m \in \{1, 2, \ldots, 9\}$ and $n \in \{1, 2, 3\}$.
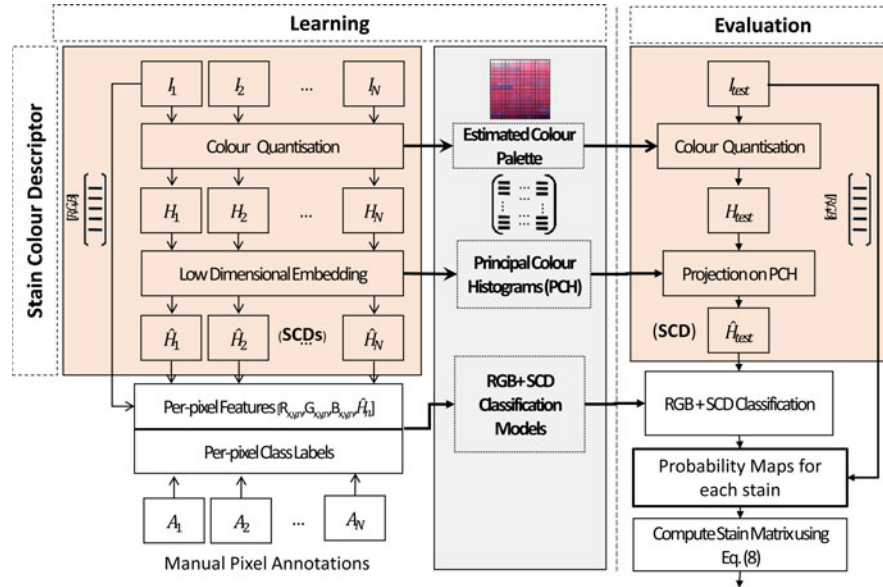

Fig. 3.   Stain matrix estimation using color classification: During learning, color palette for quantization, PCHs for low-dimensional embedding, and classification models for pixel-level color classification are learned from training images. During evaluation, the input image is quantized using the prelearned color palette, SCD is obtained by projecting the quantized image histogram on the precalculated PCH, and pixel classification is performed using the pretrained classification models that generate a probability map for each stain. Probability maps are used to define mean colors for each stain.

A. Stain Matrix Generation and CD

We use the CD framework to convert both $\Im$ and $\mathcal{X}$ images from RGB colorspace to a new colorspace defined by constituent stains. This requires estimation of image-specific stain matrix $\mathbf{S}$ each for $\Im$ and $\mathcal{X}$ images. We estimate image-specific stain matrices ($\mathbf{S}^{\Im}$ and $\mathbf{S}^{\mathcal{X}}$) using a global (perimage) SCD (see Section III-A1) and local pixel-level color information in a supervised color classification framework (see Section III-A2). Fig. 3 gives an overview of the stain matrix estimation method that consists of two phases: *learning* and *evaluation*. Learning, which is performed offline, essentially involves two steps: 1) deriving principal color histograms (PCH) from a training set of quantized image histograms to obtain SCDs; 2) learning classification models by utilizing RGB (pixel information) and SCD (whole-image color information) in a supervised classification framework to generate stain-specific probability maps

(one for each stain and bgd). These probability maps are used to estimate the color of each stain for a particular image. In the following sections, we further elaborate on this method.

1) Image-Specific SCD (SCD): Given a training set of k RGB histopathology images $I = \{I_1, I_2, \ldots, I_k\}$, we calculate a set of image-associated SCDs $\hat{H} = \{\hat{H}_1, \hat{H}_2, \ldots, \hat{H}_k\}$. For this purpose, each image I_i (where $i = \{1, \ldots, k\}$) is color quantized using Oct-tree quantization [23] to generate a set of histograms of 255 color prototypes $H = \{H_1, H_2, \ldots, H_k\}$. Oct-tree quantization [23] is a very efficient color to prototype mapping algorithm. The method works by iteratively partitioning a 3-D colorspace into eight equal-sized regions to form a tree of regions of increasingly small size. The number of leaves can be reduced by subsuming multiple leaves of tree by their common parent node (which then becomes a leaf). In our implementation, subsumption is based on the node with the fewest associated pixels until there are only 255 prototypes.
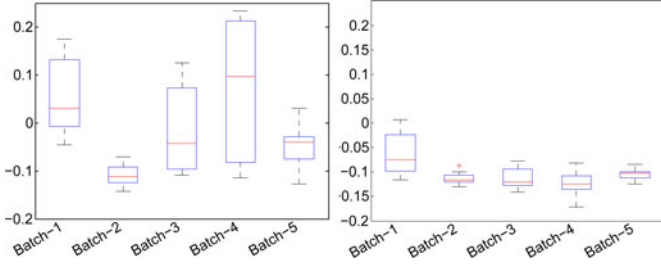
Fig. 4. Boxplots represent the range of SCDs before (*left*) and after (*right*) stain normalization on five batches of training images (used in experiment 1). Note that the SCDs are dispersed for batches of unnormalized images compared to the SCDs of normalized images.

Next, mean $\bar{H}$ and covariance Σ_h of the training set of histograms are computed to perform linear dimensionality reduction. Let E_h^r be the matrix whose columns are the first r eigenvectors of Σ_h, where the eigenvectors stand for the principal components of the training color prototype histograms and are orthonormal to each other. Low-dimensional projection of each training prototype histogram H_i into r-dimensional truncated eigenspace is performed using

$$\hat{H}_i = E_h^r(H_i - \bar{H}) \tag{6}$$

where $\hat{H}_i$ is the r-dimensional embedding of the training set color histogram H_i.

