

# EM-Based Segmentation-Driven Color Standardization of Digitized Histopathology

Ajay Basavanhally<sup>a</sup> and Anant Madabhushi<sup>b</sup>

<sup>a</sup>Rutgers, The State University of New Jersey, Piscataway, NJ, USA;

<sup>b</sup>Case Western Reserve University, Cleveland, OH, USA;

## ABSTRACT

The development of tools for the processing of color images is often complicated by nonstandardness – the notion that different image regions corresponding to the same tissue will occupy different ranges in the color spectrum. In digital pathology (DP), these issues are often caused by variations in slide thickness, staining, scanning parameters, and illumination. Nonstandardness can be addressed via standardization, a pre-processing step that aims to improve color constancy by realigning color distributions of images to match that of a pre-defined template image. Unlike color normalization methods, which aim to scale (usually linearly or assuming that the transfer function of the system is known) the intensity of individual images, standardization is employed to align distributions in broad tissue classes (e.g. epithelium, stroma) across different DP images irrespective of institution, protocol, or scanner. Intensity standardization has previously been used for addressing the issue of intensity drift in MRI images, where similar tissue regions have different image intensities across scanners and patients. However, this approach is a global standardization (GS) method that aligns histograms of entire images at once. By contrast, histopathological imagery is complicated by the (a) additional information present in color images and (b) heterogeneity of tissue composition. In this paper, we present a novel color Expectation Maximization (EM) based standardization (EMS) scheme to decompose histological images into independent tissue classes (e.g. nuclei, epithelium, stroma, lumen) via the EM algorithm and align the color distributions for each class independently. Experiments are performed on prostate and oropharyngeal histopathology tissues from 19 and 26 patients, respectively. Evaluation methods include (a) a segmentation-based assessment of color consistency in which normalized median intensity (NMI) is calculated from segmented regions across a dataset and (b) a quantitative measure of histogram alignment via mean landmark distance. EMS produces lower NMI standard deviations (i.e. greater consistency) of 0.0054 and 0.0034 for prostate and oropharyngeal cohorts, respectively, than unstandardized (0.034 and 0.026) and GS (0.031 and 0.017) approaches. Similarly, we see decreased mean landmark distance for EMS (2.25 and 4.20) compared to both unstandardized (54.8 and 27.3) and GS (27.1 and 8.8) images. These results suggest that the separation of broad tissue classes in EMS is vital to the standardization of DP imagery and subsequent development of computerized image analysis tools.

## 1. INTRODUCTION

The increasing popularity of digital pathology (DP) in recent years has been closely coupled with the deployment of whole-slide scanners in both clinical and research settings. The rapid growth of these scanners has resulted in (1) the development of computerized image analysis tools, including algorithms for object detection and segmentation,<sup>1</sup> and (2) the advent of virtual microscopy for simplifying visual analysis and telepathology for remote diagnosis.<sup>2</sup> Such tasks, however, are complicated by color nonstandardness – the phenomenon that causes similar objects to exhibit different color properties across images. Nonstandardness in medical imaging has previously been explored in the context of intensity drift in MR imaging and its relationship to the bias field artifact.<sup>3</sup> By contrast, differences in color distributions across DP images are more complex, resulting from variations in tissue thickness, staining, and illumination<sup>4</sup> (Figure 1).

Previous work in maintaining color constancy in DP images has traditionally employed normalization and calibration techniques. Normalization, the process by which color distributions of images are adjusted to fit a predetermined range, is performed independently for each image. Other approaches to normalization include

---

Corresponding author: Anant Madabhushi (anantm@case.edu)

analysis in alternate color spaces (e.g. hue-saturation-intensity) that are more invariant to the effects of color nonstandardness.<sup>5</sup> However, it is difficult to account for the non-linear intensity variations arising from the many sources of color nonstandardness. Color calibration refers to the modification of acquisition or visualization settings based on prior knowledge of imaging parameters. For instance, Yagi et al.<sup>4</sup> performed calibration of computer monitors for optimal viewing of DP; yet calibration is unfeasible for the analysis of retrospective studies on existing data cohorts. Note that the larger body of work aimed at correcting variations in illumination for images formed by reflective light (e.g. digital photography) is inappropriate for DP images that are formed by light absorption.<sup>6,7</sup>

Unlike normalization and calibration, color standardization aims to improve color constancy across a population of histology images by realigning color distributions to match a pre-defined template image. Previous applications of standardization in medical imaging include correcting intensity drift in MRI images.<sup>8,9</sup> Nyul et al.<sup>8</sup> developed a piecewise intensity standardization method that employed linear interpolation to align distributions between histogram landmarks at evenly spaced percentiles. Madabhushi et al.<sup>9</sup> further extended the model by implicitly incorporating basic spatial information via the generalized scale model; yet, their approach was limited to a connected component labeling that (1) has no particular correspondence between DP images and (2) cannot be used for tissue classes (e.g. nuclei) spread across many regions. In addition, intensity standardization has been limited to global approaches that employ a single histogram characterizes an entire image.<sup>8,9</sup> This type of global standardization (GS) is unable to account for the heterogeneity of DP images, which contain broad, independent tissue classes (e.g. stroma, epithelium, nuclei, lumen) in varying proportions, leading to skewed color distributions and errors in the standardization process.

Previous work in the intensity standardization of MRI images has yielded a number of methods for evaluating the quality of the standardization result.<sup>3,8,10</sup> Nyul et al.<sup>8</sup> described the calculation of normalized mean intensity for distinct regions in brain MRI (e.g. white matter, gray matter) and demonstrated the importance of reporting the corresponding coefficient of variation across an entire dataset. The calculation of normalized mean intensity was more closely attached to the standardization result by Madabhushi et al.,<sup>10</sup> who employed a single predefined threshold to segment all images in the dataset. Since the threshold was selected for optimal segmentation of the template image, increased consistency in the segmentation result could be directly attributed to the standardization process. More recently, Palumbo et al.<sup>3</sup> directly evaluated alignment between two histograms via a “difference-of-modes” measure. While this approach does quantify histogram alignment, it operates under the assumptions that both histograms (1) are modeled by the same underlying distribution and (2) contain a single dominant mode.

The rest of the paper is organized as follows. Section 2 details the novel contributions of this work and the methods used to develop EMS are described in Section 3. Section 4 outlines the experiments used to evaluate EMS against both global standardization and unstandardized images. Both quantitative and qualitative results are presented in Section 5 followed by concluding remarks in Section 6.

