

Image Segmentation with Implicit Color Standardization Using Spatially Constrained Expectation Maximization: Detection of Nuclei

James Monaco¹, J. Hipp², D. Lucas², S. Smith²,
U. Balis², and Anant Madabhushi^{1,*}

¹ Department of Biomedical Engineering, Rutgers University, USA

² Department of Pathology, University of Michigan, 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. Previously, we presented a novel Bayesian color segmentation algorithm for histological images that is highly robust to color nonstandardness; this algorithm employed the expectation maximization (EM) algorithm to dynamically estimate — for each individual image — the probability density functions that describe the colors of salient objects. However, our approach, like most EM-based algorithms, ignored important spatial constraints, such as those modeled by Markov random field (MRFs). Addressing this deficiency, we now present spatially-constrained EM (SCEM), a novel approach for incorporating Markov priors into the EM framework. With respect to our segmentation system, we replace EM with SCEM and then assess its improved ability to segment nuclei in H&E stained histopathology. Segmentation performance is evaluated over seven (nearly) identical sections of gastrointestinal tissue stained using different protocols (simulating severe color nonstandardness). Over this dataset, our system identifies nuclear regions with an area under the receiver operator characteristic curve (AUC) of 0.838. If we disregard spatial constraints, the AUC drops to 0.748.

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

* This work was made possible through funding from the NCI (R01CA136535-01, R01CA140772-01A1, R03CA128081-01) and the Burroughs Wellcome Fund.

(see [1] for a review) and for mitigating the analogous intensity drift in grayscale images (e.g. MRI [2]), these methods are not extensible to color images formed via light transmission as in microscopy, and thus are inappropriate for histological analysis.

We previously presented a novel Bayesian color segmentation algorithm for histological images that is highly robust to color nonstandardness [3]. With respect to Bayesian classification, color nonstandardness manifests as inter-image changes in the probability densities that describe the colors of salient objects. To account for such changes, we estimated the requisite distributions in each image individually using expectation maximization (EM). However, our methodology did not account for spatial coherency among pixels.

Within a Bayesian framework, spatial dependencies are modeled as Markov random fields. Unfortunately, using EM to estimate Markov priors is notoriously complex: the EM equations become intractable. In an effort to maintain tractability, authors have invoked simplifying assumptions; unfortunately, these assumptions tended to reduce the generality of the solutions. For example, Comer and Delp [4] demanded that the Markov prior remain constant during the EM iteration; Nikou *et al.* [5] imposed Gaussianity constraints; Marroquin [6] tied the implementation to a predetermined cost function, thus linking estimation (i.e. model fitting) and classification.

In this work, we introduce spatially-constrained EM (SCEM), a novel approach for incorporating Markov priors into the EM framework. This approach 1) employs the pseudo-likelihood [7] to simplify the EM equations without incurring any loss of generality and 2) leverages our recently presented Markov chain Monte Carlo (MCMC) method [8] to estimate the resulting complex marginal distributions. As we will discuss later, previous MCMC methods [9] are ill-equipped to perform this estimation.

We validate SCEM by integrating it into our computerized system to segment nuclei in H&E stained histopathology. To test our system, we cut seven consecutive slices from a block of paraffin embedded gastrointestinal (GI) tissue, H&E stain the slices using different protocols (to simulate the gamut of color nonstandardness), digitized the sections, and then use our algorithm to segment the nuclear regions from the background. We then compare the resulting performance to that of our previous approach, which ignored spatial constraints.

2 Spatially Constrained Expectation Maximization

2.1 Definitions and Terminology

Let $R = \{1, 2, \dots, |R|\}$ reference $|R|$ sites to be classified. Each site $r \in R$ has two associated random variables: $X_r \in \mathcal{A} \equiv \{1, 2, \dots, |A|\}$ indicating its class and $Y_r \in \mathbb{R}^D$ representing its D -dimensional feature vector. Let $\mathbf{X} = \{X_r : r \in R\}$ and $\mathbf{Y} = \{Y_r : r \in R\}$ refer to all X_r and Y_r in aggregate. The state spaces of \mathbf{X} and \mathbf{Y} are the Cartesian products $\Omega = \mathcal{A}^{|R|}$ and $\mathbb{R}^{D \times |R|}$. Instances of random variables are denoted by their associated lowercase letters, e.g. $P(\mathbf{X} = \mathbf{x})$.