This low-dimensional representation of color prototypes is what we call PCHs. Once the PCH are computed, we project each quantized image histogram on the PCH to compute the SCD, which is a very compact, efficient, and resolution-invariant representation of color distribution in a histopathology image.

Fig. 4 demonstrates the idea of SCD as global image-specific color descriptor for RGB histopathology images, grossly describing the stain color of each image. SCDs are calculated for five batches of 12 images each, before and after stain normalization. It is worth noting that there exists significant variation in SCDs within a single batch (e.g., batch 1, 3, and 4) and between the different batches before stain normalization. It can also be observed from this figure that the SCDs are relatively more homogenous within a single batch and across different batches after stain normalization.

2) Color Classification: In order to generate the stain matrix $\mathbf{S}$, we exploit pixel-level (local) RGB information present in training image set $I = \{I_1, I_2, \ldots, I_k\}$ and image-level SCDs $\hat{H} = \{\hat{H}_1, \hat{H}_2, \ldots, \hat{H}_k\}$, generated for each image in I using the procedure outlined in Section III-A1, to learn pixel classes belonging to s_1, s_2, and bgd in a supervised pixel classification-based learning framework. For supervised classification, the computationally efficient relevance vector machine (RVM) method [24] is selected as it is a sparse kernel machine that results in a model with several orders of magnitude fewer support vectors (and thus runtime) than the related support vector machine method [25]. Random forests [26] (of various sizes) were also evaluated, but their performance was inferior as they tended to overfit the data. Additionally, the RVM provides a probabilistic (rather than binary) output. As we have a three-class problem, and RVM is a two-class classifier, classification is implemented using the one-against-all approach.

TABLE I
RULES FOR CATEGORIZING AN IMAGE INTO THREE CLASSES (SEE SECTION III-B1 FOR DETAILS)

if	$P(bgd|\mathcal{F}) > T_{bgd}$	$\zeta(c) =$	BACKGROUND
elseif	$P(s_n|\mathcal{F}) > T_{fgd}$	$\zeta(c) =$	STAINED
else		$\zeta(c) =$	OTHER

Since most probabilities are close to either 0 or 1 (see Fig. 5), 0.75 for T_{bgd} and T_{fgd} encompasses any pixels that are remotely near $p = 1$.

Given a feature vector $\mathcal{F} = [R, G, B, \hat{H}]$ (generated by concatenating RGB pixels values with corresponding image-specific SCD $\hat{H}$) and pixel-level class labels A, we perform supervised machine learning to produce a probabilistic output that provides the probability of association of each pixel to a particular class. Probabilistic output of each RVM for each pixel is computed as follows:

$$P(s_n|\mathcal{F}) = \frac{P_{s_n}(s_n|\mathcal{F})}{P_{s_1}(s_1|\mathcal{F}) + P_{s_2}(s_2|\mathcal{F}) + P_{bgd}(bgd|\mathcal{F})} \tag{7}$$

where $s_n \in \{s_1, s_2, bgd\}$, $P_{s_n}(s_n|\mathcal{F})$ is the probabilistic output of the RVM model.

$P(s_n|\mathcal{F})$ is used to define the pixels of interest for each stain. Pixels assigned a probability greater than some threshold T_p are used to calculate the exemplar color for each stain. We use the mean colors $(\bar{r}_n, \bar{g}_n, \bar{b}_n)$ of pixels classified as belonging to each class to define stain color [as shown in (8)]. If only two stains are used, the cross product of $[\bar{s}_{r,1}, \bar{s}_{g,1}, \bar{s}_{b,1}]$ and $[\bar{s}_{r,2}, \bar{s}_{g,2}, \bar{s}_{b,2}]$ is used to complete the 3×3 $\mathbf{S}$ matrix,

$$\begin{bmatrix} \bar{s}_{r,n} \\ \bar{s}_{g,n} \\ \bar{s}_{b,n} \end{bmatrix} = \begin{bmatrix} -\log(\bar{r}_n + 1)/256 \\ -\log(\bar{g}_n + 1)/256 \\ -\log(\bar{b}_n + 1)/256 \end{bmatrix} \Bigg/ \begin{vmatrix} -\log((\bar{r}_n + 1)/256 \\ -\log(\bar{g}_n + 1)/256 \\ -\log(\bar{b}_n + 1)/256 \end{vmatrix}. \tag{8}$$

Fig. 5 shows examples of the probability maps produced by the stain matrix estimation method (data taken from [19]). It can be seen that if no SCD [$r = 0$ in (6)] is used (see Fig. 5, columns 2 and 5), significant probability is assigned to each class for weakly stained pixels and bgd pixels with very minor staining. These are the pixels where classification is context specific (i.e., it depends whether the overall staining is strong or weak). We showed in [19] that there was no statistical improvement (or degradation) in classification accuracy when using more than one dimension for the SCD.

