IMAGE SEGMENTATION WITH IMPLICIT COLOR STANDARDIZATION USING CASCADED EM: DETECTION OF MYELODYSPLASTIC SYNDROMES

James Monaco[⋆], *Philipp Raess*[†], *Ronak Chawla*[⋆]
Adam Bagg[†], *Mitchell Weiss*[‡], *John Choi*[§], *Anant Madabhushi*[⋆]

[⋆] Department of Biomedical Engineering, Rutgers University, Piscataway, NJ, USA
[†] Department of Pathology and Lab Medicine, University of Pennsylvania, Philadelphia, PA, USA
[‡] Division of Hematology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
[§] Department of Pathology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

ABSTRACT

Color nonstandardness — the propensity for similar objects to exhibit different color properties across images — poses a significant problem in the computerized analysis of histopathology. Though many papers propose means for improving color constancy, the vast majority assume image formation via reflective light instead of light transmission as in microscopy, and thus are inappropriate for histological analysis. In this work, we present a novel Bayesian color segmentation algorithm for histological images that is highly robust to color nonstandardness; this algorithm employs a unique instantiation of the expectation maximization (EM) algorithm to dynamically estimate — for each individual image — the probability density functions (mixtures of gamma and von Mises distributions) that describe the colors of salient objects. To validate our segmentation scheme, we employ it as part of a computerized system to detect myelodysplastic syndromes (MDS) on bone marrow specimens. Qualitative anecdotal evidence suggests that biopsies of MDS exhibit abnormalities in the arrangement of erythroid precursors (immature red blood cells). Herein, we confirm and quantify this phenomenon, using it to discriminate MDS from normal tissue: over a dataset of 53 representative regions selected from 18 patients, our classification system correctly discriminates MDS from normal tissue with an accuracy of 85% and an area under the receiver operator characteristic curve of 0.8803.

1. INTRODUCTION

Color nonstandardness, the propensity for similar objects (e.g. cells) to exhibit different color properties across images, poses a significant challenge in the analysis of histopathology images. This nonstandardness typically results from variations in tissue fixation, staining, and digitization. Though methods have been proposed for improving color constancy in images formed via reflective light (see [1] for a review) and for mitigating the analagous intensity drift in grayscale images (e.g. MRI [2]), these methods have not been extended to color images formed via light transmission as in microscopy, and thus are inappropriate for histological analysis.

In this work, we present a novel Bayesian (pixelwise) color segmentation algorithm for histological images that is highly robust to color nonstandardness. With respect to Bayesian classification, color nonstandardness manifests as inter-image changes in the probability density functions (PDFs) that describe the colors of salient objects. To account for such changes, we estimate the requisite PDFs in each image individually. This estimation is facilitated by operating in HSV color space. Unlike RGB and CIELab, HSV yields color channels that are (nearly) statistically independent, justifying the use of univariate PDFs to model each channel. This is significant: estimating PDFs becomes increasingly difficult as the number of dependent dimensions increases. Nonetheless, even in a single dimension, the HSV color channels can create complex distributions, which are difficult to model parametrically with a standard PDF (e.g. Gaussian, Weibull). Consequently, we employ mixtures of standard PDFs, significantly improving their ability to fit complex data. To estimate the parameters of these mixtures, we present cascaded expectation maximization (EM), a novel instantiation of EM [3] that facilitates the estimation of PDFs modeled by a hierarchy of mixtures.

To demonstrate the efficacy of our segmentation algorithm, we employ it as part of a computerized system to detect myelodysplastic syndromes (MDS) on digitized bone marrow specimens. MDS are blood-based neoplasms typified by the ineffective formation of blood cellular components. They can present symptomatically as anemia, neutropenia, or thrombocytopenia [4]. Diagnosis of MDS is multifactorial, requiring the analysis of bone marrow biopsy and aspirate, blood samples, and cytogenetic studies. Bone marrow biopsies from patients with MDS have been noted to exhibit abnormalities in the arrangement of erythroid precur-

This work was made possible through funding from the National Cancer Institute (R01CA136535-01, R01CA140772-01A1, R03CA128081-01).

sors (i.e. immature red blood cells). In healthy tissue, these cells usually form cohesive aggregates — called "erythroid islands"; MDS disrupts such islands (see Figures 1(a) and (c)). However, the relationship between erythroid precursor (EP) aggregation and MDS is purely anecdotal: no study has confirmed the value of EP islands in diagnosing MDS. In fact, no one has proposed a measure to quantitate this potentially valuable imaging marker.

