

# Stain Specific Standardization of Whole-Slide Histopathological Images

Babak Ehteshami Bejnordi\*, Geert Litjens, Nadya Timofeeva, Irene Otte-Höller, André Homeyer, Nico Karssemeijer, and Jeroen AWM van der Laak

**Abstract**—Variations in the color and intensity of hematoxylin and eosin (H&E) stained histological slides can potentially hamper the effectiveness of quantitative image analysis. This paper presents a fully automated algorithm for standardization of whole-slide histopathological images to reduce the effect of these variations. The proposed algorithm, called whole-slide image color standardizer (WSICS), utilizes color and spatial information to classify the image pixels into different stain components. The chromatic and density distributions for each of the stain components in the hue-saturation-density color model are aligned to match the corresponding distributions from a template whole-slide image (WSI). The performance of the WSICS algorithm was evaluated on two datasets. The first originated from 125 H&E stained WSIs of lymph nodes, sampled from 3 patients, and stained in 5 different laboratories on different days of the week. The second comprised 30 H&E stained WSIs of rat liver sections. The result of qualitative and quantitative evaluations using the first dataset demonstrate that the WSICS algorithm outperforms competing methods in terms of achieving color constancy. The WSICS algorithm consistently yields the smallest standard deviation and coefficient of variation of the normalized median intensity measure. Using the second dataset, we evaluated the impact of our algorithm on the performance of an already published necrosis quantification system. The performance of this system was significantly improved by utilizing the WSICS algorithm. The results of the empirical evaluations collectively demonstrate the potential contribution of the proposed standardization algorithm to improved diagnostic accuracy and consistency in computer-aided diagnosis for histopathology data.

**Index Terms**—Computer-aided diagnosis, H&E staining, hue-saturation-density, standardization, whole-slide image color standardizer (WSICS), whole-slide imaging.

Manuscript received June 16, 2015; revised August 25, 2015; accepted August 27, 2015. Date of publication September 04, 2015; date of current version February 01, 2016. The authors wish to acknowledge the financial support by the European Union FP7 funded VPH-PRISM project under Grant 601040. The authors acknowledge financial support from the Stichting IT Projecten Nijmegen (NT) and the Maurits en Anna de Kock foundation for image analysis equipment. *Asterisk indicates corresponding author.*

\*B. Ehteshami Bejnordi is with the Diagnostic Image Analysis Group, Radboud University Medical Center, 6500 HB Nijmegen, The Netherlands (e-mail: babak.ehteshamibejnordi@radboudumc.nl).

G. Litjens, N. Timofeeva, I. Otte-Höller, and J. A.W.M. van der Laak are with the Department of Pathology, Radboud University Medical Center, 6500 HB Nijmegen, The Netherlands.

A. Homeyer is with Fraunhofer MEVIS, 28359 Bremen, Germany.

N. Karssemeijer is with the Diagnostic Image Analysis Group, Radboud University Medical Center, 6500HB Nijmegen, The Netherlands.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2015.2476509

## I. INTRODUCTION

HISTOPATHOLOGY involves microscopic examination of stained histological slides to study presence and characteristics of disease. Tissue sections are stained with multiple contrasting dyes to highlight different tissue structures and cellular features [1]. This staining provides invaluable information to the pathologists for diagnosing and characterizing various pathological conditions. Pathologists make diagnosis of the disease based on features such as morphology and spatial arrangement of cells [2]. This task is, however, laborious and prone to subjectivity [3], [4]. Several previous studies have shown poor concordance among pathologists in histopathological grading of prostate, cervical, and breast cancer [5]–[7]. Computer-aided diagnosis (CAD) can potentially alleviate shortcomings of human interpretation. CAD can facilitate diagnosis by sieving out obviously benign slides and providing quantitative characterization of suspicious areas.

The appearance of the histological stains (e.g., the widely used hematoxylin and eosin (H&E) staining) often suffer from large variability [8]. While pathologists can effectively cope with staining variability, the performance of CAD systems can be hampered by color and intensity variations. Such variations in digital pathology images may be attributed to a number of factors, including specimen preparation and staining protocol inconsistencies (e.g., temperature of solutions), variations in fixation characteristics, inter-patient variation, and the scanner used to digitize the slides [8]. The use of standardized staining protocols and automated staining machines may improve staining quality by yielding a more accurate and consistent staining. However, eliminating all the underlying sources of variation is infeasible [9]. The problem is more acute in studies in which different laboratories share digital images. As an alternative, algorithms for automated standardization of digitized whole-slide images (WSI) have been published [9]–[16] (described in detail below). Ideally, such an algorithm takes as input an arbitrary WSI and yields a normalized version of the image, with standardized appearance of the different dyes. The algorithm should be able to eliminate all sources of variation.

Many color standardization algorithms, also referred to as color normalization algorithms, are based on stain-specific color deconvolution [17]. Stain deconvolution requires prior knowledge of reference stain vectors for every dye present in the WSI. Ruifrok *et al.* [17] suggested a manual approach to estimate color deconvolution vectors by selecting representative sample pixels for each stain class. A similar approach was used in [10] for extracting the stain vectors. Manual interaction for estimation of stain vectors, however, strongly limits its applicability

in large studies. The method devised by Macenko *et al.* [11] enables automatic extraction of reference stain vectors by finding the fringe of pixel distributions in the optical density space. The method, however, yields poor estimation of the stain vectors in the presence of strong staining variations. Several other approaches have been suggested for automatic extraction of stain vectors [9], [12]. The major drawback of these methods is that the estimation of the stain vectors relies solely on the color information present in the image. The outstanding ability of a pathologist to identify stain components is not only because of using the color information but also because of incorporating the spatial dependency of tissue structures (e.g., cell nuclei have a near elliptical shape and in H&E staining are mostly stained with hematoxylin while other tissue structures are mostly eosin stained). Neglecting to take into account the spatial dependency of tissue structures for determination of the stain vectors significantly limits the robustness of these methods in presence of severe staining variations.

Reinhard *et al.* [13] proposed one of the first techniques to standardize image colors with respect to a template image through the use of a color model. Their proposed technique aligns each of the color channels of the Lab color model [18] to the corresponding channels of a predefined template image. However, the use of a single transformation function for each channel will, in practice, rarely suffice. In general, dyes have independent contribution to the final color in the image, as each dye has its own specific reaction pattern. Consequently, using a single transformation function may lead to improper color mapping of the stain components in the standardization process. This problem can be addressed by applying separate transformations to different stain classes [10], [14] or different tissue classes [15]. The applicability of the approach described in [15] is limited to image patches that necessarily contain all the considered tissue classes. Moreover, the use of Gaussian mixture model in [15] and [10] for segmentation of tissue or stain classes lacks robustness to strong staining variations. The performance of the algorithm described in [14] also decreases when there is a considerable imbalance in the amount of different stain classes in the image. This is mainly because the accuracy of the stain classification approach utilized (based on singular value decomposition) decreases as the imbalance between the amount of different stain components increases.

In addition to the limitations discussed above, the majority of these studies have 3 major shortcomings which limit their applicability to studies using large cohorts. First of all, the efficacy of most of the algorithms was not evaluated on data from multiple laboratories (preferably both academic and non-academic) with different staining protocols, which potentially causes the most severe staining variations [14]. Secondly, most of the proposed algorithms to date have been solely focusing on standardization of patch images containing a small region within the WSI. Development of a fully automated CAD system for large-scale digital pathology requires algorithms that can deal with WSI data. Whole-slide extension of the existing patch-based standardization algorithms may not be straightforward and requires automatic extraction of training patches from the entire slide to define the required color transformations. In [16] an automated algorithm has been proposed for WSI standardization which can

handle variations caused by using different WSI scanners. However, variations caused by other sources such as staining protocols are more problematic [14] and can not be dealt with by this approach. Finally, although the major aim of most published algorithms is to enhance the performance of CAD systems, the efficacy of these algorithms was typically not evaluated on an existing CAD system. To the authors' knowledge no algorithm has been proposed to date that addresses WSI standardization in a fully automated manner in presence of all sources of variation.

This paper presents and evaluates a novel fully automatic algorithm for standardization of whole-slide H&E stained histopathological images. The algorithm, called whole-slide image color standardizer (WSICS), is based on transformation of the chromatic and density distributions for each individual stain class in the hue-saturation-density (*HSD*) color model [19]. While standardization strongly facilitates accurate recognition of stain components, a good technique for determination of stain components can help in developing a robust standardization algorithm. Such a technique should recognize different dye components in presence of various sources of stain variation. Unlike the available supervised and unsupervised pattern recognition techniques for dye recognition in the literature, which rely solely on color features, the proposed algorithm makes use of spatial information, making it robust against severe color and intensity variations. Standardization of the stain components in the WSI is achieved by aligning the chromatic and density distributions of the stain components to predefined corresponding histograms from a template WSI. The WSICS algorithm is compared to three state-of-the-art standardization algorithms both qualitatively and quantitatively. Empirical evaluation of the impact of our standardization algorithm on the performance of an already published CAD system for necrosis quantification [20] is also presented.