B. Nonlinear Mapping of Channel Statistics

For each channel of deconvolved target and source images ($\hat{\Im}$ and $\hat{\mathcal{X}}$), we calculate a set of statistics (see Section III-B1) and smoothly map the statistics of each source image channel to those of the statistics of corresponding target image channel using a spline-based nonlinear mapping (see Section III-B2). Following two sections outline the details of these steps:

1) Computed Statistics of Deconvolved Channels: Given the probability map for each stain and bgd, OD intensities in each deconvolved channel are divided into three classes ζ (where $\zeta \in$ [STAINED, BACKGROUND, OTHER]) using the rules stated in Table I. For each channel of $\Im$ and $\mathcal{X}$,
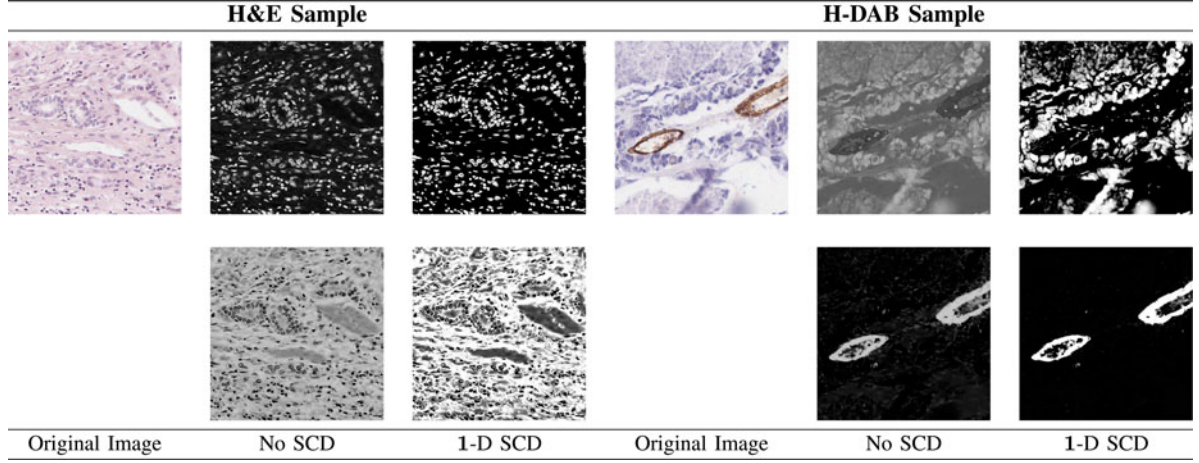
Fig. 5. Probability maps P(s_n) for typical H&E (columns 2 and 3) and H&DAB (columns 5 and 6)-stained images using classifiers with no SCD [$r = 0$ in (6)] (columns 2 and 5) and 1-D SCD [$r = 1$ in (6)] (columns 3 and 6). Top row: Haematoxylin (blue/purple). Bottom: Eosin (columns 2 and 3—pink)/DAB(columns 5 and 6—brown). (White shows high probability of belonging to the stain channel and vice versa.)
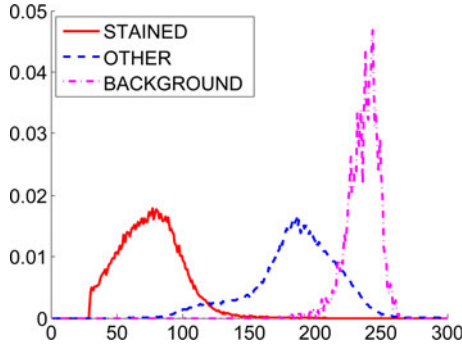

Fig. 6. Marginal distributions of three classes (STAINED, OTHER, BACKGROUND) in ζ from a single channel of a deconvolved image. Optically saturated pixels and chemically saturated pixels are excluded from these distributions to ensure percentiles are representative of the width of the distribution.
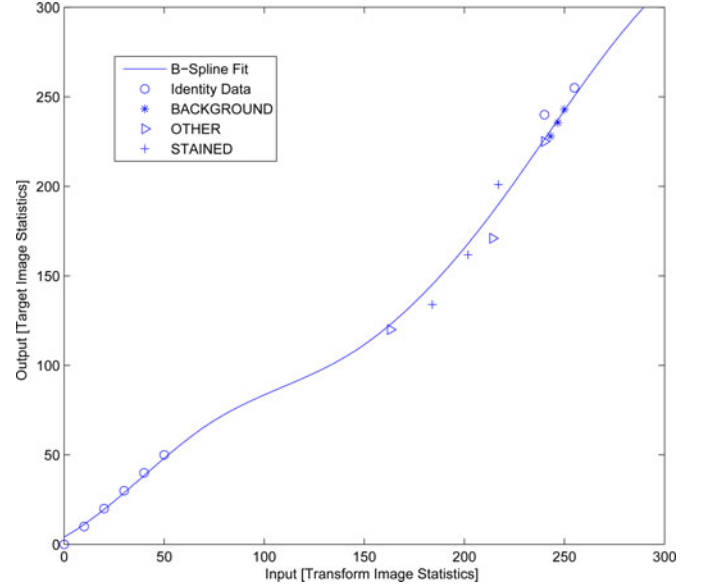

Fig. 7. Estimation of B-Spline mapping function from image statistics. Such a mapping is estimated and applied to each stain channel. Here, identity data is introduced at the extremes of the representation ([0, 10, 20, 30, 40, 50, 240, 250, 500]) to ensure black (chemically saturated) and white (optically saturated) pixels remain unchanged. Please note that in our implementation of [20], we clip the pixel intensities of deconvolved image channels in [0, 500] range.

three statistics (mean, 5th percentile, and 95th percentile) are computed for each class in ζ, producing a vector of length 9 (i.e., 3 statistics × 3 classes). The three statistics are intuitively chosen to represent the distribution associated with each class in a compact and comparable manner (see Fig. 6). Optically saturated (white) and chemically saturated pixels (black) are excluded from this process in all channels to make the percentiles more representative of the width of the distribution.