The random field \mathbf{X} is a Markov random field if its local conditional probability density functions satisfy the Markov property: $P(X_r = x_r | \mathbf{X}_{-r} = \mathbf{x}_{-r}) = P(X_r = x_r | \mathbf{X}_{\eta_r} = \mathbf{x}_{\eta_r})$, where $\mathbf{x}_{-r} = \{x_s : s \in R, s \neq r\}$ and $\eta_r \subseteq R$ is the set of sites that neighbor r . Note that where it does not cause ambiguity, we will henceforth simplify probabilistic notations by omitting the random variables, e.g. $P(\mathbf{x}) \equiv P(\mathbf{X} = \mathbf{x})$. The joint probability distribution describing \mathbf{X} and \mathbf{Y} is as follows: $P(\mathbf{y}, \mathbf{x} | \boldsymbol{\theta}) = P(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}) P(\mathbf{x} | \boldsymbol{\theta})$, where $P(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta})$ is the conditional distribution, $P(\mathbf{x} | \boldsymbol{\theta})$ is the Markov prior, and $\boldsymbol{\theta}$ is the vector of free parameters.

2.2 Derivation of Spatially Constrained Expectation Maximization

Typically, we infer $\boldsymbol{\theta}$ from the observations \mathbf{y} . The most common method of inference is maximum likelihood estimation (MLE). In an unsupervised context, EM provides an effective method for performing MLE. In brief, the EM algorithm employs the following iterative approach: at each step t , EM selects $\boldsymbol{\theta}^{t+1}$ such that $Q(\boldsymbol{\theta}^{t+1} | \boldsymbol{\theta}^t) > Q(\boldsymbol{\theta}^t | \boldsymbol{\theta}^t)$, where $Q(\boldsymbol{\theta} | \boldsymbol{\theta}^t) = \sum_{\mathbf{x}} P(\mathbf{x} | \mathbf{y}, \boldsymbol{\theta}^t) \ln P(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta})$. Typically, $\boldsymbol{\theta}^{t+1}$ is set to the $\boldsymbol{\theta}$ that maximizes $Q(\boldsymbol{\theta} | \boldsymbol{\theta}^t)$ [10].

Unfortunately, the above form of Q is not amenable to implementation. To construct a more tractable form, we present the following derivation:

$$\begin{aligned} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^t) &= \sum_{\mathbf{x}} P(\mathbf{x} | \mathbf{y}, \boldsymbol{\theta}^t) [\ln P(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}) + \ln P(\mathbf{x} | \boldsymbol{\theta})] \\ &\approx \sum_{\mathbf{x}} P(\mathbf{x} | \mathbf{y}, \boldsymbol{\theta}^t) \sum_{r \in R} [\ln P(y_r | x_r, \boldsymbol{\theta}) + \ln P(x_r | \mathbf{x}_{\eta_r}, \boldsymbol{\theta})] \quad (1) \\ &= \sum_{r \in R} \sum_{x_r} P(x_r | \mathbf{y}, \boldsymbol{\theta}^t) \ln P(y_r | x_r, \boldsymbol{\theta}) + \\ &\quad \sum_{r \in R} \sum_{\mathbf{x}_{r \cup \eta_r}} P(\mathbf{x}_{r \cup \eta_r} | \mathbf{y}, \boldsymbol{\theta}^t) \ln P(x_r | \mathbf{x}_{\eta_r}, \boldsymbol{\theta}) \quad (2) \end{aligned}$$

Equation (1) results from 1) making the common assumption that each observation Y_r is conditionally independent given its associated state X_r , i.e. $P(\mathbf{y} | \mathbf{x}) = \prod_{r \in R} P(y_r | x_r)$ and 2) replacing the Markov prior with its pseudo-likelihood representation [7]: $P(\mathbf{x} | \boldsymbol{\theta}) \approx \prod_{r \in R} P(x_r | \mathbf{x}_{\eta_r}, \boldsymbol{\theta})$. Equation (2) follows from changing the order of summation and then summing out the superfluous variables. Note that the first term in (2) is the formulation proposed by Comer and Delp [4]. It is the second term that allows us to estimate the MRF parameters during the SCEM iteration.

In summary, the expectation step in SCEM requires estimating the marginal probabilities $P(\mathbf{x} | \mathbf{y}, \boldsymbol{\theta}^t)$ and $P(\mathbf{x}_{r \cup \eta_r} | \mathbf{y}, \boldsymbol{\theta}^t)$ using the current estimate $\boldsymbol{\theta}^t$. The maximization step involves determining the $\boldsymbol{\theta}^{t+1}$ that maximizes (2). The difficulty lies in estimating the marginals needed for the expectation step.