## 2. BRIEF OVERVIEW AND NOVEL CONTRIBUTIONS

In this paper, Expectation Maximization (EM) based segmentation-driven standardization (EMS) is employed to improve color constancy across multiple prostate and oropharyngeal histopathology images. In order to mitigate the challenges associated with tissue heterogeneity, each image is first partitioned into broad tissue classes (e.g. nuclei, epithelium, stroma, lumen) using the EM algorithm<sup>11</sup> – an unsupervised clustering algorithm that separates image pixels into a pre-determined number of Gaussian components (Figure 2). Corresponding tissue classes across different images are automatically matched by minimizing pairwise distances in the RGB color space between all classes. Histograms are constructed using pixels from each tissue class of a *test* image and aligned to the corresponding tissue class in the *template* image (Algorithm 1). For comparison, evaluation is also performed on globally standardized (GS) images, in which color distributions are aligned for entire images without isolating tissue classes.

Novel contributions of EMS are that it

- Aligns color distributions of broad tissue classes (e.g. nuclei, stroma) that are first partitioned via EM; by contrast, previous global methods perform standardization using a histogram of the entire image.

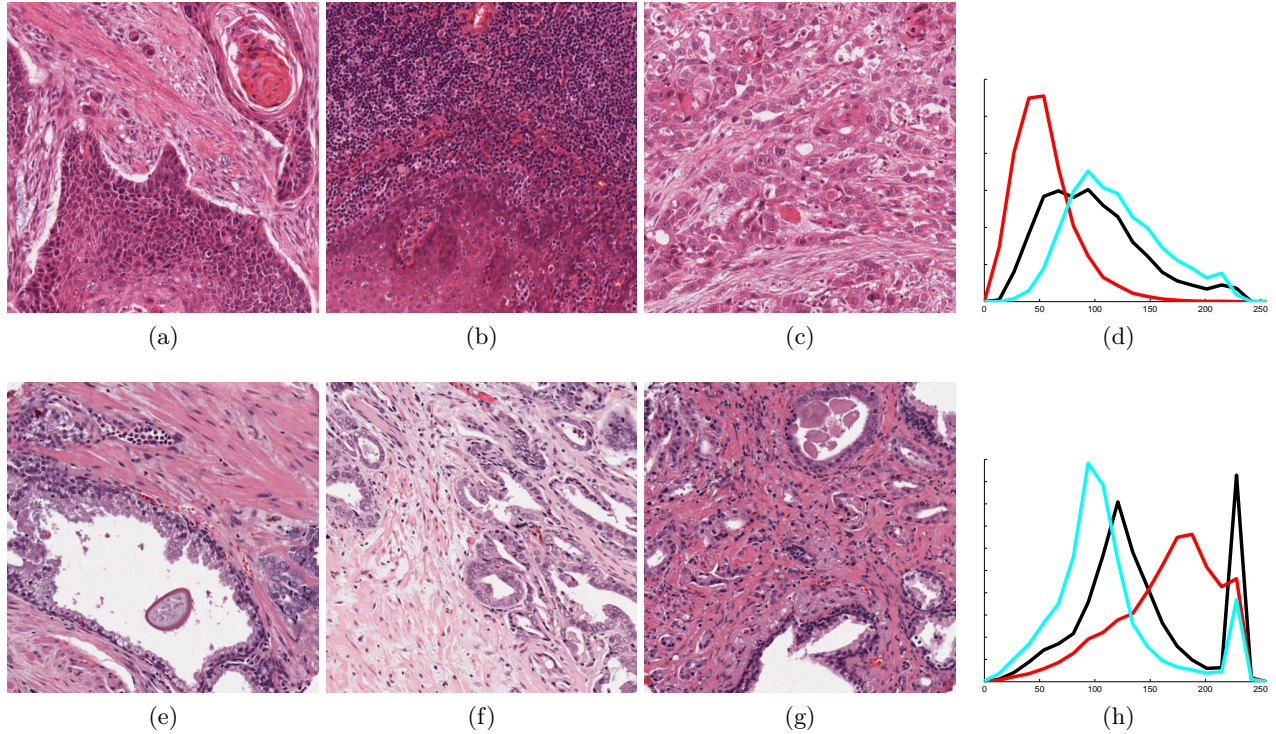


Figure 1. H & E stained histopathology images of (a)-(c) oropharyngeal and (e)-(g) prostate cancers demonstrate color nonstandardness across tissue specimens resulting from variations in slide preparation (e.g. tissue thickness and staining). These variations are reflected by (d), (h) corresponding histograms of the green color channel, in which each image occupies different ranges of the color spectrum.

- Can be used retrospectively since it is independent of scanners, staining protocols, and institutions.

### 3. METHODS

For all methods, an image scene  $\mathcal{C}_a = (C, \mathbf{f})$  is a 2D set of pixels  $c \in C$  and  $\mathbf{f}$  is the associated function that assigns RGB values. In addition, the function  $g$  is used to assign intensity values (from the hue-saturation-intensity color space) for evaluation purposes.

#### 3.1 EM-based Partitioning of Broad Tissue Classes

The EM framework is employed to first partition each image into broad tissue classes. First, the pixels in each image scene  $\mathcal{C}_a$  are modeled as a Gaussian mixture of  $\kappa$  components, where  $K \in \{1, 2, \dots, \kappa\}$ . We optimize the model parameter set  $\gamma^i = \{\mu_K^i, \sigma_K^i, \mathbf{p}_K^i : \forall K\}$ , comprising the mean  $\mu_K$ , covariance  $\sigma_K$ , and *a priori* probability  $\mathbf{p}_K$  at iteration  $i$ . The mixture is initialized to  $\gamma^0$  via  $\kappa$ -means clustering of RGB values over all  $c \in C$ . The Expectation step calculates the posterior probability

$$p^i(K|\mathbf{f}(c)) = \frac{\mathbf{p}_K^i N(\mathbf{f}(c)|\mu_K^i, \sigma_K^i)}{\sum_{j=1}^{\kappa} \mathbf{p}_K^i N(\mathbf{f}(c)|\mu_j^i, \sigma_j^i)},$$