## II. OVERVIEW OF THE PROPOSED WHOLE-SLIDE STANDARDIZATION METHOD

The WSICS algorithm takes as input a WSI and outputs a standardized image with staining characteristics similar to a predefined template WSI. We interface our standardization algorithm with a 2D WSI, using our own developed open source library [21] which is built on top of the OpenSlide library [22]. This library allows us to read a large set of WSI file formats (e.g., .tif, .vsi, .mrxs, .svs, etc.).

Our proposed algorithm initially classifies the pixels in the WSI into different dye classes and then applies standardization on the chromatic and density components of the *HSD* model [19]. Fig. 1 presents an overview of the WSICS algorithm. The proposed algorithm has 6 basic steps:

- 1) Applying *HSD* transform.
- 2) Automatic extraction of samples for the hematoxylin, eosin and background classes from the WSI and deriving the chromatic and density distributions of these classes.
- 3) Transforming the 2D chromatic distribution for each dye class to match the chromatic distribution of the corresponding class from a template slide.
- 4) Transforming the density distribution for each dye class to match the density distribution of the corresponding class from a template slide.

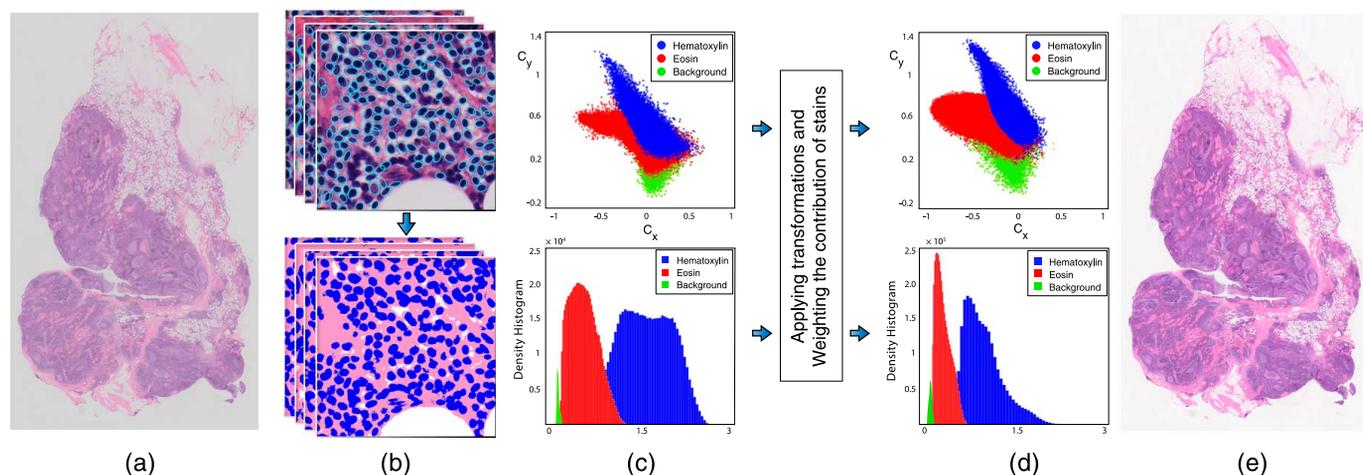


Fig. 1. Illustration of selected steps of the WSICS algorithm. (a) Original WSI of a lymph node tissue section. Several regions containing tissue are randomly selected for automatic extraction of representative samples for the hematoxylin, eosin and background class. (b) The selected regions of interest are classified into: pixels absorbing mostly hematoxylin, pixels absorbing mostly eosin, and background pixels. (c) The chromatic distribution and density histogram of pixels are defined in the *HSD* model to be transformed to match a template WSI. (d) The result of transforming the chromatic and Density distributions after application of weights. (e) Reconstructing the *RGB* image by applying reverse *HSD* transform using the standardized chromatic and density components.

- 5) Weighting the contribution of stains for each pixel and obtaining final chromatic and density transformations.
- 6) Applying inverse *HSD* transform.

Detailed description of the WSICS algorithm's steps are discussed below.

#### A. *HSD* Transform

The algorithm first applies the hue-saturation-density (*HSD*) color transformation [19]. In [19] we showed that the *HSD* model is better-suited for analysis of transmitted light microscopy compared to the *RGB* and *HSI* models. The *HSD* model transforms *RGB* data into two chromatic components ( $c_x$  and  $c_y$ ; which are independent of the amount of stain) and a density component ( $D$ ; linearly related to the amount of stain). Theoretical *RGB* intensities obtained from varying stain densities should result in a single point in the  $c_x c_y$  chromaticity plane of the *HSD* transform [19]. However, in practice, the use of broad-band camera filters and existence of pixel inhomogeneity (stain variability over the specimen area occupied by a pixel) lead to dispersion of chromatic data in this plane for each dye component. As a result, the chromatic data of the pixels stained with a particular dye component will form a distribution, which is represented by  $F(c_x, c_y)$ . In H&E staining the chromatic distribution of the hematoxylin stain  $F_H(c_x, c_y)$ , and the chromatic distribution of the eosin stain  $F_E(c_x, c_y)$  have a significant overlap.

#### B. Deriving the Chromatic and Density Distributions of Hematoxylin, Eosin, and Background

WSIs are generally stored in a multi-resolution pyramid structure. Image files contain multiple downsampled versions of the original image. Each image in the pyramid is stored as a series of tiles, to facilitate rapid retrieval of subregions of the image which enable us to quickly identify regions that are rich in tissue.

For a WSI, extraction of representative samples for hematoxylin, eosin and background classes starts with identifying

the tiles containing more than 75% of tissue (non-background pixels) on the lowest magnification. The pixel size of the image at this magnification is  $3.88 \mu\text{m} \times 3.88 \mu\text{m}$ . Each tile is a  $64 \times 64$  pixel image. A pixel is classified as background if its overall density is lower than 0.2.

Next, we randomly select one of the tiles identified in the lowest magnification and apply Restricted Randomized Hough Transform [23] to detect candidate nuclei in the corresponding area in the highest magnification. If the number of detected nuclei surpasses a predefined threshold (200 nuclei) the image is classified into different dye components, and the labeled pixel samples are stored. This process is repeated for a large number of randomly selected tiles until a predefined number of samples (in this study 3 million pixels) are acquired for each class.

To make the sampling process robust against the inhomogeneities present in some slides, it is possible to sample less pixels in each tile image but instead sample from more random patches from the entire WSI. This, however, comes with an additional computation cost.

Classification of pixels into different dye components requires prior knowledge of  $c_x c_y$  color vectors for every dye in the WSI. To enable fully automated classification, the algorithm automatically extracts training samples for each stain class (the class absorbing mostly hematoxylin and the class absorbing mostly eosin) from the image, thus obviating the need for manually labeled training data. Our technique for generating training samples makes use of prior shape knowledge (cell nuclei are usually ellipse shaped). Incorporating spatial information such as prior shape knowledge into the classification problem provides robustness to color and intensity variations.