2.3 Estimating the Marginals Densities

The (large) range Ω of \mathbf{X} precludes determining $P(x_r | \mathbf{y}, \boldsymbol{\theta})$ and $P(\mathbf{x}_{r \cup \eta_r} | \mathbf{y}, \boldsymbol{\theta})$ via the direct marginalization of $P(\mathbf{x} | \mathbf{y}, \boldsymbol{\theta})$. However, it is possible to use the Gibbs

sampler [11] (or the Metropolis-Hastings algorithm) to generate a Markov chain $(\mathbf{X}^0, \mathbf{X}^1, \mathbf{X}^2, \dots | \mathbf{y}, \boldsymbol{\theta})$ whose elements are random samples of $P(\mathbf{x} | \mathbf{y}, \boldsymbol{\theta})$ — and consequently, samples of $P(x_r | \mathbf{y}, \boldsymbol{\theta})$ and $P(\mathbf{x}_{r \cup \eta_r} | \mathbf{y}, \boldsymbol{\theta})$. Using these samples, a Monte Carlo procedure can then estimate the requisite marginals. Unfortunately, previous MCMC approaches, which employed a histogramming strategy for density estimation [9], are ill-equipped to estimate distributions for which the sample space is large, such as $P(\mathbf{x}_{r \cup \eta_r} | \mathbf{y}, \boldsymbol{\theta})$.

We recently presented a more effective Rao-Blackwellized MCMC density estimator [8]. Consider the general marginal $P(\mathbf{x}_{\tilde{R}})$, where $\mathbf{x}_{\tilde{R}} = \{x_r : r \in \tilde{R} \subseteq R\}$. Our estimator is as follows:

$$\hat{P}(\mathbf{X}_{\tilde{R}} = \boldsymbol{\lambda} | \mathbf{y}) = \frac{1}{c \cdot m} \sum_{j=1}^c \sum_{k=b+1}^{b+m} P(\mathbf{X}_{\tilde{R}} = \boldsymbol{\lambda} | \mathbf{x}_{\eta_{\tilde{R}}}^{j,k}, \mathbf{y}), \quad (3)$$

where $\boldsymbol{\lambda} \in A^{|\tilde{R}|}$, c is the number of Markov chains, b is the number of iterations needed for the Markov chain to reach equilibrium, m is the number of iterations past equilibrium required to accurately estimate $P(\mathbf{x}_{\tilde{R}})$, and $\mathbf{x}_{\eta_{\tilde{R}}}^{j,k}$ is the states of all sites that neighbor \tilde{R} in Markov chain j at iteration k .

By averaging over the functional forms $P(\mathbf{x}_{\tilde{R}} | \mathbf{x}_{\eta_{\tilde{R}}}, \mathbf{y})$ — instead of the samples themselves as with typical histogramming [9] — each “sample” in (3) updates $\hat{P}(\mathbf{x}_{\tilde{R}} | \mathbf{y})$ for all $\mathbf{x}_{\tilde{R}} \in A^{|\tilde{R}|}$, greatly decreasing the number of samples needed for an accurate estimate. Furthermore, using multiple Markov chains (instead of the typical one [9]) improves robustness to the presence of multiple modes in $P(\mathbf{x}_{\tilde{R}})$, each of which can trap the Markov chain.

3 Segmenting Nuclei on H&E Stained GI Tissue

We begin with an overview of the algorithm. Step 1) SCEM adapts the parameterized distributions to model the color — hue in HSV color-space — and spatial properties of nuclear and stromal pixels. Step 2) Using the MCMC estimator in (3), we calculate the probability that a given pixel is nuclei or stroma. Step 3) We threshold the pixel-wise probabilities to create hard classifications.

3.1 Color and Spatial Distributions of Nuclear and Stromal Pixel

First, we restate the segmentation problem using the nomenclature presented in Section 2.1. Let the set $R = \{1, 2, \dots, |R|\}$ reference the $|R|$ pixels in an image. Each pixel $r \in R$ has two associated random variables: $X_r \in A \equiv \{1, 2\}$ indicating its class as either nuclei $\{1\}$ or stroma $\{2\}$ and $Y_r \in [0, 2\pi]$ representing its hue (color) in HSV space. We assume all X_r are identically distributed; we assume all Y_r are conditionally independent given X_r and identically distributed.