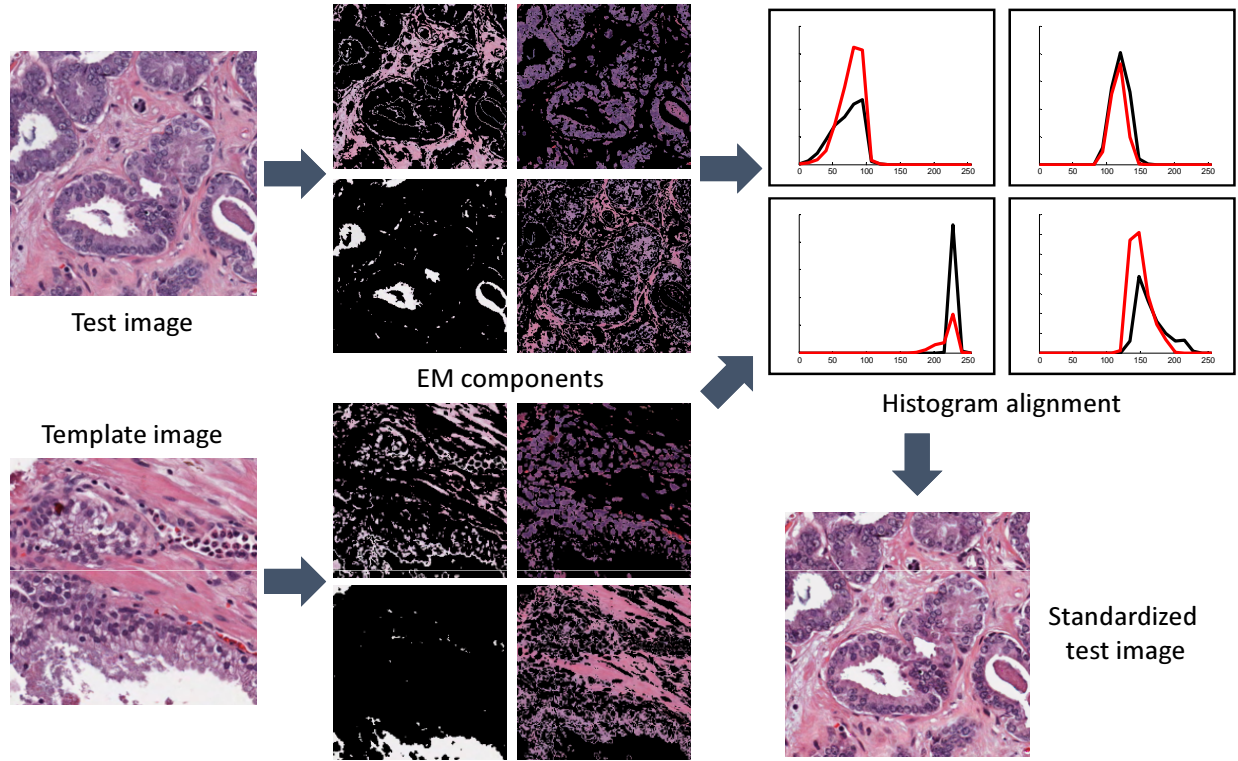


Figure 2. EM-based standardization (EMS) first decomposes the test and template images into a pre-determined number of components via the Expectation Maximization algorithm. The distribution for each component in the test image is aligned independently to the corresponding component in the template image. The standardized components are subsequently recombined to create a standardized test image.

where  $N(\mathbf{f}(c)|\mu_K, \sigma_K)$  represents the value of Gaussian component  $K$  at RGB value  $\mathbf{f}(c)$ . The Maximization step uses  $p^i$  to calculate  $\gamma^{i+1} = \{\mu_K^{i+1}, \sigma_K^{i+1}, \mathbf{p}_K^{i+1}\}$ .<sup>11</sup>

$$\begin{aligned}\mu_K^{i+1} &= \frac{\sum_{c \in C} p(K|\mathbf{f}) \mathbf{f}}{\sum_{c \in C} p(K|\mathbf{f})} \\ \sigma_K^{i+1} &= \frac{\sum_{c \in C} p(K|\mathbf{f}) (\mathbf{f} - \mu_K^i) (\mathbf{f} - \mu_K^i)^T}{\sum_{c \in C} p(K|\mathbf{f})} \\ \mathbf{p}_K^{i+1} &= \frac{1}{|C|} \sum_{c \in C} p(K|\mathbf{f})\end{aligned}$$

The EM algorithm converges when  $\|(\mathcal{L}^{i+1} - \mathcal{L}^i)/\mathcal{L}^i\| < \epsilon$ , where  $\mathcal{L}^i$  is the log likelihood of the Gaussian mixture model with parameters  $\gamma^i$  and  $\epsilon = 10^{-5}$  determined empirically. The appropriate class  $K^* = \operatorname{argmax}_K p(K|\mathbf{f}(c))$  is found for each pixel  $c \in C$  by identifying the maximum posterior probability over all  $K \in \{1, 2, \dots, \kappa\}$ . Hence, we are able to define a subset of pixels  $D_K \subset C$  corresponding to tissue class  $K$  from image scene  $\mathcal{C}_a$ .

### 3.2 Determining Correspondence of Tissue Classes from Different Images

Since the EM algorithm performs tissue partitioning in an unsupervised manner, correspondence between tissue classes is not guaranteed across different images. For instance, the background (i.e. white) regions may coincide with the first EM component in one image and the final component in a second image. In this paper, the identification of a corresponding tissue class between two images is performed automatically. Let mean RGB values  $\mu_{K,a}$  and  $\mu_{K,b}$  be defined for tissue class  $K$  in image scenes  $\mathcal{C}_a$  and  $\mathcal{C}_b$ , respectively. The first pair of matching tissue classes is identified by minimizing the pairwise Euclidean distances between mean RGB values

such that

$$\operatorname{argmin}_{i,j \in \{1,2,\dots,\kappa\}} \|\mu_{i,a} - \mu_{j,b}\|.$$

The matching tissue classes are set aside and this process is subsequently repeated  $\kappa$  times until all tissue classes have been matched.

### 3.3 EM-based Segmentation-Driven Color Standardization (EMS) for Digital Pathology Images

We first describe the generalized scheme for color standardization using landmark-based piecewise linear interpolation<sup>8</sup> followed by an explanation of the EMS approach, which aligns color distributions of the broad tissue classes identified by the EM algorithm in Section 3.1.