To quantitatively explore the connection between EP islands and MDS, this work introduces a computerized system that quantitates EP clusters on bone marrow specimens stained for an alpha hemoglobin stabilizing protein (ASHP), which specifically colors EPs brown. More precisely, our system 1) detects and segments clusters of EP cells, 2) extracts features characterizing their appearance, and 3) uses these features to classify each sample as either MDS or normal.

2. CASCADED EXPECTATION MAXIMIZATION

2.1. Notation and Definitions

Let $R=\{1,2,\ldots,N\}$ reference N sites to be classified. Each site $r \in R$ has two associated random variables: $X_r \in \Lambda \equiv \{1,2,\ldots,M\}$ indicating its class and $Y_r \in \mathbb{R}^D$ representing its D-dimensional feature vector. Let $\mathbf{X}=[X_1,X_2,\ldots,X_N]$ and $\mathbf{Y}=[Y_1,Y_2,\ldots,Y_N]$ refer to all random variables X_r and Y_r in aggregate. Cascaded EM (CEM) assumes that all X_r and Y_r are independent and identically distributed (i.i.d).

2.2. Derivation of Cascaded Expectation Maximization

Bayesian methods for estimating $\mathbf{X}$ typically require the conditional PDFs $P(\mathbf{Y}|\mathbf{X},\boldsymbol{\theta})$. These PDFs are often modeled parametrically, using maximum likelihood estimation (MLE) to determine the parameters $\boldsymbol{\theta}$ which best explain $\mathbf{Y}$. A generalized form of the log likelihood (which is maximized during MLE) is $L(\boldsymbol{\theta}|\mathbf{Y},\mathbf{w}) = \sum_{r\in R} w_r \ln P(Y_r|\boldsymbol{\theta})$, where $\mathbf{w} = [w_1,...,w_N]$ and $w_r \in \mathbb{R}^+$ are positive real numbers weighting each sample. For typical MLE, we have $\mathbf{w}\equiv 1$.

The CEM algorithm provides an effective means of determining the maximum likelihood (ML) estimate. Using an iterative approach, at each step t the CEM algorithm chooses a value $\boldsymbol{\theta}^{(t+1)}$ such that $Q(\boldsymbol{\theta}^{(t+1)}|\boldsymbol{\theta}^{(t)}) > Q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t)})$, where

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= \sum_{\lambda\in\Lambda}\sum_{r\in R} w_r P(X_r=\lambda|Y_r,\boldsymbol{\theta}^{(t)})\times \\ &\qquad [\ln P(X_r=\lambda|\pi_\lambda) + \ln P(Y_r|X_r=\lambda,\boldsymbol{\theta}_\lambda)] \\ &= \sum_{\lambda\in\Lambda}\sum_{r\in R} w_r P(X_r=\lambda|Y_r,\boldsymbol{\theta}^{(t)}) \ln \pi_\lambda + \\ &\qquad \sum_{\lambda\in\Lambda} L(\boldsymbol{\theta}_\lambda|\mathbf{Y},X=\lambda,\mathbf{w}_\lambda), \end{aligned} \tag{1}$$

$\boldsymbol{\theta} = [\boldsymbol{\theta}_1 ... \boldsymbol{\theta}_L, \pi_1,\ldots,\pi_L]$, $\boldsymbol{\theta}_\lambda$ are those parameters needed to model the conditional distribution of Y_r given $X_r=\lambda$, $w_{\lambda,r} = w_r P(X_r=\lambda|Y_r,\boldsymbol{\theta}_\lambda^{(t)})$, and $\pi_\lambda = P(X_r=\lambda|\pi_\lambda)$. Note that this result requires that X_r and Y_r be i.i.d.