2) Nonlinear Mapping of Deconvolved Channels: The principle behind the transform function is to map the statistics of the source image channel $\hat{\mathcal{X}}_n$ to those of the statistics of the corresponding channel in the target image $\hat{\mathfrak{S}}_n$. A B-spline is used to ensure a smooth mapping function. The spline parameters are estimated from the nine input–output pairs of values plus identity pairs at the extremes of the representation to ensure black (chemically saturated) and white (optically saturated) pixels remain unchanged. The B-spline parameters (knot values) are estimated by solving a linear system using Tikhonov regularization [27] with an identity mapping prior. This process is illustrated in Fig. 7.

C. Reconstruction

Once each of the stain channels of $\hat{\mathcal{X}}$ is normalized independently, they are recombined on a per-pixel basis as follows:

$$\begin{bmatrix} X_r^{\mathrm{norm}}(c) \\ X_g^{\mathrm{norm}}(c) \\ X_b^{\mathrm{norm}}(c) \end{bmatrix} = \begin{bmatrix} 255 \times \prod_{n=1}^{3} e^{-\hat{X}_n^{\mathrm{norm}}(c) s_{r,n}^{\Im}} \\ 255 \times \prod_{n=1}^{3} e^{-\hat{X}_n^{\mathrm{norm}}(c) s_{g,n}^{\Im}} \\ 255 \times \prod_{n=1}^{3} e^{-\hat{X}_n^{\mathrm{norm}}(c) s_{b,n}^{\Im}} \end{bmatrix} \quad (9)$$
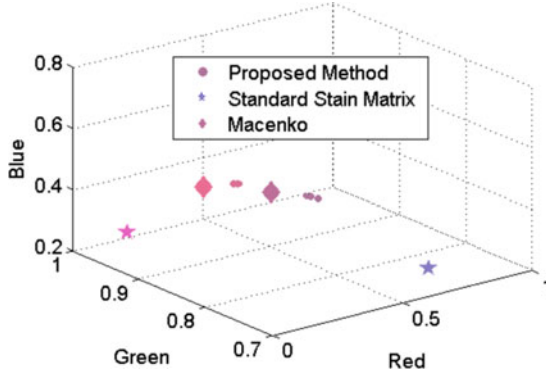
Fig. 8. Stain vectors (for two stains H&E) estimated for a tissue specimen using seven different values of T_p (0.7,0.75,0.8,0.85,0.9,0.95,0.99) represented with increasingly big size of circle. The color of each symbol corresponds to what would be produced by the stain vector. The stars represent the standard stain vectors [20]. Diamonds represent stain vectors estimated using [9].

where $c \in \mathcal{C}$ refers to a pixel on a 2-D grid $\mathcal{C}$, $\hat{\mathcal{X}}_n^{\text{norm}}$ refers to the normalized stain channel n, and $s_{\alpha,n}^{\Im}$ (where $\alpha \in \{r,g,b\}$) is the stain vector associated with nth channel of stain matrix $\mathbf{S}^{\Im}$.

IV. VALIDATION EXPERIMENTS

Two sets of experiments were performed to evaluate the utility of our proposed method. In the first experiment, an RVM classifier is trained on RGB data from each target image and tested on images from different tissue batches that are stain normalized using variety of stain normalization methods. The main aim of this experiment is to determine the color consistency of images with the target before and after stain normalization. In the second experiment, we demonstrate that stain normalization can improve the performance of histopathology image analysis algorithms, especially when data come from different scanners. In addition to the proposed method, a range of color-normalization approaches is evaluated within these experiments.

Model Parameter Selection: We build PCH and classification models from manually annotated training data which consists of five batches (four liver and one oesophagus) of 12 images each, with about 1000–3000 pixels. So, the total number of pixels used for training is about $48 \times 1500 \times 3 = 220K$ pixels. Training images are manually annotated by DM and verified by an expert pathologist (DT) and care is taken to ensure that only those pixels are marked which actually belong to the stain in question. Weakly stained pixels are deliberately avoided to keep the training data as clean as possible.

Using cross-validation experiments, we observed a significant increase ($\approx 16\%$) in classification accuracy if 1-D SCD is used along with RGB pixel information. It was further observed that the results do not show any statistically significant improvement in classification accuracy if 2-D or 3-D SCDs are used (see [19] for further validation). Therefore, in all of our experiments, we used 1-D SCD.

Fig. 8 demonstrates the sensitivity of the threshold value T_p used to estimate stain matrix. Stain vectors (for two stains H&E) are estimated and plotted for a tissue specimen using different values of T_p. Notice that all the recovered stain vectors

form tight clusters which lead us to conclude that the proposed method is not sensitive to the value of T_p. Moreover, it can also be observed that the stain vectors estimated using [9] and the proposed method are relatively closer to one another than the standard stain vectors. In all of our experiments, we used $T_p = 0.99$.