#### 3.3.1 Generalized Color Standardization

Let  $\mathcal{D}_a \subset \mathcal{C}_a$  and  $\mathcal{D}_b \subset \mathcal{C}_b$  correspond to sub-scenes in test image  $\mathcal{C}_a$  and template image  $\mathcal{C}_b$ . For a single color channel (i.e. red, green, or blue), landmarks  $\{r_{\min}, r_{10}, r_{20}, \dots, r_{90}, r_{\max}\}$  and  $\{s_{\min}, s_{10}, s_{20}, \dots, s_{90}, s_{\max}\}$  are defined at the minimum and maximum, as well as evenly-spaced percentiles  $\{10, 20, \dots, 90\}$ , of all pixel values in  $\mathcal{D}_a$  and  $\mathcal{D}_b$ , respectively. Pixel values from the test image in the range  $[r_{\min}, r_{10}]$  are interpolated to match the corresponding landmarks  $[s_{\min}, s_{10}]$  in the template image. After repeating this process for all sets of adjacent landmarks in all color channels, the standardized pixels are recombined to construct a standardized test scene  $\hat{\mathcal{D}}_a$ .

---

#### Algorithm 1 GeneralizedStandardization()

---

**Input:** Template image  $\mathcal{D}_b$ . Test image  $\mathcal{D}_a$  to be standardized.

**Output:** Standardized image  $\hat{\mathcal{D}}_a$ .

- 1: **for** RGB channels  $i \in \{R, G, B\}$  in  $\mathcal{D}_a$  and  $\mathcal{D}_b$  **do**
  - 2:   Define  $\{r_{\min}, r_{10}, r_{20}, \dots, r_{90}, r_{\max}\}$  as landmarks in  $\mathcal{D}_a$ .
  - 3:   Define  $\{s_{\min}, s_{10}, s_{20}, \dots, s_{90}, s_{\max}\}$  as landmarks in  $\mathcal{D}_b$ .
  - 4:   Interpolate pixel values from range  $[r_{\min}, r_{10}]$  to range  $[s_{\min}, s_{10}]$ . Repeat for all sets of adjacent landmarks.
  - 5: **end for**
  - 6: Recombine standardized RGB channels to construct standardized image  $\hat{\mathcal{D}}_a$ .
- 

#### 3.3.2 Class-Specific Color Standardization of Broad Tissue Classes

Algorithm 2 shows how EMS extends the generalized standardization approach by incorporating prior domain knowledge of tissue composition in DP (Algorithm 2). First, the EM algorithm is applied to partition each image into  $\kappa$  tissue classes (Section 3.1) and corresponding tissue classes between test and template images are automatically matched (Section 3.2). For each tissue class, pixels from the test and template images are standardized using the piecewise linear interpolation method presented in Algorithm 1. Subsequently, standardized pixels from all tissue classes are recombined to create a standardized test image.

## 4. EXPERIMENTAL DESIGN

### 4.1 Data Cohort

The EM-based color standardization scheme is evaluated on digitized hematoxylin & eosin stained histopathology images from independent prostate (N=19) and oropharyngeal (N=26) cohorts, in which each image was taken from a different patient (Table 1). All images were digitized via a whole slide scanner at a spatial resolution of  $1 \mu\text{m}/\text{pixel}$  and cropped to be  $500 \times 500$  pixels in size. Both cohorts were empirically determined to have  $\kappa = 4$  broad tissue classes corresponding to nuclei, epithelium, stroma, and background (i.e. white space). In addition, one image from each cohort is designated as a template image to which all other (test) images are aligned (Figure 3). It is important to note that a single “ideal” template image does not exist for all applications; instead, the template is selected based on its performance in terms of the desired task (e.g. nuclear segmentation).

---

**Algorithm 2** EMbasedStandardization()

---

**Input:** Template image  $\mathcal{C}_b$ . Test image  $\mathcal{C}_a$  to be standardized. Number of EM components  $\kappa$ .

**Output:** Standardized image  $\mathcal{C}'_a$ .

- 1: Apply EM algorithm to partition pixels from both  $\mathcal{C}_a$  and  $\mathcal{C}_b$  into  $\kappa$  tissue classes.
- 2: Pair matching tissue classes between  $\mathcal{C}_a$  and  $\mathcal{C}_b$ .
- 3: **for**  $K \in \{1, 2, \dots, \kappa\}$  **do**
- 4:   Define sub-scenes  $\mathcal{D}_a^K \subset \mathcal{C}_a$  and  $\mathcal{D}_b^K \subset \mathcal{C}_b$  corresponding to EM component  $K$ .
- 5:   Perform GeneralizedStandardization() using  $\mathcal{D}_a^K$  and  $\mathcal{D}_b^K$  as test and template images, respectively (Alg. 1).
- 6: **end for**
- 7: Create standardized image  $\mathcal{C}'_a = \{\mathcal{C}'_a^K : \forall K \in \{1, 2, \dots, \kappa\}\}$  by recombining standardized sub-scenes from all  $\kappa$  components of the test image.

---

Cohort	# images	Staining	Resolution	Size
Prostate	19	hematoxylin & eosin	1 $\mu\text{m}/\text{pixel}$	500 $\times$ 500 pixels
Oropharyngeal	26			

Table 1. A description of the prostate and oropharyngeal data cohorts used in this paper.

## 4.2 Comparative Strategy: Global Standardization

In addition to the comparison against unstandardized images, the ability of EMS to align color distributions is evaluated against global standardization (GS). GS is a straightforward approach to color standardization that does not account for the heterogeneous tissue structure in DP images. Instead of partitioning each image into multiple tissue classes, a single histogram is constructed from all pixels in a test image and aligned to the entire histogram from the template image. Specifically, the GS approach can be considered a modified application of GeneralizedStandardization() from Algorithm 1 using entire image scenes  $\mathcal{C}_a$  and  $\mathcal{C}_b$  used as the test and template images, respectively.

## 4.3 Performance Evaluation Measures

### 4.3.1 Qualitative Segmentation Consistency Across Images

A qualitative evaluation of color standardization in DP images is performed by observing the consistency of tissue segmentation across several images in a cohort.<sup>9</sup> For image  $\mathcal{C}_a$ , we segment pixels corresponding to nuclei  $D = \{c : c \in \mathcal{C}, g(c) \in [0, \psi]\}$ , where  $\psi$  is a threshold in the intensity channel  $g$  from the hue-saturation-intensity (HSI) color space. Given intensities that occupy the range  $g(c) \in [0, 255]$ , thresholds of  $\psi = 115$  for the prostate cohort and  $\psi = 145$  for the oropharyngeal cohort were selected empirically for their ability to provide a basic nuclear segmentation in their respective template images (Figure 3). Visually, images from a standardized cohort should yield a more consistent segmentation of nuclei (in comparison to the template image) than the original set of unstandardized images.