Thus, the maximization of Q can be subdivided into the $M+1$ independent maximization problems. The first sum in (1) can be solved to obtain the prior probabilities. The second sum contains M independent weighted ML equations, which — depending upon their parametric forms — can solved analytically (e.g. Gaussian), using standard numerical techniques (e.g. gamma distribution), or with CEM (e.g. a mixture distribution). If CEM is used, the ML equation again separates as in (1), creating a cascading tree-like structure.

3. DISCRIMINATING MDS FROM NORMAL TISSUE ON BONE MARROW BIOPSIES

We begin with an overview of the algorithm. Step 1) The algorithm identifies the adipose tissue (the white holes) by thresholding the value and saturation at each pixel; this approach is standard and further details are omitted. Step 2) Using the HSV color values, a Bayesian classifier labels each non-adipose pixel as either stained or non-stained. The algorithm then traces the boundaries of all groups of contiguous pixels classified as stained (Figures 1(b), (d)). Step 3) Features are extracted from the boundaries. Step 4) Each image is classified as MDS or normal based on its features values.

3.1. Segmentation of Erythroid Precursor Islands

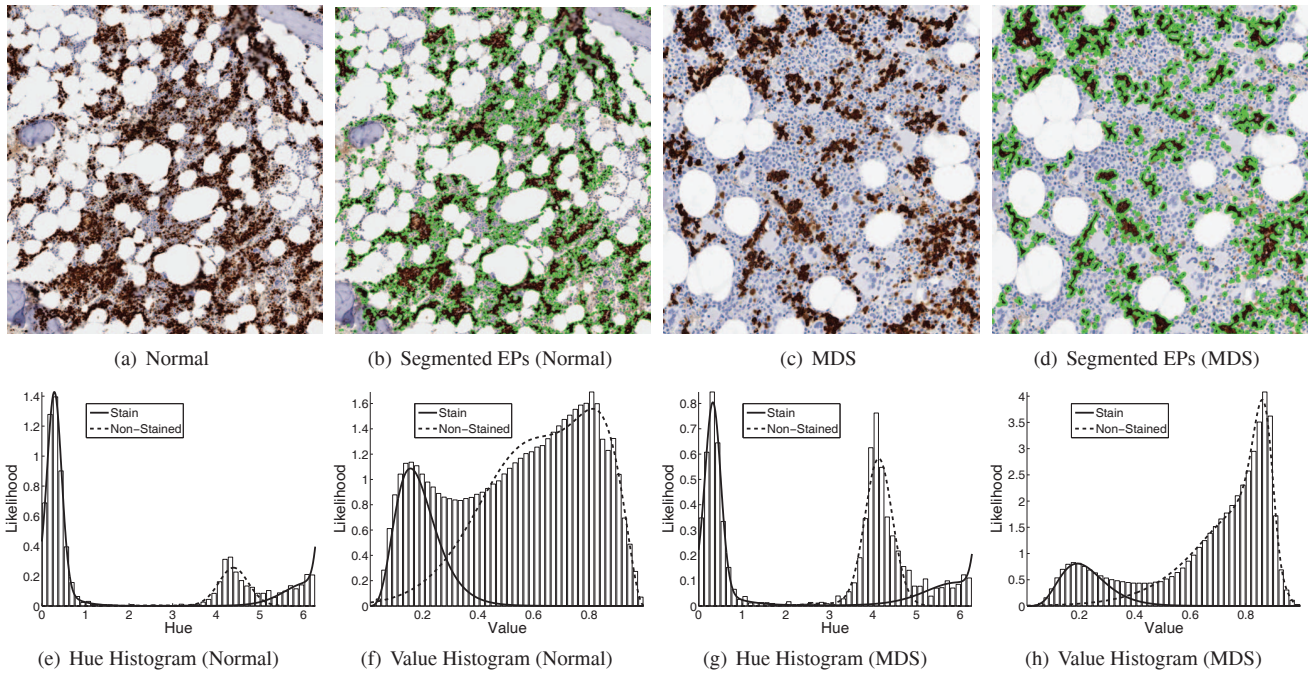
The goal of segmentation is to delineate the boundaries of EP islands. It is hypothesized that the size and shape of these islands may discriminate MDS from normal tissue. Segmentation is achieved by classifying each pixel as either stained or non-stained, and then tracing the boundaries of contiguous stained pixels (boundary tracing details are omitted).