T_{bgd}, T_{fgd}, and parameters for excluding optically and chemically saturated pixels were determined to be relatively insensitive to threshold values (within sensible ranges) by cross validation on the training set. In all of our experiments, we used the value of 0.75 for both parameters.

A. Experiment 1: Per-Pixel Classification Following Color Normalization

Dataset: The dataset for this experiment consists of five batches of 12 images each (60 images, ≈ 0.5 Million labeled pixels). Pixel classes (s_1, s_2, and bgd) are annotated for approximately 1000–3000 pixels of each class in each image using an interactive tool. Four batches contain liver tissues, with the fifth containing oesophageal tissue. These batches were prepared at different times using different chemical batches by a range of technicians within our laboratories. All tissues are formalin fixed, paraffin embedded, H&E counterstained, and scanned at $20/40\times$ magnification using an Aperio (AP) XT scanner.

Design: For each fold of a five-fold cross validation, two RVM-based classifiers were generated (as described in Section III-A2). One—used for color normalization—was trained on four-folds of training data (48 images) with a 1-D SCD ($r = 1$). The other RVM was used to test for color consistency in the normalized images and was thus only trained on a single target image with no SCD. Test images were classified by the second RVM classifier and the results were compared to the ground truth (GT) annotations to determine the color consistency of normalized images with the target image. Images were color normalized to the target image using six different methods: NN (no stain normalization), ER [7], HS [8], SK [14], MM [9], and our proposed stain normalization method. Classification accuracy is determined by assigning each classified pixel to the class with the highest probability. The kernel bandwidth parameter for RVM classifier (0.005 in all experiments) is determined by cross validation on a small training set consisting of one image from each batch.

Results and Discussion: Statistical summaries of these experiments are shown in Table II. It can be seen that color normalization improves overall classification performance for the H&E stained data in general (except when using the SK method). This indicates pixels of corresponding stain types are more similar to the target images after normalization than before. The best performing method is the proposed method. Results without SCD are not presented as the method fails to estimate CD matrices in approximately 19% of cases. This is because of insufficient number of pixels of high probability for each stain class.

B. Experiment 2: Stain Normalization as a Preprocessing Step in Histopathology Image Analysis

Segmentation of areas containing tumor cells in standard H&E histopathology images of breast (and several other tissues)

TABLE II
INTERBATCH NORMALIZATION RESULTS: FIVE-FOLD CROSS VALIDATION
ACCURACY STATISTICS FOR THE RVM CLASSIFIER TRAINED ON A SINGLE
IMAGE AND TESTED ON 48 IMAGES FROM DIFFERENT TISSUE BATCHES

Criteria	NN	ER	HS	SK	MM	Proposed
Mean	0.84	0.93	0.88	0.79	0.95	**0.96**
Std. Dev.	0.06	0.01	0.06	0.12	0.01	**0.009**
Min.	0.76	0.91	0.78	0.61	0.92	**0.95**
Max.	0.90	0.95	0.93	0.89	0.97	**0.98**
p-value.	-	0.01	0.06	0.09	0.008	**0.006**

Mean, standard deviation, minimum and maximum of accuracy are
presented, along with the p value from a paired t-test (one sided)
comparing the classification accuracy on normalized versus original
ges. Best results are presented in bold. NN (no normalization); ER [7];
HS [8]; SK [14]; MM [9].

is a key task for detection and quantification of mitotic cells from the standard H&E slides with a high degree of accuracy, without need for special stains [22], [28]. Tumor segmentation is also vital for automated scoring of IHC-stained slides to restrict the scoring or analysis to areas containing tumor cells only and avoid potentially misleading results from analysis of stromal regions. We demonstrate how the performance of a tumor segmentation algorithm [18] varies when it is preceded by different stain normalization methods as a preprocessing step especially when data come from different scanners.

Dataset: We use the public MITOS dataset of breast histology images [29]. The dataset is composed of 50 breast histology high power field (HPF) images from five patients scanned at $40\times$ magnification using two different scanners: AP and Hamamatsu (HM). Since same tissue sample is scanned using two different scanners, there are in total 100 HPF images. In order to account for interobserver variability, all images are hand segmented by two expert pathologists to mark the boundary of tumor areas in each HPF. The average degree of disagreement (interobserver variability) between the two pathologists on GT images is $11.55\% \pm 0.05$. We generate all experimental results on two criteria: 1) considering pathologist-1's markings (AP/HM-1) as GT; 2) considering pathologist-2's markings (AP/HM-2) as GT.

Design: All images in the dataset were color normalized by a range of methods in the same way as experiment 1 using a target image scanned using an AP scanner. The segmentation was performed using [18] and evaluated by comparing against the manual annotations of two expert pathologists using Dice coefficient. The classifier for color normalization using the presented method was trained on the data from experiment 1.

Evaluation and Discussion: Segmentation performance was evaluated using the Dice coefficient; a widely used pixel-wise accuracy measure. Given a segmented image (X) and pathologist's marked GT image (Y), Dice coefficient is defined by $\frac{2|X \cap Y|}{|X|+|Y|}$. The measure provides values between 0.0 to 1.0 (1.0 indicates identical segmentations).