### 4.3.2 Quantitative Evaluation via Normalized Median Intensity

The segmentation results from Section 4.3.1 can also be evaluated quantitatively via normalized median intensity (NMI), which is employed to characterize color constancy from a segmented region across all images in a dataset. Using the segmented nuclear region  $D \in \mathcal{C}$  for an image  $\mathcal{C}_a$ , NMI is defined as

$$\text{median} \left( \frac{g(D)}{\max g(D)} \right),$$

where  $g(D)$  is the set of intensities for all pixels isolated by the thresholding process. Intensity values across all images in a cohort are considered to be more consistent as (1) the standard deviation and (2) the coefficient of variation (i.e. standard deviation divided by mean) of NMI decreases.

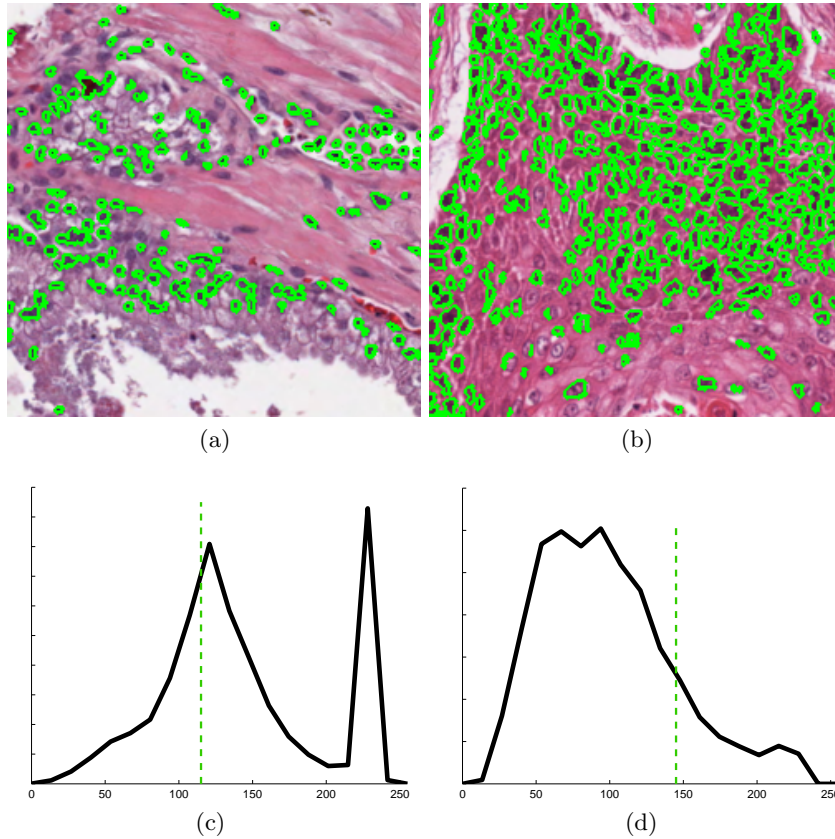


Figure 3. The template images selected for (a) prostate and (b) oropharyngeal cohorts are shown along with segmented nuclei (green outline). (c), (d) Corresponding green color channel distributions are shown along with a dotted green line denoting the location of the empirically-selected threshold used to segment the nuclei.

### 4.3.3 Quantitative Evaluation via Histogram Landmark Distance

Another measure of improved standardization is the distance between corresponding histogram landmarks (i.e. percentiles  $\{10, 20, \dots, 90\}$ ) in the template and test images (Figure 4), whereby histograms are considered to have improved alignment as the mean landmark distance decreases. Using the notation defined in Algorithm 1, the mean histogram landmark distance between a test image  $\mathcal{C}_a$  and template image  $\mathcal{C}_b$  can be defined as  $\phi(a, b) = \frac{1}{9} \sum_{j \in \{10, \dots, 90\}} \|r_j - s_j\|$ , where  $r_j$  and  $s_j$  are corresponding landmarks in  $\mathcal{C}_a$  and  $\mathcal{C}_b$ , respectively. For a set of  $H$  histograms, pairwise landmark distance  $\Phi = \{\phi(a, b) : \forall a, b \in \{1, \dots, H\}, a \neq b\}$  is calculated between all histograms and mean distance is reported for the cohort.

## 5. RESULTS AND DISCUSSION

### 5.1 Qualitative Evaluation of Consistency in Nuclear Segmentation

Qualitative evaluation is performed by visualizing the effect of standardization on segmentation of nuclei in the test images (Figure 5). The inconsistent segmentation results between the template images (Figures 3(a), (b)) and unstandardized test images (Figures 5(a), (d)) clearly demonstrates the inherent color nonstandardness that affects DP images. A more consistent segmentation of nuclear regions is visible after GS (Figures 5(b), (e)) and is further improved by the application of EMS (Figures 5(c)-(f)). The improvement seen by employing EMS suggests that separation of tissue classes may be vital to the development of algorithms for the segmentation of primitives (e.g. nuclei).

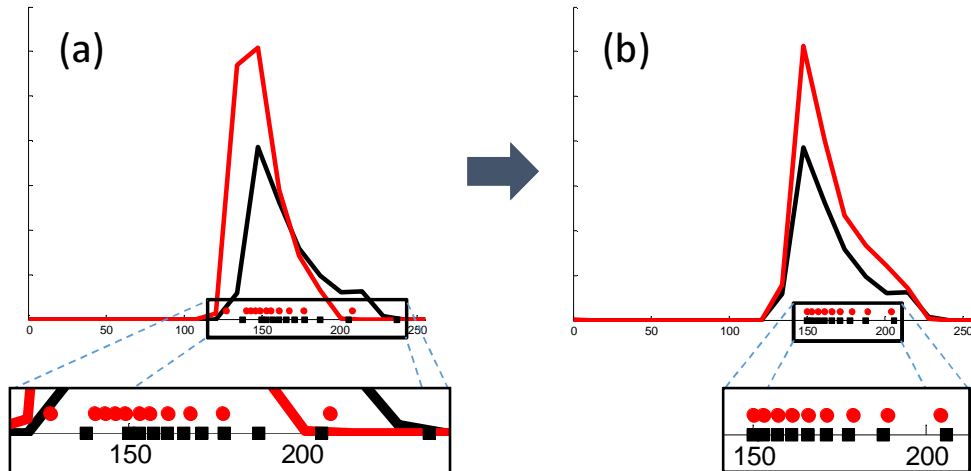


Figure 4. Histograms representing a single EM component are shown for template (black) and test (red) images both (a) before standardization and (b) after EMS has been applied. The insets illustrate the evenly-spaced percentiles  $\{10, 20, \dots, 90\}$  used as landmarks for histogram alignment during the standardization process.