The process for classifying different tissue components is illustrated in Fig. 2. The first step involves locating the nuclei and estimating their boundaries with ellipses by utilizing the Restricted Randomized Hough Transform [23] on the Canny edge detected image over the density component of the *HSD* model (see Fig. 2(b)). To remove artifacts among the detected candidate nuclei, percentile thresholding is applied on the average

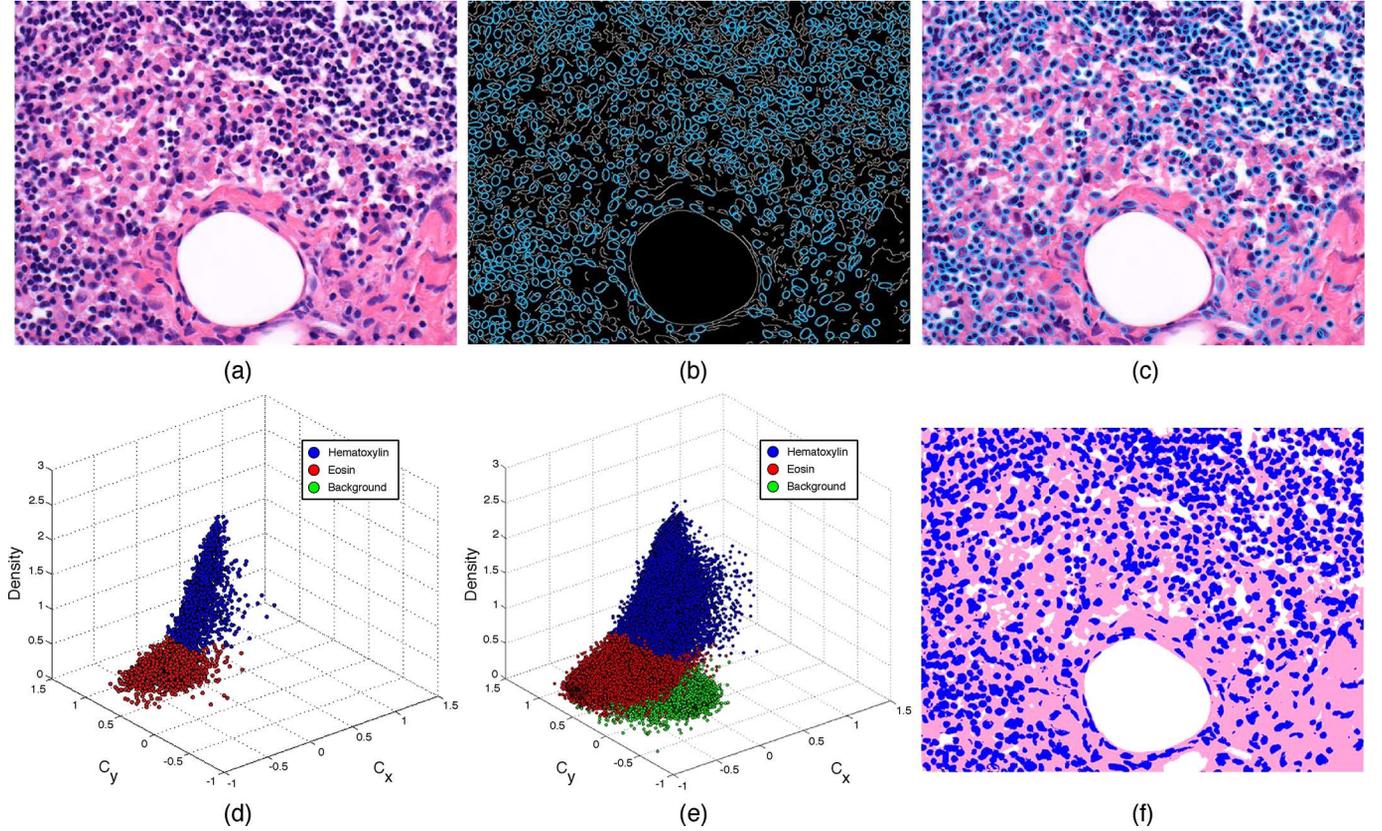


Fig. 2. Illustration of the classification steps. (a) Sample image of a lymph node tissue section. (b) Utilizing Canny edge detector on the density image in conjunction with the Restricted Randomized Hough Transform to detect and estimate the boundaries of the nuclei. (c) The result of detected nuclei overlaid on the original image after artifact rejection. (d) Scatter plot showing  $c_x$ ,  $c_y$ , and density features of samples extracted from hematoxylin and eosin class. (e) Scatter plot after classifying the entire image. Note that the data points associated with the background class have been obtained by thresholding. (f) The classification result produced by our method.

optical density of the red camera channel ( $D_r$ ) and the average overall density ( $D$ ). The optical density of a channel is defined as:

$$D_{ch} = -\ln\left(\frac{I_{ch}}{I_{o,ch}}\right)$$

where  $I_{ch}$  is the intensity of channel  $ch$  (which can be  $R$ ,  $G$ , or  $B$  in RGB color model), and  $I_{o,ch}$  is the intensity of channel  $ch$  when no stain is present. The overall density ( $D$ ) refers to the density component of the HSD model. Among the detected candidates, objects with very low average  $D$  (e.g., background) and objects presenting very low average  $D_r$  (e.g., red blood cells) are removed by applying these thresholds. We found the 8th percentile to be a suitable threshold for both average  $D_r$  and average  $D$  measures. A random selection of samples from the pixels belonging to the detected nuclei constructs conservative representative training samples for the tissue class mostly absorbing hematoxylin. As shown in Fig. 2(c), our algorithm does not require exhaustive detection of all the nuclei, but rather a proportion of the nuclei within the image.

In the next step, all the pixels with very low overall density ( $D < 0.2$ ) are classified as background (devoid of stain). Provided that background pixels are white, the optical density of the red, green, and blue camera channels of the pixel should each be lower than 0.25 to be classified as background.

Training samples for the eosin class are obtained by first removing all the background pixels and the candidate pixels for the hematoxylin class (before artifact rejection) and then applying 5th percentile thresholding on the  $D_r$  of the remaining pixels. Note that the training samples for the eosin class are randomly selected from these pixels and are equal in number to the training samples from the hematoxylin class. The pixels selected during the sampling process will serve as ground truth data for a classifier. By applying a threshold on the optical density of the red channel, we deliberately avoid sampling weakly stained pixels in the eosin class and try to retain samples that are as clean as possible.

Finally, a binary k-NN classifier ( $k = 7$ ) is trained using the extracted samples from the hematoxylin and eosin class. The chromatic information ( $c_x$  and  $c_y$ ) and the density of each pixel in the HSD color model are used as classification features. The class labels for all the remaining pixels in the randomly selected training patches are predicted using the trained classifier, yielding the final pixel classification result.

The derived chromatic and density distributions for each of the stain classes (as shown in Fig. 2(e)) are used for subsequent chromatic and density transformations.

### C. Non-Linear Transformation of Chromatic Information

To align the chromatic distribution of each of the classified dye components to match the corresponding class distribution

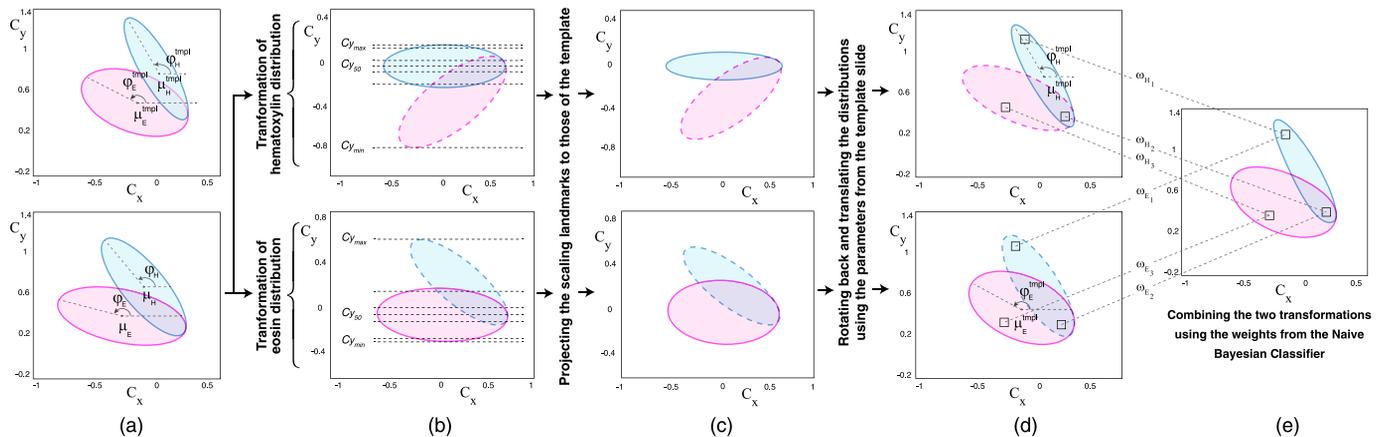


Fig. 3. Illustration of the non-linear transformation of hematoxylin and eosin chromatic distributions. Note that the transformation applied to the background class is not shown in this figure. (a) The chromatic distribution of hematoxylin (blue) and eosin (pink) staining for the template slide (top left) and the slide to be standardized (bottom left). The angle ( $\phi_i$ ) and the mean ( $\mu_i$ ) of the H&E distributions are extracted, where  $i \in \{Hematoxylin, Eosin\}$ . (b) Extraction of the scaling parameters after translation of each distribution to the origin and rotation along  $\phi_i$ . The extracted parameters comprise of the minimum, 1st, 25th, 50th, 75th, and 99th percentiles, and maximum of the projected values along each axis (for simplicity this is only shown for parameters along the  $c_y$  axis). (c) The result of scaling the hematoxylin and eosin distributions by using the scaling parameters derived from the template slide. (d) Rotation of the distributions along  $\phi_i^{tmpl}$  and translating them to  $\mu_i^{tmpl}$ . (e) The final transformation obtained using the weighted combination of the two transformations in (d).

in a template slide, we perform 2D registration of the color information in the  $c_x c_y$  chromatic plane. Let  $F_i(c_x, c_y)$  denote the chromatic distribution for class  $i$  in the image, where  $i \in \{Hematoxylin, Eosin, Background\}$ . We define the registration problem as finding the transformation function  $T_i$  such that:

$$F_i(T_i(c_x, c_y)) \sim F_i^{template}(c_x, c_y). \quad (1)$$

The procedure for registration of the chromatic distribution has two steps: (1) extraction of statistical parameters from the template slide and (2) aligning the 2D chromatic distribution. Each of the steps are described in the following subsections.