Fig. 9 provides illustration of different stain normalization methods considered in this study. It can be seen that when ER stain normalization is applied (on both AP and HM HPF images), output image quality degrades significantly particularly in areas where we have fat tissue (white regions) and flat stromal

structures of the tissue. For HM HPF images, the effect of stain normalization is almost invisible in the case of ER and SK. With MM, the effect of image normalization is relatively discernible, however there are two problems: 1) the color distribution of normalized image is different from the target image as the method normalized both source and target images; 2) for an HM HPF image, it introduces significant visual artifacts in white regions (see electronic version for better visibility). The reason being the linear transformation function which do not faithfully normalize extremely white and extremely dark pixels. With the proposed method, we obtain visually superior color normalization: staining of both Eosin- and Hematoxylin-rich regions normalized to the target image without introducing visual artifacts.

Fig. 10 presents tumor segmentation performance in terms of the Dice coefficient for different stain normalization methods. The results of AP HPF images as expected are fairly consistent as the algorithm is trained on data from this variation. However, for HM HPF images, the effect of stain normalization using our method can be seen clearly. The proposed algorithm outperforms all other methods for HM HPF images. This result suggests that color normalization could be used to build methods that work across different tissue types scanned with scanners from different manufacturers.

C. Computational Efficiency

To quantitatively evaluate the computational efficiency of the proposed method compared to ER, SK, and MM, we run all of these algorithms on a set of 100 images, and the average time is reported in Table III. All timings are calculated on a 3.1 GHz Windows 7 machine running MATLAB 2013b with 8GB of RAM.

V. DISCUSSION AND FUTURE WORK

Color inconsistency between tissue sections within and between laboratories, or between different scanners is a significant issue in histopathology. We have demonstrated the importance of color consistency in two application areas (per-pixel color-based segmentation and texture-based tumor segmentation), and presented a method for ensuring consistency by color normalization to a target image. The presented method is qualitatively superior to the state of the art in both application areas. Additionally, the presented method results in less image artifacts than existing approaches due to its robustness (at estimating deconvolution vectors) and appropriateness (using a nonlinear transform regularized to identity at the extremes). From both experiments, it is clear that RGB-histogram-based methods do not perform well at this task if source and target images are significantly different in content. The method of Reinhard *et al.*, based on linear normalization in Lab colorspace, is attractive in its simplicity but is based on the false assumption of unimodal color distribution in each channel. This can result in bgd appearing stained after normalization and poor normalization of the least dominant channel(s). CD-based approaches come out the best in both experiments presented, with the proposed method outperforming Macenko's method in both cases. The appropriateness of CD-based approaches should be obvious, as
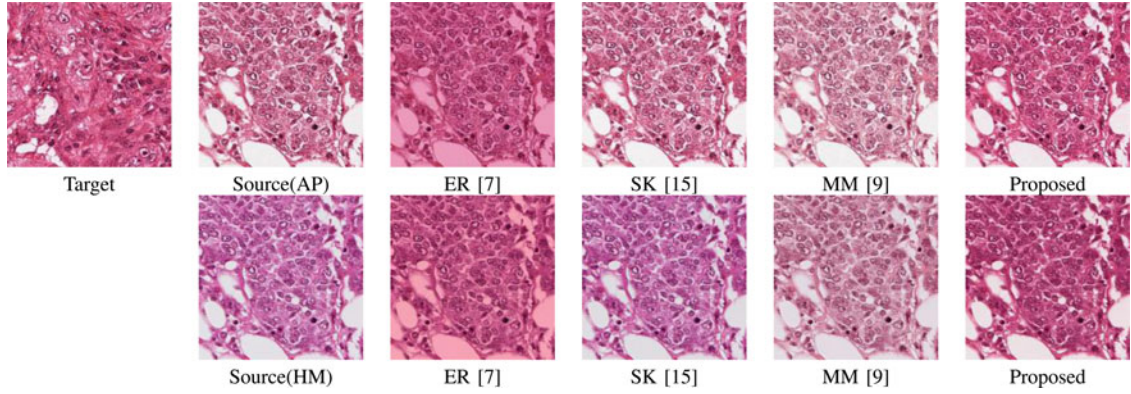
Fig. 9.   Illustration of different stain normalization methods: Top row (left to right): target image, source AP HPF image, results of normalization using ER [7], SK [14], MM [9], and the proposed methods, respectively. Bottom row (left to right): the same HPF scanned using HM scanner, results of normalization using ER, SK, MM, and the proposed methods, respectively. HPF, AP, and HM.
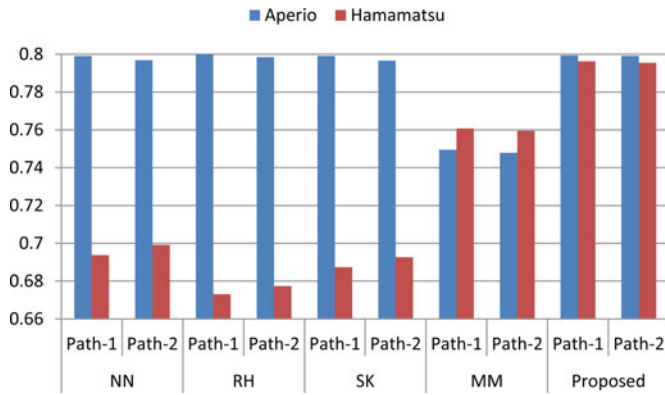

Fig. 10.   Dice coefficient representing the agreement between algorithm's output and pathologist's markings of AP and HM HPF images, when each image is preprocessed using different stain normalization methods. Path-1 and Path-2 refers to the GT marking by pathologist-1 and pathologist-2, respectively; NN; ER Reinhard [7]; SK [14]; MM [9].