## 5.2 Quantitative Evaluation of Segmentation Consistency via NMI

The qualitative results presented in Figure 5 are also evaluated quantitatively by calculating the normalized median intensity (NMI) of the segmented regions.<sup>8</sup> In terms of NMI, EMS produces improved color constancy compared to the original images, with considerably lower NMI standard deviation (SD) of 0.0054 vs. 0.0338 and NMI coefficient of variation (CV) of 0.0062 vs. 0.0393 in the prostate cohort (Table 2). In addition, EMS yields more consistent results than GS, demonstrating an order of magnitude improvement in SD and CV of 0.0305 and 0.0354, respectively. All corresponding results for the oropharyngeal cohort show similar improvement after standardization.

	Prostate		Oropharyngeal	
	SD	CV	SD	CV
Original	0.0338	0.0393	0.0261	0.0302
GS	0.0305	0.0354	0.0166	0.0193
EMS	<b>0.0054</b>	<b>0.0062</b>	<b>0.0034</b>	<b>0.0039</b>

Table 2. Standard deviation (SD) and coefficient of variation (CV) of normalized median intensity (NMI) is calculated across all images in the prostate and oropharyngeal cohorts.

## 5.3 Quantitative Evaluation of Histogram Landmark Alignment

The performance of EMS is further supported by histograms of the green color channel (from the RGB color space) for both prostate and oropharyngeal cohorts (Figure 6 and 7). Examining the prostate cohort, it is visually clear that unstandardized images have highly misaligned color distributions for both the global histogram (Figure 6(a)) and for the EM component corresponding to nuclei (Figures 6(d)). While both GS and EMS yield improved alignment over unstandardized images, a closer examination of the GS distributions suggests that higher pixel values (denoted by the black dashed rectangle in Figure 6(b)) frequently suffer from poor alignment. This is because GS is unable to account for the large variations in the amount of white space (e.g. luminal areas, adipose tissue, slide background) across histopathology images. By contrast, EMS does not suffer from this issue (Figures 6(c)) since it partitions the white regions in each image and aligns their distributions independently. The histograms in Figure 7 suggest that similar results and trends hold true for the oropharyngeal cohort.

The improved alignment of EMS distributions over both unstandardized and GS distributions is confirmed quantitatively by calculating the histogram landmark distance for individual EM components (Figures 6(d)-(f)). Using the EM component corresponding to nuclei in the prostate histopathology images, EMS yields a



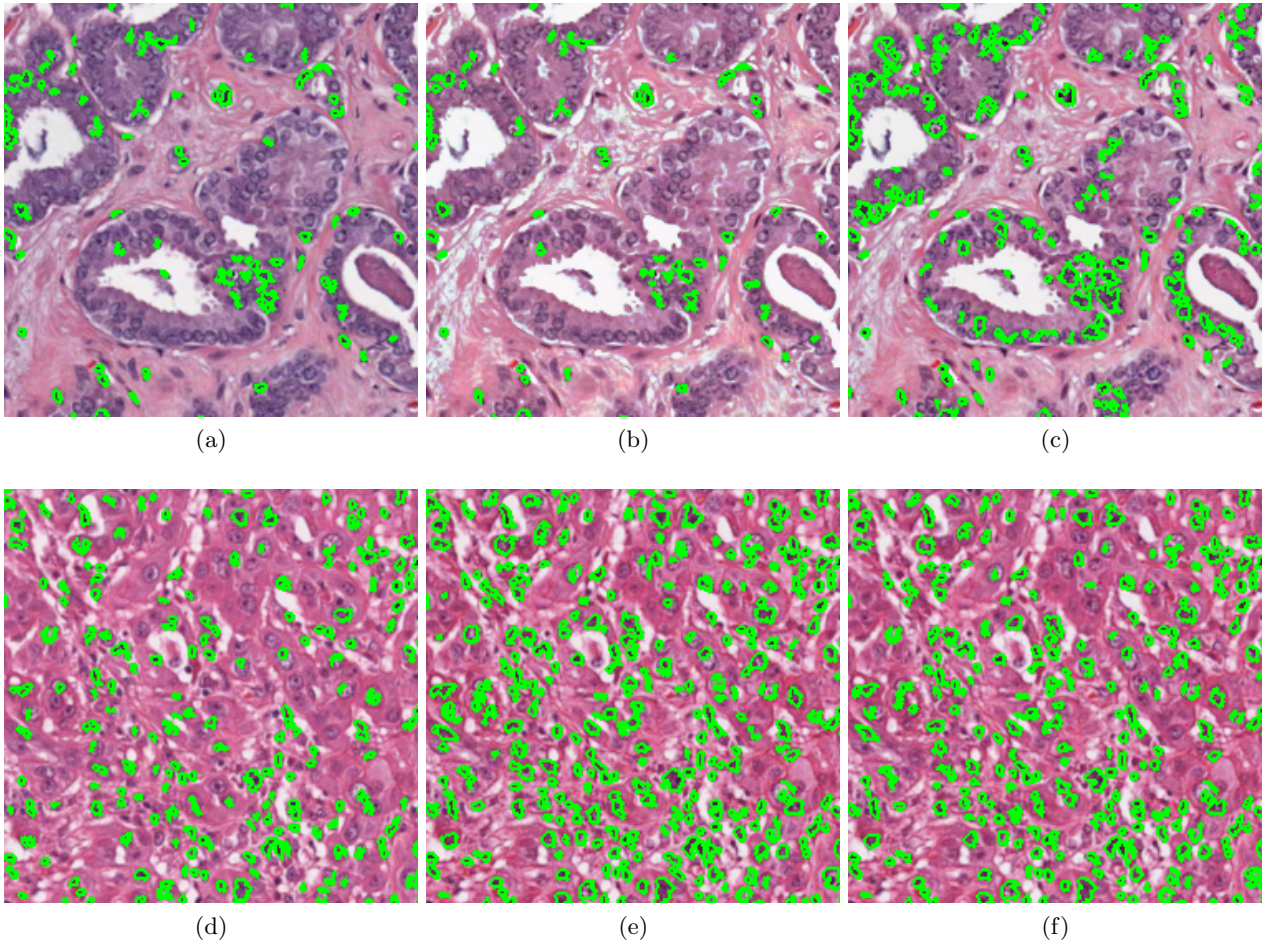


Figure 5. H & E stained test images for (a)-(c) prostate and (d)-(f) oropharyngeal cancers are shown. Segmented nuclei (green outline) are shown for images that are (a), (d) unstandardized (b), (e) globally standardized (GS), and (c), (f) standardized via EMS.