1) *Template Parameter Extraction*: Training samples for the two stain classes in the template image are obtained automatically as described in Section II-A above. From these, 3 sets of parameters are calculated (two for the stain classes and one for the background class). These parameters are the mean  $\mu_i^{tmpl}$ , angle  $\phi_i^{tmpl}$ , and scaling landmarks of the  $F_i^{template}$  distribution. Let  $\Sigma_i^{tmpl} = cov(c_x, c_y)$  denote the covariance matrix of  $F_i^{template}$ . The angle  $\phi_i^{tmpl}$  of this distribution with respect to the  $c_x$  axis is derived by calculating the major eigenvector of  $\Sigma_i^{tmpl}$ . To compute the scaling parameters, the entire  $F_i^{template}$  distribution is translated to the origin. This is followed by a rotation step along  $\phi_i^{tmpl}$  to maximize the variance along the  $c_x$  axis. Finally, the scaling landmarks are defined after projection of the rotated distribution onto each of the  $c_x$  and  $c_y$  axes, comprising the minimum, 1st, 25th, 50th, 75th, and 99th percentiles, and maximum of the projected values along each axis.

The only parameters for the background class are the mean values of the  $c_x$  and  $c_y$  channels. This is mainly because the white background class does not require major color standardization but rather density standardization which is achieved by adjusting its density component in the *HSD* model.

2) *Alignment of 2D Color Histogram*: At this step, the 2D chromatic distribution for each of the three classes in the WSI to be standardized are aligned with the chromatic distribution of

the corresponding class in the template WSI. Fig. 3 illustrates different steps for the non-linear transformation of chromatic information. Let  $F(c_x, c_y)$  denote the chromatic distribution of the WSI to be standardized. We apply three separate transformations  $T_i(c_x, c_y)$ , each time focusing at standardizing the chromatic distribution of a particular class  $i$ .

Let  $F_i(c_x, c_y)$  denote the chromatic distribution of the pixels belonging to class  $i$  sampled from the slide to be standardized. For different pixel classes in this slide, statistical parameters are extracted identical to those previously extracted from the template slide. The process for standardization of each stain class  $i$  starts with translating the entire  $F(c_x, c_y)$  distribution by subtraction of the mean of  $F_i(c_x, c_y)$  distribution and rotation along the major eigenvector of  $\Sigma_i$ , where  $\Sigma_i$  denotes the covariance matrix of  $F_i(c_x, c_y)$  (see Fig. 3(a) and (b)). Then we apply piece-wise linear scaling to match landmarks from the current distribution to those of the template slide. The result is shown in Fig. 3(c). In the next step, the scaled distribution is rotated back along the  $\phi_i^{tmpl}$  angle to be aligned with the major eigenvector of the corresponding distribution in the template WSI. The final step in the transformation of the  $F_i(c_x, c_y)$  involves translation of the distribution to  $\mu_i^{tmpl}$ . The standardized chromatic distributions with focus on transforming hematoxylin and eosin classes are shown in the first and second row of Fig. 3(d), respectively.

The color transformation of the background class is yielded by subtracting  $\mu_b$  from and adding  $\mu_b^{tmpl}$  to the  $F(c_x, c_y)$  distribution, where  $\mu_b$ , and  $\mu_b^{tmpl}$  denote the mean of the background class in  $F(c_x, c_y)$ , and  $F^{template}(c_x, c_y)$  distributions, respectively. At the end of this step we have 3 separate transformation functions, one for each class.

#### D. Weighing the Contribution of Classes

Provided that in H&E staining, pixels may contain a mixture of stains, it is important to take into account the contribution of each stain for each pixel in our final transformation. Therefore, we define the final transformation as the weighted

combination of the transformation functions associated for each class. To generate per pixel weights accounting for relative absorption of each stain, a naive Bayesian classifier is trained. The ground truth data for training the classifier originates from the automatically classified samples extracted in step 1 of the WSICS algorithm. By using the chromatic and density components of the *HSD* transform as features, we calculate the posterior probability of each pixel in the WSI belonging to each class. Finally, the weighted combination of the three transformations gives:

$$(c'_x, c'_y) = \sum_{i=1}^3 \omega_i * T_i(c_x, c_y) \quad (2)$$

where  $i \in \{Hematoxylin, Eosin, Background\}$  and  $\omega_i$  denotes the weight for class  $i$ , and  $(c'_x, c'_y)$  denotes the final transformed chromatic vector in the *HSD* model. The result is shown in Fig. 3(e).

The naive Bayesian classifier was chosen to generate the membership degree of each pixel to each of the stain classes because of its straightforward probabilistic interpretation and its relative simplicity, requiring no parameter tuning in contrast to more complex classifiers. Given the nature of the classification task (3 features and millions of samples) a more complex classifier is not needed.

#### E. Density Standardization

The density component of the *HSD* color model was also transformed to match the density profile of the template image using a weighted combination of linear transformations with respect to each class. Applying linear transformation of the densities for separate classes without using the weights may lead to severe artifacts. The reason for this is that pixels lying in the vicinity of classification boundary but belonging to different classes will be standardized with separate transformation functions, hence transforming into new density values which might differ significantly. To address this problem, the weights generated by the naive Bayesian classifier in the previous section were used to create a smooth density standardization. In the first step a weighted mean and a weighted standard deviation were computed for each of the density distributions corresponding to the hematoxylin, eosin and background classes separately. For each of these classes, the distributions were standardized by matching the mean and standard deviation of the distribution to the corresponding class statistics in the template image. The transformed density distribution for class  $i$  is therefore determined by:

$$D'_i = \frac{D_i - \mu_i}{\sigma_i} \times \sigma_i^{tmpl} + \mu_i^{tmpl} \quad (3)$$

where  $i \in \{Hematoxylin, Eosin, Background\}$ ,  $\mu_i$  and  $\sigma_i$  are the weighted mean and the weighted standard deviation of the density distribution for class  $i$  in the WSI to be standardized and  $\mu_i^{tmpl}$  and  $\sigma_i^{tmpl}$  are the corresponding values in the template WSI.  $D_i$  and  $D'_i$  denote the densities before and after transformation.

By using the weights obtained from the naive Bayesian classifier, a weighted combination of the linear transformations for

the three classes were computed to yield the final density transformation:

$$D' = \sum_{i=1}^3 \omega_i \times D'_i \quad (4)$$

where  $\omega_i$  denotes the posterior probability of class  $i$ ,  $D'_i$  denotes the standardized density associated with the distribution of class  $i$ , and  $D'$  is the final standardized density component.

#### F. Inverse HSD Transform

In the final step, the standardized chromatic components  $c'_x$  and  $c'_y$ , and the standardized density component  $D'$  were used to get back to *RGB* model. This was achieved by following the reverse *HSD* transformation illustrated in [19]. The output of this step is the standardized WSI with the staining characteristics similar to the predefined template WSI.

### III. EMPIRICAL EVALUATION

#### A. Histopathology Image Dataset

Two histopathological image datasets were used for empirical evaluation of the proposed algorithm. The first dataset consisted of 125 digitized H&E stained WSIs of lymph nodes from 3 patients. These slides were serially sectioned and stained in 5 different Dutch pathology laboratories, each using their own routine staining protocols. The set up included staining of slides on different days of the week. Staining protocol variations between laboratories include temperature, concentration, staining time and manufacturer of different solutions. Frequency of refreshing staining solutions may also differ between laboratories. All slides were digitized using a CCD *RGB* camera (Zeiss AxioCam HRC) mounted on a light microscope (Zeiss AxioPlan 2im) with a  $40 \times$  objective lens. Each image has square pixels of size  $0.256 \mu\text{m} \times 0.256 \mu\text{m}$  in the microscope image plane.