TABLE III
RUN TIMES FOR STAIN NORMALIZATION OF IMAGES OF VARIOUS SIZES

Method	256×256	512×512	1024×1024
ER	0.20s	0.27s	0.48s
SK	0.13s	0.15s	0.19s
MM	0.06s	0.27s	1.15s
Proposed	0.21s	0.81s	2.04s

All timings are calculated on a 3.1GHz Windows 7 machine running MATLAB 2013*b* with 8GB of RAM. ER [7]; SK [14]; MM [9].

Fig. 11.   Demonstration of the need for the image-specific stain matrix: (*top left*) RGB image; (*top right*) channel-3 obtained by performing CD using standard stain matrix [20]; and (*bottom row*) using image-specific stain matrices estimated using [9] and the proposed method in Section III-A2, respectively. Amount of information content in each channel is measured in terms of entropy (bits per pixel). Ideally the third channel should be empty as only two stains are used in the image.

chemical processes are largely independent for each stain, and CD separates out the effect of variation of each stain so it can be corrected independently. Of the two CD-based approaches evaluated, the method presented herein is significantly the superior. This could be for two reasons: 1) Better, or more robust, deconvolution matrix estimation, and/or 2) a more appropriate mapping function. An investigation on a small number of images revealed that both the proposed (supervised classification based) method and Macenko *et al.*'s (unsupervised) method generally produced more appropriate deconvolution matrices than
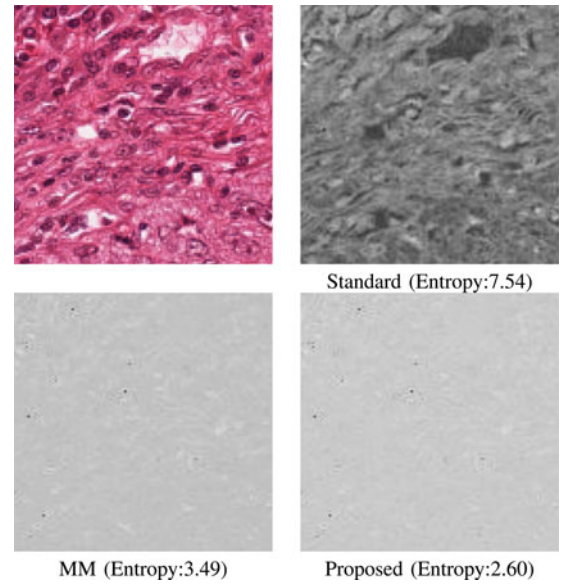
simply using standard matrices, with little data leaking into the third channel (see Fig. 11). The proposed method performed marginally better under this criterion than the approach of Macenko *et al.*, with consistent performance over all images. However, we speculate that the inappropriateness of the simple linear mapping function used in the study (especially at the extremes of distributions) is the main reason for the superior performance of our approach, as our method seems reasonably invariant to choice of thresholds (and thus slight variation in deconvolution matrices), and the matrices produced by Macenko *et al.* are not radically different in most cases.

It is worth pointing out that in experiment 2, the tissue type (breast tissue) was significantly different to the tissue used for training the classifier for our method (liver and oesophagus).

However, the performance on this new set was nonetheless accurate and robust. This suggests that the proposed method is potentially appropriate for a wide variety of applications without re-training.

One important point to raise about the use of any color normalization method which is based on a target image is the choice of an appropriate target image. The choice of target image in our experiments was either random (Experiment-1), or based on manual selection (Experiment-2). In practice, however, careful choice of the target image based on quantitative measures (e.g., cross validation accuracy of some method) applied to a set of normalized images may be required, as performance of a method on a single target image is not necessarily indicative of performance on a normalized image set.

It remains true that often the best way of estimating CD vectors is to apply a single stain to control tissue [20]. This is the method used to estimate the standard matrices supplied with the publicly available implementations of [20]. However, it is often hard to ensure that this control tissue is representative of tissue being analyzed (especially over time, or if analysis involves tissue from multiple labs/technicians). Color normalization methods, such as those presented in this paper, offer an opportunity to utilize such carefully estimated stain matrices over a wider range of tissues. We anticipate application of these techniques to a wide range of problems in histopathology image analysis in the near future. In addition to our success in increasing color consistency, the building blocks of our color normalization (the classifier and SCD) could be used for a wide variety of tasks in histopathology image analysis from segmentation to similarity-based retrieval, content filtering, and visualization applications.

ACKNOWLEDGMENT

REFERENCES

[1] H. Lyon, A. De Leenheer, R. Horobin, W. Lambert, E. Schulte, B. Van Liedekerke, and D. Wittekind, "Standardization of reagents and methods used in cytological and histological practice with emphasis on dyes, stains and chromogenic reagents," *Histochem. J.*, vol. 26, no. 7, pp. 533–544, 1994.

[2] G. Finlayson, S. Hordey, and P. Hubel, "Color by correlation: A simple, unifying approach to color constancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1209–1216, Nov. 2001.

[3] P. Hamilton, P. Bartels, D. Thompson, N. Anderson, and R. Montironi, "Automated location of dysplastic fields in colorectal histology using image texture analysis," *J. Pathol.*, vol. 182, no. 1, pp. 68–75, 1999.