Cohort	Unstandardized vs. GS	Unstandardized vs. EMS	GS vs. EMS
Prostate	< 0.0001	< 0.0001	< 0.0001
Oropharyngeal	0.0645	< 0.0001	< 0.0001

Table 3. P-values from application of the Wilcoxon rank-sum test to pairwise histogram landmark distances between all images in the prostate and oropharyngeal cohorts. All reported p-values have undergone Bonferroni correction for multiple comparisons.<sup>13</sup>

significantly lower mean landmark distance of 2.25, compared to 54.8 and 27.1 for unstandardized and GS distributions, respectively (Figures 6(d)-(f)). The significance of this comparison is verified by application of the non-parametric Wilcoxon rank-sum test<sup>12</sup> in conjunction with a null hypothesis that pairwise histogram landmark distances between unstandardized images are not different from the distances calculated from standardized images (Table 3). Similarly, oropharyngeal images (Figures 7(d)-(f)) demonstrate significantly lower distances for EMS (4.2) compared to unstandardized (27.3) and GS (8.8) distributions, suggesting that EMS is able to more accurately account for the different proportions of tissue types (e.g. nuclei, epithelium, stroma, lumen) present in histopathology imagery.

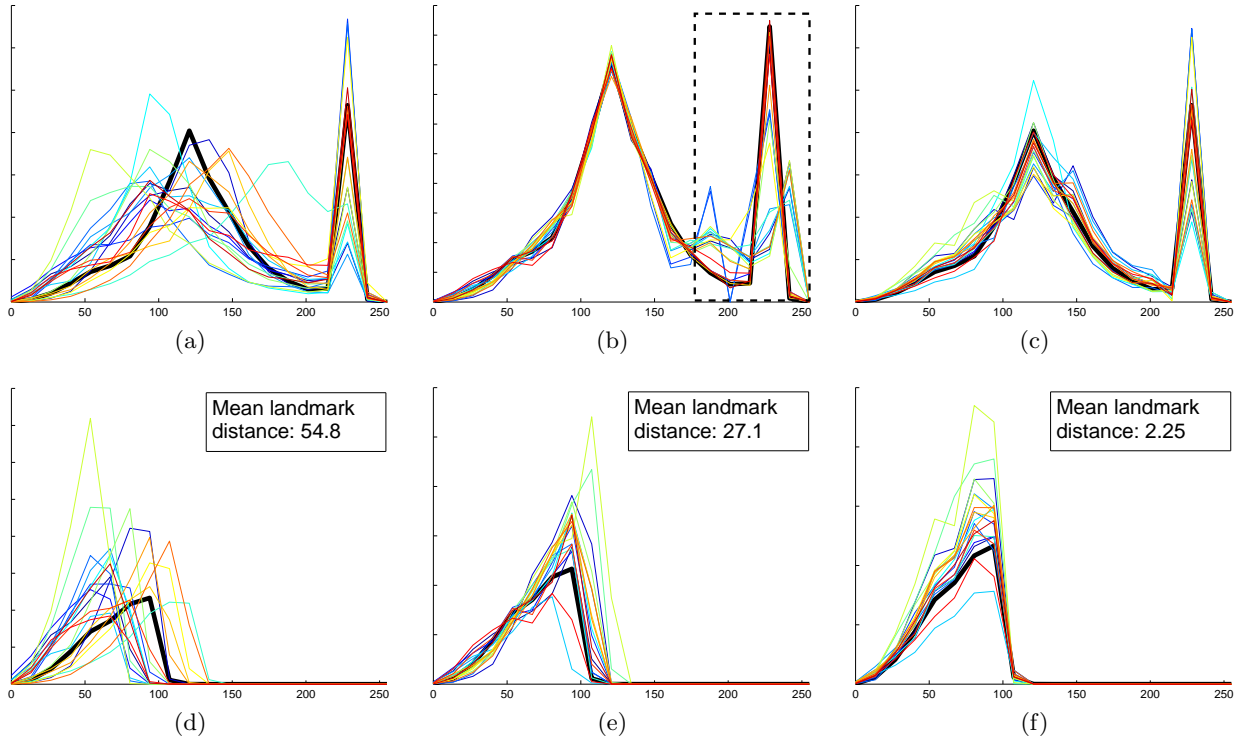


Figure 6. Distributions of the green color channels are shown for all images in the prostate cohort. Results are shown for (a), (d) unstandardized, (b), (e) GS, and (c), (f) EMS images. Alignment is shown for histograms representing (a)-(c) entire images and (d)-(f) an individual EM component (i.e. tissue class) along with the mean pairwise landmark distance over all images. In each figure, the histogram of the template image is represented by a thick black line and misalignment associated with GS is highlighted by a black box with dashed line.

## 6. CONCLUDING REMARKS

Digital pathology (DP) is a rapidly growing field involving the conversion of histopathology specimens into high-resolution digital images that are amenable to computer-aided analysis. The development of computerized algorithms, however, is greatly affected by color nonstandardness, where similar objects (e.g. nuclei) will have different color properties across images. Color nonstandardness occurs for several reasons related to both slide preparation (e.g. variations in slice thickness, staining protocol) and digitization (e.g. variations in illumination across slide scanning systems). Previous work in maintaining color constancy has traditionally focused on (1) calibration of slide scanning equipment or computer monitors and (2) natural images formed by reflective light, neither of which are appropriate for standardization of DP images. Related work in intensity standardization for MRI images yielded globally optimal solutions; yet, the heterogeneity of DP tissue classes (e.g. epithelium, stroma) requires a more flexible, localized approach to standardization. These issues were addressed for MRI images, in part, by a generalized scale approach that assumed correspondence between the largest connected component in the target and template images. However, this assumption is invalid for DP images and would lead to inconsistent partitioning of DP images. The EMS approach presented in this paper extends beyond the generalized scale approach by performing a fully localized, independent standardization of different tissue classes. We achieve this by (1) partitioning each image into broad tissue classes using the EM algorithm, (2) employing a piecewise linear transformation to align corresponding tissue classes to a template image, and (3) recombining the standardized pixels from each tissue class to create a standardized image. In this paper, EMS has been shown to yield improved color constancy over both non-standardized images and GS in terms of several evaluation approaches, including qualitative segmentation results, NMI of segmented regions, and mean histogram landmark alignment. Future work will entail evaluation on larger cohorts and other tissues as well as incorporating spatial information to improve separation of tissue classes.