3.1.1. Pixelwise Classification via MAP Estimation

We first state the Bayesian segmentation problem using the nomenclature presented in Section 2. Let the set $R = \{1,2,\ldots,N\}$ reference the N non-adipose pixels to be classified. Each pixel $r \in R$ has two associated random variables: $X_r \in \Lambda \equiv \{1,2\}$ indicating its class as either stained (1) or non-stained (2) and $Y_r \equiv [H_r\, S_r\, V_r]^\mathsf{T} \in [0,1]^3$ representing its color in HSV space — with each dimension normalized to the interval $[0,1]$. We assume all X_r and Y_r are i.i.d.; we also assume each color dimension is independent of the others.

Classification is accomplished via maximum *a posteriori* (MAP) estimation, which determines the $\mathbf{X}$ that maximizes

$$\begin{aligned} P(\mathbf{X}|\mathbf{Y}) &= \frac{P(\mathbf{Y}|\mathbf{X})P(\mathbf{X})}{P(\mathbf{Y})} \propto P(\mathbf{Y}|\mathbf{X})P(\mathbf{X}) \\ &= \prod_{r\in R} P(H_r|X_r)P(S_r|X_r)P(V_r|X_r)P(X_r). \end{aligned}$$



(a) Normal (b) Segmented EPs (Normal) (c) MDS (d) Segmented EPs (MDS)

(e) Hue Histogram (Normal) (f) Value Histogram (Normal) (g) Hue Histogram (MDS) (h) Value Histogram (MDS)

Fig. 1. Normal (a) and MDS (c) digitized marrow specimens. The brown cells are EPs stained with AHSP. As compared to MDS, the EPs in the normal specimen demonstrate an increased (though subtle) tendency to form dense "erythroid islands." (b), (d) Segmentation boundaries of stained EPs clusters. (e), (g) Hue histograms overlaid with estimated PDFs. (f), (h) Value histograms overlaid with estimated PDFs.

Thus, MAP estimation requires three conditional probability distributions and one prior probability distribution. We assume a non-informative prior: $P(X_r) \equiv 1/2$. Additionally, the saturation channel, which was found to be non-informative, is ignored. To model the conditional distributions of the hue and value, we use parametric mixture models defined by the modeling parameters $\boldsymbol{\theta}^h$ and $\boldsymbol{\theta}^v$. These parameters are estimated for each image using CEM. Since each conditional distribution is independent of the other, $\boldsymbol{\theta}^h$ and $\boldsymbol{\theta}^v$ can be determined independently.

Note that CEM converges to a local maximum of the log likelihood, and thus the initial conditions are important. In this application, we know *a priori* the approximate shape and location of each conditional distribution (e.g. the stained pixels are brown, establishing a range of reasonable hues). Empirically choosing the initial conditions was straightforward.