[4] A. Ruiz, O. Sertel, M. Ujaldon, U. Catalyureko, J. Saltz, and M. Gurcan, "Pathological image analysis using the GPU: Stroma classification for neuroblastoma," in *Proc. IEEE Int. Conf. Bioinformat. Biomed.*, 2007, pp. 78–85.

[5] H. Qureshi, O. Sertel, N. Rajpoot, R. Wilson, and M. Gurcan, "Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification," in *Proc. Med. Image Comput. Comput.-Assist. Intervention*, 2008, pp. 196–204.

[6] A. N. Basavanhally, S. Ganesan, S. Agner, J. P. Monaco, M. D. Feldman, J. E. Tomaszewski, G. Bhanot, and A. Madabhushi, "Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 3, pp. 642–653, Mar. 2010.

[7] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 34–41, Sep./Oct. 2001.

[8] A. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.

[9] M. Macenko, M. Niethammer, J. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro*, 2009, pp. 1107–1110.

[10] J. D. Hipp, J. Y. Cheng, M. Toner, R. G. Tompkins, and U. J. Balis, "Spatially invariant vector quantization: A pattern matching algorithm for multiple classes of image subject matter including pathology," *J. Pathol. Informat.*, vol. 2, no. 1, pp. 13–18, 2011.

[11] J. Kong, O. Sertel, H. Shimada, K. Boyer, J. Saltz, and M. Gurcan, "Computer-aided grading of neuroblastic differentiation: Multi-resolution and multi-classifier approach," in *Proc. Int. Conf. Image Process.*, 2007, pp. 525–528.

[12] O. Sertel, J. Kong, H. Shimada, U. Catalyurek, J. Saltz, and M. Gurcan, "Classification of stromal development on whole-slide images," *Proc. SPIE*, p. 69150P, 2008.

[13] Y. Wang, S. Changa, L. Wu, S. Tsai, and Y. Sun, "A color-based approach for automated segmentation in tumor tissue classification," in *Proc. Conf. IEEE Eng. Med. Biol. Soc.*, 2007, pp. 6577–6580.

[14] S. Kothari, J. H. Phan, R. A. Moffitt, T. H. Stokes, S. E. Hassberger, Q. Chaudry, A. N. Young, and M. D. Wang, "Automatic batch-invariant color segmentation of histological cancer images," in *Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro*, 2011, pp. 657–660.

[15] D. Magee, D. Treanor, D. Crellin, M. Shires, K. Smith, K. M. K, and P. Quirke, "Color normalization in digital histopathology images," in *Proc. Opt. Tissue Image Anal. Microsc., Histopathol. Endosc.*, 2009, pp. 100–111.

[16] M. Niethammer, D. Borland, J. Marron, J. Woosley, and N. E. Thomas, "Appearance normalization of histology slides," in *Machine Learning in Medical Imaging*. New York, NY, USA: Springer-Verlag, 2010, pp. 58–66.

[17] M. M. R. Krishnan, P. Shah, C. Chakraborty, and A. K. Ray, "Statistical analysis of textural features for improved classification of oral histopathological images," *J. Med. Syst.*, vol. 36, no. 2, pp. 865–881, 2012.

[18] A. M. Khan, H. El-Daly, and N. Rajpoot, "RanPEC: Random projections with ensemble clustering for segmentation of tumor areas in breast histology images," in *Proc. Med. Image Understand. Anal.*, 2012, pp. 17–23.

[19] D. Magee, D. Treanor, P. Chomphuwiset, and P. Quirke, "Context aware color classification in digital microscopy," in *Proc. Med. Image Understand. Anal.*, 2010, pp. 1–5.

[20] A. Ruifrok and D. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal. Quantitat. Cytol. Histol.*, vol. 23, pp. 291–299, 2001.

[21] C. Taylor and R. Levenson, "Quantification of immunohistochemistry-issues concerning methods, utility and semiquantitative assessment II," *Histopathology*, vol. 49, no. 4, pp. 411–424, 2006.

[22] A. M. Khan, H. El-Daly, and N. M. Rajpoot, "A Gamma-Gaussian mixture model for detection of mitotic cells in breast cancer histopathology images," *J. Pathol. Informat.*, vol. 4, pp. 11–14, 2013.

[23] M. Gervautz and W. Purgathofer, "A simple method for color quantization: Octree quantization," in *New Trends in Computer Graphics*. Berlin, Germany: Springer-Verlag, 1988.

[24] M. Tipping and A. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *Proc. Int. Workshop Artif. Intell. Statist.*, 2003, pp. 3–6.

[25] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.

[26] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.

[27] P. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. Philadelphia, PA, USA: SIAM, 1999.

[28] A. M. Khan, H. El-Daly, E. Simmons, and N. M. Rajpoot, "HyMaP: A hybrid magnitude-phase approach to unsupervised segmentation of tumor areas in breast cancer histology images," *J. Pathol. Informat.*, vol. 4, no. 2, pp. 1–4, 2013.

[29] L. Roux, D. Racoceanu, N. Loménie, M. Kulikova, H. Irshad, J. Klossa, F. Capron, C. Genestie, G. Le Naour, and M. Gurcan, "Mitosis detection in breast cancer histological images," *J. Pathol. Informat.*, vol. 4, pp. 8–14, 2013.