The second image dataset comprised three batches, each containing 10 H&E stained histological slides of rat liver with different amounts of confluent necrosis. The slides were stained in University Hospital Jena, Germany, and scanned with a Hamamatsu Nanozoomer Scanner at  $40 \times$  objective magnification. A human expert generated ground truth data on this dataset by annotating necrotic tissue in each section at small tile levels of size  $256 \times 256$  pixels in the highest resolution.

#### B. Experiments and Results

To evaluate the performance of the WSICS algorithm three experiments were performed:

- 1) A quantitative, comparative evaluation of the performance of the proposed algorithm versus two competing algorithms in calculating stain vectors
- 2) A qualitative and quantitative evaluation of the performance of the proposed standardization algorithm in comparison with other methods in achieving color constancy
- 3) A quantitative evaluation of the effect of employing the proposed whole-slide standardization algorithm on the performance of a necrosis quantifying CAD system [20]

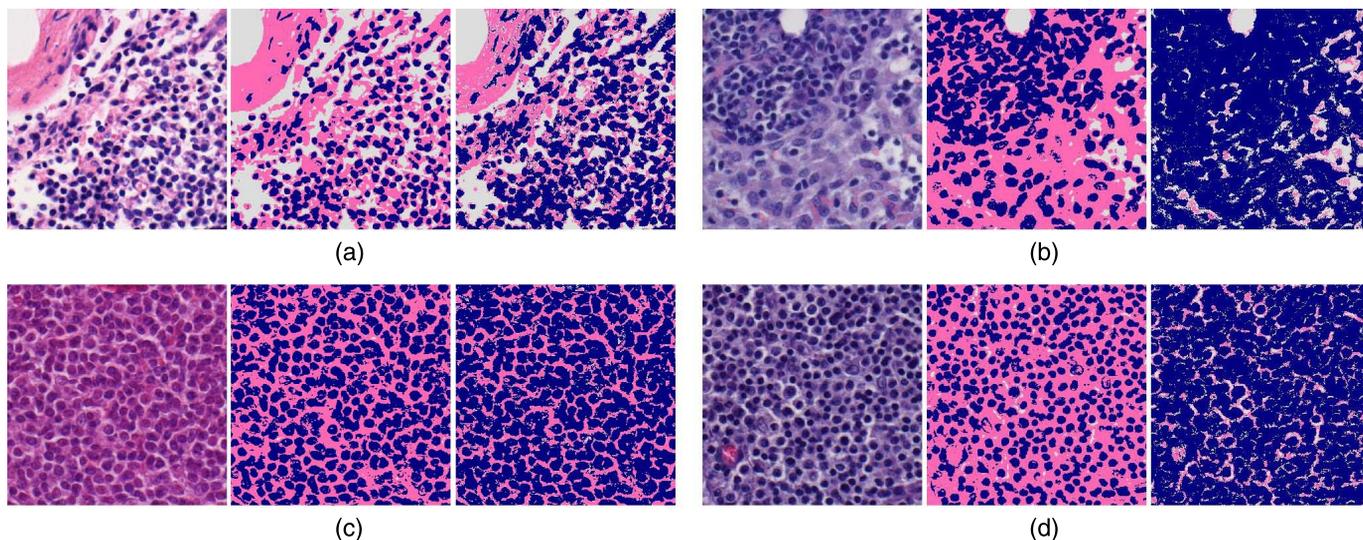


Fig. 4. Comparison of the pixel classification performances between the method by Khan *et al.* [12] and our proposed algorithm. (a) – (d) show 4 different classification results produced by the two algorithms. For each sub-figure the image on the left is the original image to be classified. The middle image presents the classification result by our algorithm and on the right is the result produced by the algorithm of Khan *et al.* [12].

1) *Experiment 1*: The aim of this experiment was to quantitatively evaluate the performance of each standardization algorithm in extracting stain vectors using a subset of the lymph node dataset. For the experiment, we randomly selected 5 different slides for each laboratory and took one sample FOV image from each slide. This yielded a total of 25 images from 25 slides. For each of the images, a large number of pixels were manually annotated to give representative ground truth pixels for hematoxylin and eosin classes. Subsequently, hematoxylin and eosin stain vectors were calculated using Ruifrok's color deconvolution method [17].

The performance of our proposed algorithm is compared to that of two state-of-the-art algorithms for stain deconvolution: the appearance normalization algorithm by Macenko *et al.* [11], and the nonlinear mapping approach to stain normalization by Khan *et al.* [12]. The algorithm by Macenko *et al.* [11] tries to find the fringe of pixel distribution in the optical density space to determine the stain vectors. The algorithm by Khan *et al.* utilizes a pretrained relevance vector machine (RVM) classifier to classify the pixels in the image into different stain components in the image. The stain vectors are then calculated from the set of labeled pixels for each stain class. In this regard, this method works similar to our proposed algorithm, however, our algorithm does not require manual training of the classifier. The other fundamental difference of our method is that it mainly uses shape information for identifying hematoxylin pixels, and samples from pixels outside of ellipse-shaped objects for eosin.

Fig. 4 shows the results of classifying pixels as hematoxylin, eosin, or background, which are intermediate steps in our color standardization algorithm and that of Khan *et al.* [12]. As seen in this figure, the algorithm by Khan performs poorly on the third and fourth example which can be related to the fact that the staining colors in these images deviate from the image batches that were originally used for training the RVM classifier. Contrary to this approach, our fully automatic algorithm effectively derives the hematoxylin and eosin distributions and

TABLE I  
AVERAGE EUCLIDEAN DISTANCES BETWEEN THE MANUALLY DETERMINED STAIN VECTORS IN THE ORIGINAL IMAGES AND THE STAIN VECTORS COMPUTED BY DIFFERENT ALGORITHMS.  $d_H$  AND  $d_E$  DENOTE THE AVERAGE EUCLIDEAN DISTANCES FOR THE HEMATOXYLIN AND EOSIN VECTORS, RESPECTIVELY.

Method	$d_H$ (mean $\pm$ SD)	$d_E$ (mean $\pm$ SD)
Mackenco [11]	0.2120 $\pm$ 0.0314	0.1764 $\pm$ 0.0296
Khan [12]	0.0393 $\pm$ 0.0134	0.1187 $\pm$ 0.0297
Proposed	<b>0.0135 <math>\pm</math> 0.0093</b>	<b>0.0186 <math>\pm</math> 0.0225</b>

classifies the pixels very accurately mainly due to using shape information.

The mean and the standard deviation (SD) of the Euclidean distances between the stain vectors from the annotated data and the stain vectors derived by each of the algorithms is used as a measure to compare the efficacy and robustness of the algorithms in extracting the correct stain vectors. Table I presents the results. Our proposed algorithm is performing considerably better in calculating stain vectors for both hematoxylin and eosin stains.

2) *Experiment 2*: The aim of this experiment was to qualitatively and quantitatively evaluate the performance of the WSICS algorithm. We focus on inter-laboratory variations of the H&E staining in the lymph node dataset, as this is a major concern in large scale application of CAD in pathology. The performance of our proposed algorithm is compared to that of three previously published algorithms: global standardization (GS) by bagci *et al.* [24], the appearance normalization algorithm by Macenko *et al.* [11], and the nonlinear mapping approach to stain normalization by Khan *et al.* [12]. Five representative field-of-view (FOV) images were acquired from each WSI yielding a total of 625 images. Each image is of size 1388  $\times$  1040 pixels. The results of the standardization performed by different algorithms are shown in Fig. 5. The image shown in the top left of Fig. 5.a was used as the template image to extract parameters required by different standardization algorithms.