3.1.2. Parametric Models of Conditionals Distributions

Hue: Since hue is an angle (normalized to fit in the interval $[0, 1]$), we require distributions appropriate for modeling "wrapped" data. The von Mises is the natural choice:

$$f_h(h|\mu, \kappa) = \exp\{\kappa \cos(2\pi[h - \mu])\}/I_0(\kappa),$$

where μ is the mean, κ is the concentration (i.e. shape), and I_0 is the modified Bessel function. Unfortunately, a single von Mises is not sufficiently complex to account for the complex shapes of the conditional distributions $P(H_r|X_r = 1)$ and $P(H_r|X_r = 2)$. That is, a single von Mises will not accurately fit the observed data. An easy was to expand its modeling capability is to form a mixture of von Mises distributions. Thus, we model the two conditional PDFs as follows:

$$P(H_r|X_r = \lambda, \boldsymbol{\theta}^h_\lambda) = \alpha^h f_h(h_r|\boldsymbol{\theta}^h_{\lambda,1}) + (1-\alpha^h) f_h(h_r|\boldsymbol{\theta}^h_{\lambda,2}),$$

where $\alpha^h \in [0, 1]$ is the mixing term, $\boldsymbol{\theta}^h_\lambda = [\alpha^h, \boldsymbol{\theta}^h_{\lambda,1}, \boldsymbol{\theta}^h_{\lambda,2}]$, and $\boldsymbol{\theta}^h_{\lambda,j} = [\mu_{\lambda,j}, \kappa_{\lambda,j}]$ for $j \in \{1, 2\}$. Figures 1(e) and (g) depict histograms of the hue channel overlaid with their estimated conditional distributions.

Value: The value also ranges from zero to one, but does not wrap like the hue. We begin with the gamma distribution:

$$f_v(v|k, \phi) = v^{k-1} \exp\{v/\phi\}/[\phi^k \Gamma(k)],$$

where $k > 0$ is the shape parameter, $\phi > 0$ is the scale parameter, and Γ is the gamma function. As with the hue, the gamma distribution alone is insufficient to model the conditional PDFs. Thus, we again form mixtures:

$$P(V_r|X_r = 1, \boldsymbol{\theta}^v_1) = \alpha^v f_v(v_r|\boldsymbol{\theta}^v_{1,1}) + (1-\alpha^v) f_v(v_r|\boldsymbol{\theta}^v_{1,2})$$

$$P(V_r|X_r = 2, \boldsymbol{\theta}^v_2) = \alpha^v f_v(1-v_r|\boldsymbol{\theta}^v_{2,1}) + (1-\alpha^v) f_v(1-v_r|\boldsymbol{\theta}^v_{2,2}),$$

where $\alpha^v \in [0, 1]$, $\boldsymbol{\theta}^v_\lambda = [\alpha^v, \boldsymbol{\theta}^v_{\lambda,1}, \boldsymbol{\theta}^v_{\lambda,2}]$, and $\boldsymbol{\theta}^v_{\lambda,j} = [k_{\lambda,j}, \phi_{\lambda,j}]$ for $j \in \{1, 2\}$. Note that since stained pixels yield low values and non-stained pixels exhibit high values, $P(V_r|X_r = 1, \boldsymbol{\theta}^v_1)$ is a mixture of gamma PDFs and $P(V_r|X_r = 2, \boldsymbol{\theta}^v_2)$ is a mixture of gamma PDFs that have been reflected and shifted. Figures 1(f) and (h) depict histograms of the value channel overlaid with their estimated conditional PDFs.

Feature	Accuracy	AUC
Mean Area	70%	0.6453
Standard Deviation of Area	62%	0.5613
Area (25^{th} Percentile)	68%	0.6503
Area (50^{th} Percentile)	70%	0.6866
Area (75^{th} Percentile)	79%	0.8433
Area (80^{th} Percentile)	81%	0.8113
Area (85^{th} Percentile)	81%	0.8661
Area (90^{th} Percentile)	85%	0.8803
Mean Compactness	68%	0.6638
Relative Area of Stain	62%	0.7170

Table 1. Features describing EP islands and their performance discriminating MDS from normal tissue.

3.2. Feature Extraction and Image Classification

From each segmented boundary, we extract the area and compactness. We then aggregate these measures, yielding features that describe the entire image instead of each individual segmentation: mean area, order statistics (i.e. percentiles) of the area, mean compactness, standard deviation of areas, and "relative area of stain" which is ratio of the total area of the stained regions to the total non-stained area. To classify each image as MDS or normal, we simply threshold the features of interest. (Note that for some features, measurements below the prescribed threshold indicate MDS, while for others, measurements exceeding the threshold indicate MDS.)