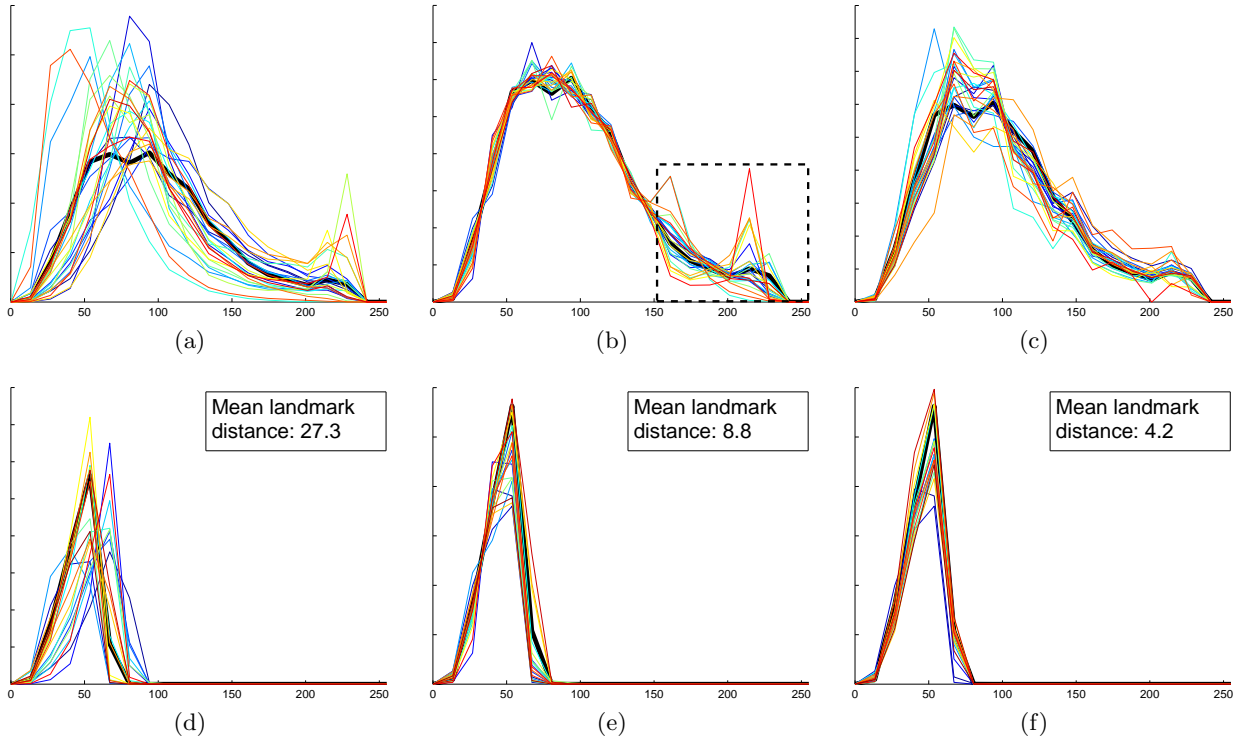


Figure 7. Distributions of the green color channels are shown for all images in the oropharyngeal cohort. Results are shown for (a), (d) unstandardized, (b), (e) GS, and (c), (f) EMS images. Alignment is shown for histograms representing (a)-(c) entire images and (d)-(f) an individual EM component (i.e. tissue class) along with the mean pairwise landmark distance over all images. In each figure, the histogram of the template image is represented by a thick black line and misalignment associated with GS is highlighted by a black box with dashed line.

## ACKNOWLEDGMENTS

This work was made possible by grants from the National Institute of Health (R01CA136535, R01CA140772, R43EB015199, R21CA167811), National Science Foundation (IIP-1248316), and the QED award from the University City Science Center and Rutgers University.

## REFERENCES

- [1] Gurcan, M. N., Boucheron, L., Can, A., Madabhushi, A., Rajpoot, N., and Yener, B., “Histopathological image analysis: A review,” *IEEE Reviews in Biomedical Engineering* **2**, 147–171 (2009).
- [2] Evans, A., Wilbur, D., Pfeifer, J., Kaplan, K., Collins, L., Pantanowitz, L., Valenstein, P., and Colgan, T., “Review of the current state of whole slide imaging in pathology,” *J. Pathol. Informat.* **2**(1), 36 (2011).
- [3] Palumbo, D., Yee, B., O’Dea, P., Leedy, S., Viswanath, S., and Madabhushi, A., “Interplay between bias field correction, intensity standardization, and noise filtering for t2-weighted mri,” *Conf Proc IEEE Eng Med Biol Soc* **2011**, 5080–5083 (2011).
- [4] Yagi, Y., “Color standardization and optimization in whole slide imaging.,” *Diag Pathol* **6 S.1**, S15 (2011).
- [5] Doyle, S., Feldman, M., Tomaszewski, J., and Madabhushi, A., “A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies,” *IEEE Trans Biomed Eng* **59**(5), 1205–1218 (2012).
- [6] Finlayson, G. D., Hordley, S. D., and Hubble, P. M., “Color by correlation: a simple, unifying framework for color constancy,” *IEEE Trans. Pattern Anal. and Machine Intel.* **23**(11), 1209–1221 (2001).
- [7] Reinhard, E., Ashikhmin, M., Gooch, B., and Shirley, P., “Color transfer between images,” *IEEE Computer Graphics and Applications* **21**(5), 34–41 (2001).

- [8] Nyúl, L. G., Udupa, J. K., and Zhang, X., “New variants of a method of mri scale standardization.,” *IEEE Trans Med Imaging* **19**, 143–150 (Feb 2000).
- [9] Madabhushi, A. and Udupa, J. K., “New methods of mr image intensity standardization via generalized scale.,” *Med Phys* **33**, 3426–3434 (Sep 2006).
- [10] Madabhushi, A. and Udupa, J. K., “Interplay between intensity standardization and inhomogeneity correction in mr image processing,” *IEEE Trans Med Imaging* **24**(5), 561–576 (2005).
- [11] Ramme, A. J., DeVries, N., Kallemyn, N. A., Magnotta, V. A., and Grosland, N. M., “Semi-automated phalanx bone segmentation using the expectation maximization algorithm.,” *J Digit Imaging* **22**, 483–491 (Oct 2009).
- [12] Wilcoxon, F., “Individual comparisons by ranking methods,” *Biometrics Bulletin* **1**(6), pp. 80–83 (1945).
- [13] Abdi, H., [*Encyclopedia of Measurement and Statistics*], Thousand Oaks (2007).