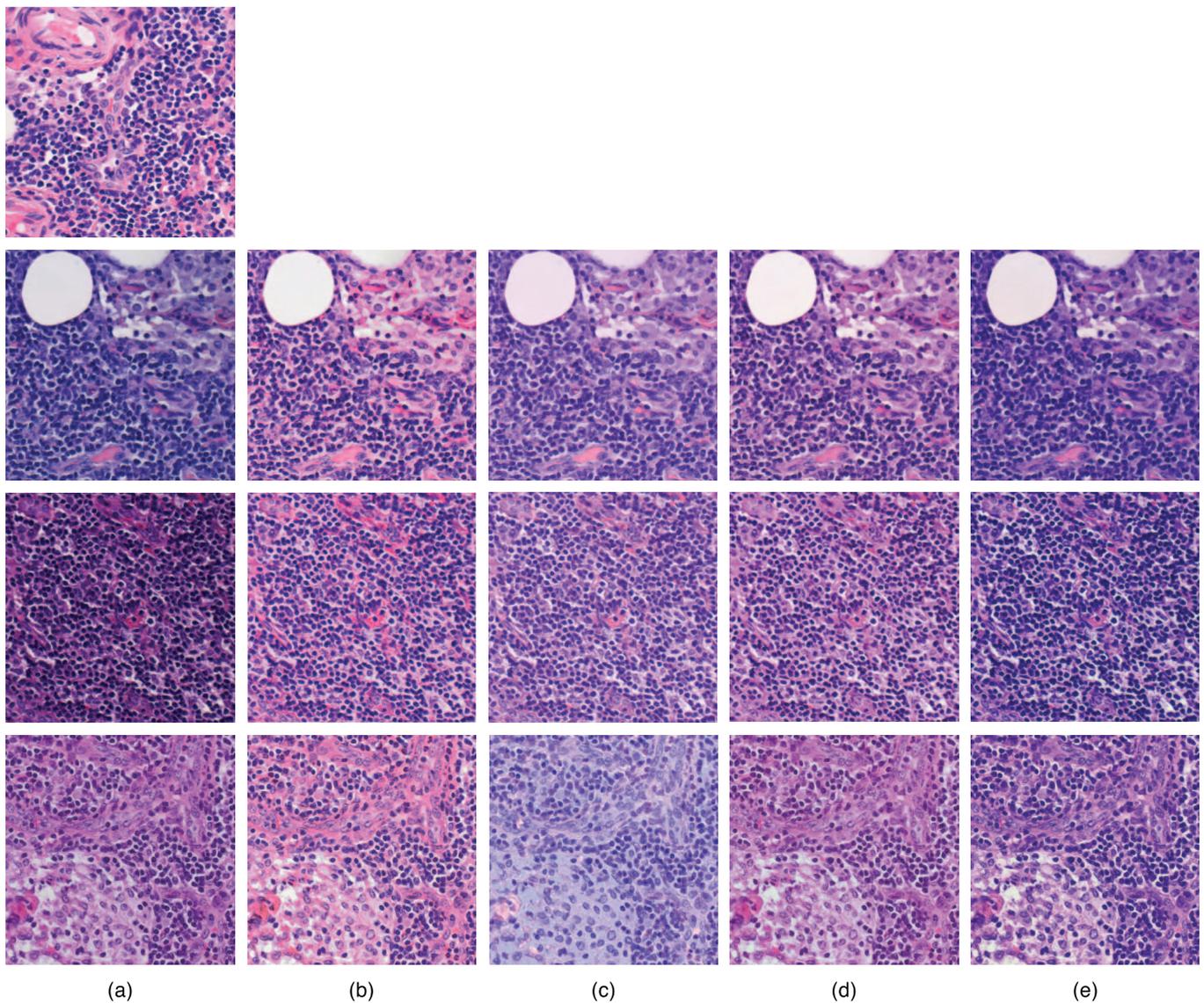


Fig. 5. Illustration of the performance of different stain standardization algorithms. The top left image has been used as the template image. Column (a) three original images sampled from different slides stained in different laboratories. (b) the result of standardization using the WSICS, (c) the algorithm devised by Macenko *et al.* [11], (d) Bagci *et al.* [24], and (e) the algorithm by Khan *et al.* [12].

The three standardized images by the method proposed by Macenko *et al.* [11] have more color variability compared to the other methods. Moreover, it can be seen that the color of the images standardized by the WSICS algorithm have the highest similarity to the color of the template image as compared to the other approaches.

In this experiment, the choice of the template slide was based on the opinion of two pathologists, who studied a large number of slides from each laboratory. The major criteria for them to select a high quality staining are: (1) high contrast between hematoxylin and eosin staining (2) visibility of the nuclear texture. The majority of the slides stained in Lab 1 were found to meet these criteria and therefore a slide from this batch was selected as the template slide in this experiment.

Quantitative analysis of standardization results is based on color constancy of nuclear staining and eosin staining independently. To evaluate the color constancy of the nuclear staining, nuclei were first detected using fast radial symmetry trans-

form [25]. The detected candidate nuclei were subsequently segmented using a marker-controlled watershed algorithm as illustrated in [26]. Quantitative measures of the area and elliptical shape were computed for each candidate nucleus. The elliptic variance descriptor ( $E_{var}$ ) [27] was used to measure how closely the borders of a fitted ellipse agree with those of the segmented nucleus-like object. Objects that were too small (area < 200) or irregular ( $E_{var} > 0.13$ ) were rejected as artifacts. The normalized median intensity (NMI) measure [15], [28] was then chosen to evaluate color constancy of the nuclei. This measure enables comparison of the intensity statistics over a population of images. The NMI measure is defined as:

$$NMI(I) = \frac{Median\{U(i)\}_{i \in I}}{P_{95}\{U(i)\}_{i \in I}} \quad (5)$$

where  $U(i)$  denotes the average of the  $R$ ,  $G$ ,  $B$  values for the pixel  $i$  in image  $I$ , and  $P_{95}$  denotes the 95th percentile. Note

TABLE II  
STANDARD DEVIATION AND COEFFICIENT OF VARIATION OF NMI FOR ALL THE IMAGES IN THE FIVE LABORATORIES.

Method	Laboratory 1		Laboratory 2		Laboratory 3		Laboratory 4		Laboratory 5	
	NMI SD	NMI CV								
<b>Original</b>	0.0206	0.0405	0.0254	0.0485	0.0305	0.0641	0.0279	0.0466	0.0201	0.0348
<b>Bagci [24]</b>	0.0157	0.0294	0.0184	0.0336	0.0163	0.0307	0.0256	0.0422	0.0172	0.0303
<b>Macenko [11]</b>	0.0180	0.0285	0.0161	0.0256	0.0092	0.0148	0.0203	0.0317	0.0127	0.0200
<b>Khan [12]</b>	0.0154	0.0313	0.0194	0.0385	0.0199	0.0390	0.0214	0.0398	0.0139	0.0254
<b>WSICS</b>	<b>0.0083</b>	<b>0.0147</b>	<b>0.0075</b>	<b>0.0131</b>	<b>0.0081</b>	<b>0.0141</b>	<b>0.0087</b>	<b>0.0155</b>	<b>0.0052</b>	<b>0.0092</b>

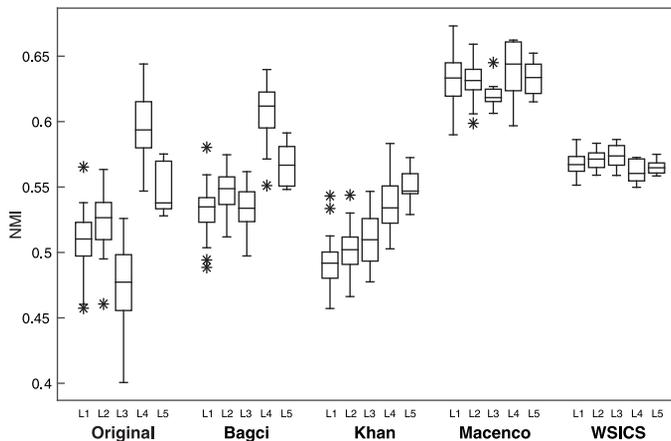


Fig. 6. Box plots of the normalized median intensity values of the slides stained in five laboratories for all the methods in Experiment 3.

that to increase the robustness of the NMI measure against noisy pixels in the image, instead of dividing the median term by the maximum, we divide it by the 95th percentile. The standard deviation of the NMI values (SDN) and coefficient of the variation (i.e., standard deviation divided by mean) of the *NMI* values (CVN) were computed for the images in different laboratories before and after standardization using different methods. The results are shown in Table II. Note that small values for SDN and CVN indicate that nuclei in images from different laboratories have similar color distributions (i.e., qualitatively look the same) after stain standardization. In all cases, the WSICS algorithm yielded the smallest SDN and CVN. The box plots of the NMI values for each laboratory is shown in Fig. 6. The box plot shows that the spread of NMI values about the median (inter-quartile range) is the smallest for the proposed algorithm compared to the competing state-of-the-art algorithms. In addition, the distribution of NMI values across different laboratories is considerably more stable for the WSICS algorithm.

To evaluate the color constancy of the eosin staining we used part of the lymph node dataset that was used in experiment 1. Provided that automatic segmentation of eosin stained tissue structures is more complicated, we evaluate the color constancy of eosin staining within regions that were manually annotated as such. The results are shown in Table III. Note that the SDN and CVN are computed over all the 25 images from the 5 laboratories. Overall, WSICS yielded the smallest SDN and CVN.

The computation time for each algorithm to standardize an image of size  $1388 \times 1040$  (averaged over 20 images) is presented in Table IV. The computation time required by WSICS

TABLE III  
STANDARD DEVIATION AND COEFFICIENT OF VARIATION OF NMI FOR THE EOSIN DYE FOR THE 25 IMAGES CONSIDERED.