4. RESULTS AND DISCUSSION

4.1. Dataset

Our collaborating pathologists selected 18 bone marrow biopsies: nine normal and nine MDS. Diagnoses were made using current clinical protocol, which considers — in addition to the bone marrow biopsies — peripheral blood samples, bone marrow aspirate, and cytogenetic information. One section of each specimen was immunohistochemically stained for AHSP, counterstained with hematoxylin, and then digitized at 20x resolution (0.5 μm per pixel) using an Aperio scanner. Finally, a pathologist delineated two to three representative regions per image, resulting in 53 total sub-images.

4.2. Performance Evaluation and Discussion

To test our system, we segmented the stained regions and extracted the described features for all 53 sub-images. Since our system is completely unsupervised, there was no need to establish training and testing sets. To assess the value of each feature individually, we classified each sub-image as either MDS or normal by applying a threshold to the specific feature measurement. Table 1 details the classification performance for each feature — measured using accuracy and the area under the receiver operator characteristic curve (AUC).

To qualitatively illustrate the importance of color standardization we provide Figure 2, which depicts the high variability captured by the PDFs describing the value channel of the stained pixels. Note that since color standardization is performed implicitly, our algorithm does not need to align these PDFs with an established reference distribution (as is typical of explicit methods [2]). Finally, note that we did not directly evaluate segmentation performance due to the difficulty of obtaining expert segmentations; however, segmentation efficacy is measured indirectly by overall algorithm performance.

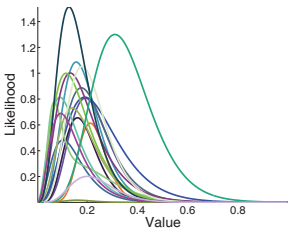


Fig. 2. Estimated PDFs of value channel for the stained EPs. PDFs were estimated using one sub-image from each of the 18 patients. Note the high inter-image variability.

The results in Table 1 demonstrate (over our dataset) that the area of erythroid islands — when assessed using the 90^{th} percentile — can discriminate normal from MDS tissue with an accuracy and AUC of 85% and 0.8803, respectively. This suggests that the formation of EP islands — noted anecdotally and qualitatively by pathologists — is in fact a valid phenomenon that merits further investigation. We hypothesize that the 90^{th} percentile is more discriminative than other measurements, such as the mean, because of the following: though normal tissue produces many small EP islands, an appreciable number of the islands are very large, in concordance with anecdotal observations. MDS, on the other hand, rarely produces large islands. Thus, the difference in island size becomes more pronounced as the area percentile increases.

5. CONCLUDING REMARKS

In this paper, we introduced a Bayesian method for segmenting histological images that is robust to color nonstandardness. For each image, our method dynamically estimates the PDFs describing the colors of salient objects in HSV space. Since these univariate PDFs were too complex to model using standard densities, we employed mixtures of standard densities. We also introduced cascaded EM, a novel instantiation of EM that facilitates the fitting of such mixture models. Finally, to evaluate the efficacy of our segmentation method, we integrated it into a system to discriminate MDS from normal tissue on digitized bone marrow biopsies.

6. REFERENCES

[1] G Finlayson *et al.*, "Color by correlation: A simple, unifying framework for color constancy," *PAMI*, vol. 23, pp. 1209–1221, 2001.

[2] A Madabhushi *et al.*, "Comparing mr image intensity standardization against tissue characterizability of magnetization transfer ratio imaging," *J Magn Reson Imag*, vol. 24, pp. 667–675, 2006.

[3] A Dempster *et al.*, "Maximum likelihood from incomplete data via the em algorithm," *J Roy Stat Soc*, vol. 39, pp. 1–38, 1977.

[4] K Foucar, *Bone Marrow Pathology*, ASCP Press, 2001.