Method	NMI SD	NMI CV
<b>Original</b>	0.0563	0.0748
<b>Bagci [24]</b>	0.0200	0.0247
<b>Macenko [11]</b>	0.0362	0.0439
<b>Khan [12]</b>	0.0434	0.0555
<b>WSICS</b>	<b>0.0191</b>	<b>0.0220</b>

TABLE IV  
COMPUTATION TIME (IN SECONDS) FOR PRESENTED METHODS FOR STANDARDIZING AN IMAGE OF SIZE  $1388 \times 1040$ .

Method	Macenko [11]	Khan [12]	Bagci [24]	WSICS
<b>Processing time</b>	1.89	123.73	0.87	18.94

to create a look-up table for standardizing a WSI of lymph node is  $10 \pm 1$  minutes.

3) *Experiment 3*: In this experiment, the performance of an already published CAD system [20] for quantifying necrosis was evaluated before and after standardization of the slides. A dataset comprising three batches of H&E stained histological WSIs of rat liver sections with different amounts of confluent necrosis was available. The CAD system described in [20] using a Random Forest classifier was used to detect necrotic tissue in WSI. This system utilizes local binary pattern (LBP) [29] and pixel value statistics features for each of the individual channels of the RGB and HSV color models. The performance of this system was assessed using a leave-one-batch-out cross-validation scheme. At each cross-validation round, CAD was trained with all slides from two of the batches and validated on the third batch. The same assessment was carried out after standardizing the entire slides using one of the slides in the training set as template image. The performance of the CAD system was then assessed for each cross-validation round in terms of the area under the receiver operating characteristic (ROC) curve [30] at the patch level of size  $256 \times 256$ . Fig. 7 shows the ROC curves for each cross-validation round. The area under ROC curves with and without employing WSICS were compared using the bootstrap test [31]. This test was used to test the null hypothesis that the CAD system performs equally well with and without standardization, versus the alternative hypothesis that it does not. The AUC results for each cross-validation round and the corresponding p-values are summarized in Table V. The p-value for the test was smaller than 0.01 for all the 3 ROC curve pairs,

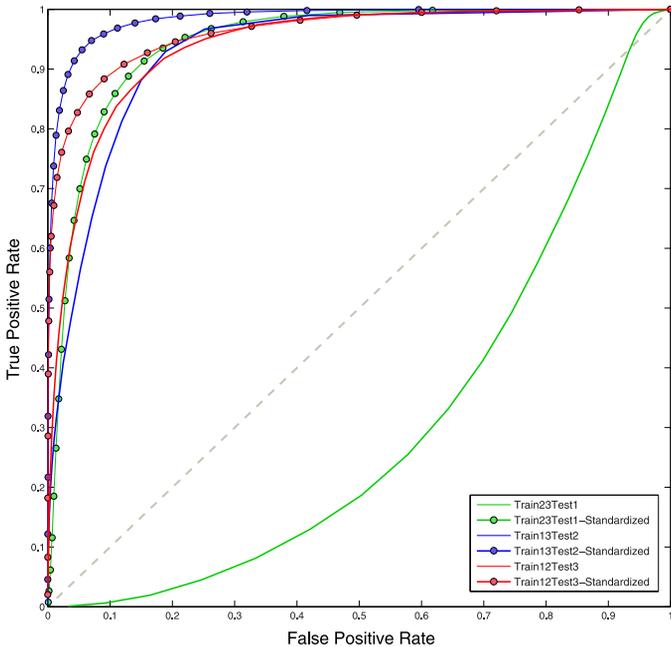


Fig. 7. Performance of the CAD system with and without standardization for different cross-validation rounds.

TABLE V  
SUMMARY OF THE AUC RESULTS FOR EACH CROSS-VALIDATION ROUND OF EXPERIMENT 3 AND THE CORRESPONDING p-VALUES OF THE BOOTSTRAP TEST TO COMPARE AUCs.

Train set	Test set	AUC	AUC with WSICS	p-value
1 and 2	3	0.939	0.963	< 0.01
1 and 3	2	0.930	0.985	< 0.01
2 and 3	1	0.310	0.944	< 0.01

thus providing evidence that CAD performance is increased by applying the WSICS algorithm.

#### IV. DISCUSSION AND CONCLUSION

In this paper, we presented a novel algorithm, called whole-slide image color standardizer (WSICS), for standardization of whole-slide histopathological images. We showed that the WSICS algorithm outperforms previously published algorithms. Even more importantly, we showed that the algorithm was capable of significantly improving the quality of an existing CAD system, rendering it applicable even for specimens exhibiting staining characteristics which strongly deviated from the specimens it was trained on.

The main characteristics of our algorithm are: 1) fully automated detection of stain components in WSIs enabling unsupervised operation, 2) robust stain classification by making use of spatial information, 3) the *HSD* color model for transformation of chromatic and density distributions, and 4) application of weights to create a smooth and artifact-free whole-slide standardization. The algorithm was shown to be very robust against all sources of staining variation.

Our algorithm for classification of tissue components avoids manual training of the classifier by learning samples of each class from the image at hand using prior shape information. This

yields fully automated, objective and reproducible classification results on image data with various sources of variation. Unlike the algorithms in the literature which rely solely on color information to identify stain components our algorithm incorporates spatial information which makes it significantly more robust. Fig. 4 compares the classification result obtained by our proposed algorithm and the algorithm by Khan *et al.* Our approach performed remarkably well on the entire image data provided by 5 different (academic and non-academic) laboratories containing severe staining variation. We observed no classification failure over the entire dataset. In the classification approach in [10], in contrast, the automatically extracted reference stain vectors had to be replaced in more than 10% of the images due to segmentation failures. Our previous experience with unsupervised approaches such as EM-based segmentation [14], [15] shows that segmentation may occasionally fail. This failure is mainly due to the cases in which the chromatic distribution of the channels overlap significantly. The algorithm proposed by Macenko *et al.* [11], moreover, becomes unstable in images with poor contrast and insufficient data for each stain.

Existing color transformation approaches align different channels of a particular color model, independently, using separate 1D transformations [15], [10], [13]. These approaches, however, assume that the channels of the color model utilized are independent which generally does not hold. The use of a more advanced color model called *HSD* which was specifically designed for transmission light microscopy enables independent transformation of chromatic and density information. We performed non-rigid registration of 2D chromatic distribution using several transformations which provides better aligning of the color information compared to separate 1D transformation of each channel. The use of class-dependent weights in combining these transformations yields a smooth standardization result. Consequently, our method consistently shows improved color constancy compared to existing methods.

We presented 3 experiments to evaluate the efficacy of the WSICS method. In the first experiment, we presented a quantitative, comparative evaluation of the performance of the proposed algorithm versus two state-of-the-art algorithms in extracting stain vectors by Macenka *et al.* [11] and Khan *et al.* [12]. The results show that for both hematoxylin and eosin dyes, the derived stain vectors by the WSICS algorithm are substantially more accurate and highly comparable with the stain vectors computed from manually annotated regions in the image.

In the second experiment we presented qualitative and quantitative evaluations of our method relative to three state-of-the-art methods: global standardization (GS) by Bagci *et al.* [24], the appearance normalization by Macenko *et al.* [11], and the non-linear mapping approach by Khan *et al.* [12]. Qualitative assessment of the results show the efficacy of our algorithm in enhanced color constancy of the histology images.

The results shown in Fig. 5 demonstrate that the images with severe staining variation can be standardized to resemble the template image using the WSICS algorithm. Compared to the three state-of-the-art methods, our method performs considerably better in standardizing eosin staining which is in correspondence with the results achieved in experiment 1. The algorithm devised by Macenko [11] yields poor result with artifacts in the

third example which is due to wrong estimation of the stain vectors. This algorithm tries to find the fringe of pixel distribution in the optical density space to determine the stain vectors. As seen in the third example shown in Fig. 5(a) there is a significant difference in the staining of blood cells compared to cytoplasmic/stromal staining. The poor contrast between nuclear staining and cytoplasmic/stromal staining has led to poor standardization results by this method. The Global Standardization algorithm, on the other hand, achieves smooth standardization output without any artifacts (see Fig. 5(d)). However, the algorithm is clearly unable to match the staining quality to the template image. This is mainly due to the usage of a single transformation function for standardizing the image which does not correspond with the existence of multiple components in the data. The algorithm proposed by Khan *et al.* [12] yields artifact-free standardized images. However, the quality of the standardized images by this method significantly deviate from the template image. This algorithm uses an RVM classifier to classify the pixels into different stain components in the image. Using a pre-trained classifier makes this algorithm unstable in case where the color of the dyes in the test specimen deviate from the image batches that the RVM classifier was initially trained on. As a result, the estimation of stain vectors may fail. This was also observed in experiment 1. As shown in Fig. 4, heavy pollution of the hematoxylin population with pixels from connective tissue and cytoplasm leads to wrong estimation of the stain vectors by this algorithm. As a result, all the eosin stained structures will have a purplish-blue color after standardization. Contrary to this approach, our algorithm effectively defines the hematoxylin and eosin distributions by incorporating spatial information. This, in turn, leads to better standardization performance by our proposed algorithm.

The results of the quantitative assessment, as summarized in Tables II and III, show that for the entire lymph node dataset stained in 5 different laboratories, the algorithm proposed here outperformed the competing standardization methods by yielding the lowest standard deviation and coefficient of variation of the NMI measure.

The application of whole-slide standardization to computer-aided diagnosis of histopathology data has so far remained elusive in the literature due to technical complexities in dealing with whole-slide images. The focus has been limited to investigating the contribution of standardization algorithms that work at the small image patch level. In the third experiment, we evaluated the impact of our proposed whole-slide standardization algorithm on the performance of a necrosis quantification CAD system. The performance of the CAD system for quantifying necrosis was assessed in a leave-one-batch-out cross validation experiment, before and after standardization. The performance of the CAD system was better in all cross-validation rounds after utilizing the WSICS algorithm. In particular, the performance was significantly improved for the case that the CAD system was trained on batch 2 and 3 and tested on the first batch. The ROC curve shown in Fig. 7, illustrates that the performance without standardization is worse than random guessing. The reason for this can be related to the significant difference in the staining color of the slides in batch 1 relative to other batches which results in the viable tissue having inten-

sity and color ranges similar to necrotic tissue in other batches, and vice-versa. This is, however, effectively addressed using the proposed method. Hence, the substantial impact of employing the proposed whole-slide standardization algorithm on the performance of an already published CAD system further demonstrates its efficacy and reliability.

One limitation of the current study is that although the WSICS algorithm has been designed to standardize whole-slide images, the comparison of the color constancy of the images standardized by different algorithms in experiment 2 is limited to image patches (sub-images from the WSI). This is because the competing algorithms have been designed to standardize patch images only.

The WSICS algorithm has been specifically described for standardization of H&E stained images. The algorithm, however, can be adapted to work with other histological staining techniques such as immunohistochemistry (IHC). Identification of the stain components is more straight forward in IHC. The major reason is that in IHC, in contrast to H&E staining, the chromatic distribution of the stain components have a small overlap in the chromatic plain of the *HSD* model. In future work, we will concentrate on adapting the proposed algorithm to be utilized on other staining techniques and investigate the possibility of extending our method for standardizing WSIs with more than two stain components.

#### ACKNOWLEDGMENT

The authors wish to acknowledge support from the histology laboratory of the Department of Pathology, Radboud University Medical Center, Nijmegen, the Netherlands and Prof. Dr. med. U. Dahmen and Dr. med. O. Dirsch, University Hospital Jena, Jena, Germany.

#### REFERENCES

- [1] F. Ghaznavi, A. Evans, A. Madabhushi, and M. Feldman, "Digital IMAGING in pathology: Whole-slide IMAGING and beyond," *Annu. Rev. Pathol., Mechan. Disease*, vol. 8, pp. 331–359, 2013.
- [2] M. N. Gurcan *et al.*, "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.
- [3] S. M. Ismail *et al.*, "Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia," *Br. Med. J.*, vol. 298, no. 6675, p. 707, 1989.
- [4] A. Andrion *et al.*, "Malignant mesothelioma of the pleura: Interobserver variability," *J. Clin. Pathol.*, vol. 48, no. 9, pp. 856–860, 1995.
- [5] J. Epstein, W. Allsbrook, Jr., M. Amin, and L. Egevad, "Update on the gleason grading system for prostate cancer: Results of an international consensus conference of urologic pathologists," *Adv. Anatom. Pathol.*, vol. 13, no. 1, p. 57, 2006.
- [6] M. H. Stoler *et al.*, "Interobserver reproducibility of cervical cytologic and histologic interpretations: Realistic estimates from the ascus-lsil triage study," *JAMA*, vol. 285, no. 11, pp. 1500–1505, 2001.
- [7] C. A. Roberts *et al.*, "Interpretive disparity among pathologists in breast sentinel lymph node evaluation," *Am. J. Surg.*, vol. 186, no. 4, pp. 324–329, 2003.
- [8] J. D. Bancroft and M. Gamble, *Theory and Practice of Histological Techniques*. New York: Elsevier Health Sci., 2008.
- [9] 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: Springer, 2010, pp. 58–66.
- [10] D. Magee *et al.*, "Colour normalisation in digital histopathology images," in *Proc. Optical Tissue Image Anal. Microsc., Histopathol. Endosc. MICCAI Workshop*, 2009, pp. 20–24.
- [11] M. Macenko *et al.*, "A method for normalizing histology slides for quantitative analysis, *emph.*" in *Proc. IEEE Int. Symp. Biomed. Imag., From Nano to Macro*, 2009, pp. 1107–1110.

- [12] A. Khan, N. Rajpoot, D. Treanor, and D. Magee, "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1729–1738, Jun. 2014.
- [13] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 34–41, Sep. 2001.
- [14] B. E. Bejnordi, N. Timofeeva, I. Otte-Höller, N. Karssemeijer, and J. A. van der Laak, "Quantitative analysis of stain variability in histology slides and an algorithm for standardization," *Proc. SPIE Med. Imag.*, pp. 904 108–904 108, 2014.
- [15] A. Basavanthally and A. Madabhushi, "EM-based segmentation-driven color standardization of digitized histopathology," in *Proc. SPIE Med. Imag.*, 2013, pp. 86 760G–86 760G.
- [16] P. A. Bautista, N. Hashimoto, and Y. Yagi, "Color standardization in whole slide IMAGING using a color calibration slide," *J. Pathol. Inf.*, vol. 5, 2014.
- [17] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Analyt. Quant. Cytol. Histol. Int. Acad. Cytol. Am. Soc. Cytol.*, vol. 23, no. 4, pp. 291–299, 2001.
- [18] R. S. Hunter, "Accuracy, precision, and stability of new photoelectric color-difference meter," *J. Opt. Soc. Am.*, vol. 38, no. 12, pp. 1094–1094, 1948.
- [19] J. A. van der Laak, M. M. Pahlplatz, A. G. Hanselaar, and P. de Wilde, "Hue-saturation-density (HSD) model for stain recognition in digital images from transmitted light microscopy," *Cytometry*, vol. 39, no. 4, pp. 275–284, 2000.
- [20] A. Homeyer *et al.*, "Practical quantification of necrosis in histological whole-slide images," *Comput. Med. Imag. Graph.*, vol. 37, no. 4, pp. 313–322, 2013.
- [21] G. Litjens, Automated Slide Analysis Platform (ASAP) 2015 [Online]. Available: <https://github.com/GeertLitjens/ASAP>
- [22] A. Goode, B. Gilbert, J. Harkes, D. Jukic, and M. Satyanarayanan, "Openslide: A vendor-neutral software foundation for digital pathology," *J. Pathol. Informat.*, vol. 4, 2013.
- [23] Z. Cheng and Y. Liu, "Efficient technique for ellipse detection using restricted randomized Hough transform," in *Proc. Int. Conf. Inf. Technol., Coding Comput.*, 2004, vol. 2, pp. 714–718.
- [24] U. Bagci and L. Bai, "Registration of standardized histological images in feature space," *Proc. SPIE Med. Imag.*, vol. 6914, pp. 69 142V–1, 2008.
- [25] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 959–973, Aug. 2003.
- [26] R. Moshavegh *et al.*, "Automated segmentation of free-lying cell nuclei in Pap smears for malignancy-associated change analysis," in *Proc. Annu. Int. Conf. IEEE EMBC*, 2012, pp. 5372–5375.
- [27] M. Peura and J. Iivarinen, "Efficiency of simple shape descriptors," in *Proc. 3rd Int. Workshop Vis. Form*, 1997, vol. 443, p. 451.
- [28] L. G. Nyúl, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans. Med. Imag.*, vol. 19, no. 2, pp. 143–150, Feb. 2000.
- [29] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [30] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine," *Clin. Chem.*, vol. 39, no. 4, pp. 561–577, 1993.
- [31] J. A. Hanley and B. J. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases," *Radiology*, vol. 148, no. 3, pp. 839–843, 1983.