Conditional Distribution. The conditional distribution models the color properties of nuclear and stromal pixels. Since these biological structures stain differently (blue and red, respectively), their hue in HSV color-space provides a

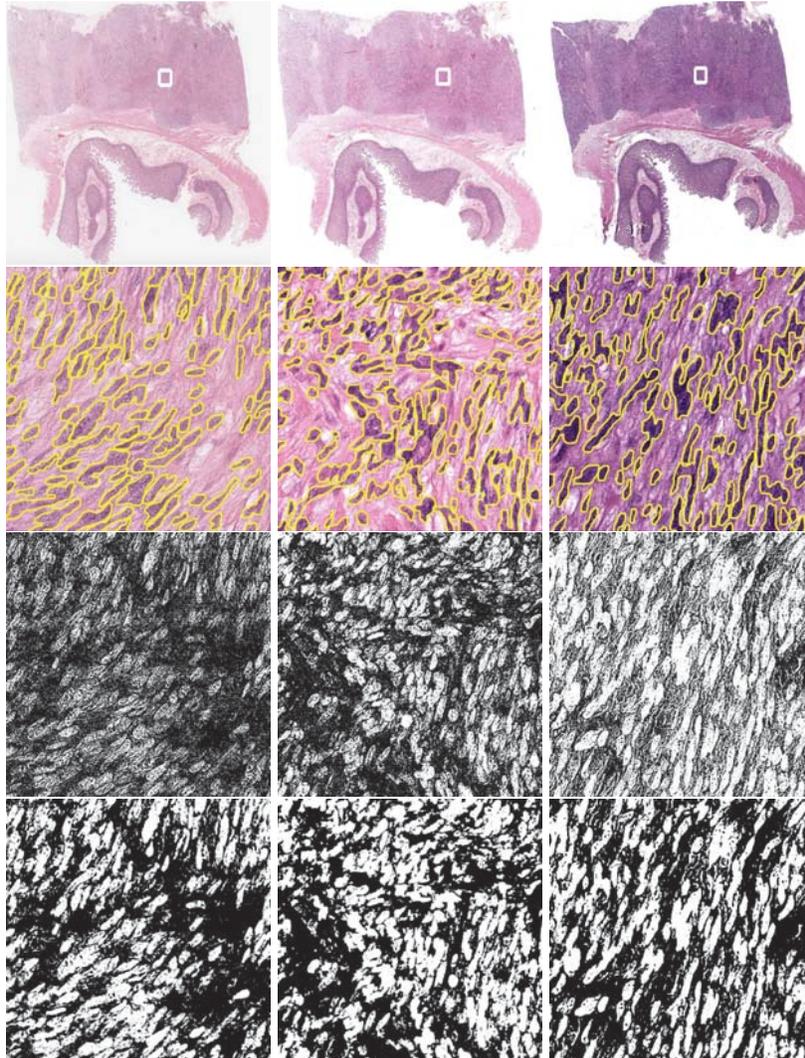


Fig. 1. Row 1) H&E stained GI specimens: (left to right) under-stained in both H and E ($\downarrow H \downarrow E$), normal (HE), and over-stained with both H and E ($\uparrow H \uparrow E$). Row 2) Magnified view of white boxes in the first row (boxes in Row 1 shown at 5x size to ensure visibility); yellow lines indicate manually delineated nuclear boundaries used as ground-truth. Row 3) Probability maps: $P(X_r = 1 | y_r, \theta_{hue})$. Row 4) Probability maps: $P(X_r = 1 | y, \theta)$.

reasonably discriminative feature. To model hue, an angular variable, we select the von Mises density. This leads to the following conditional distribution:

$$P(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}) = \prod_{r \in R} P(y_r | x_r, \boldsymbol{\theta}) = \prod_{r \in R} \exp\{\kappa_{x_r} \cos(y_r - \mu_{x_r})\} / [2\pi I_0(\kappa_{x_r})] \quad (4)$$

where μ_{x_r} is the mean, κ_{x_r} is the concentration (i.e. shape), and I_0 is the modified Bessel function; the subscript x_r indicates that the means and concentrations are functions of the class x_r .

Prior Distribution. The Markov prior models spatial coherency. We establish this prior by defining its local conditional probability density functions, electing to use the Ising formulation:

$$P(x_r | \mathbf{x}_{\eta_r}) = \frac{1}{Z(\eta_r, \beta)} \exp \left\{ \beta \sum_{s \in \eta_r} \delta(x_r - x_s) \right\}, \quad (5)$$

where $\beta \in \mathbb{R}$ and $Z(\eta_r, \beta) = \sum_{x_r} \exp \left\{ \beta \sum_{s \in \eta_r} \delta(x_r - x_s) \right\}$ is the normalizing constant. The neighborhood η_r is the typical 4-connected region.

3.2 Parameter Estimation Using SCEM

The parameter vector for the MRF model is $\boldsymbol{\theta} = [\boldsymbol{\theta}_{hue} \boldsymbol{\theta}_{mrf}]^T = [\mu_1 \kappa_1 \mu_2 \kappa_2 \beta]^T$. Due to variabilities in staining, these parameters will vary from image to image. Consequently, we use SCEM to estimate $\boldsymbol{\theta}$ for each image. Inserting (4) and (5) into (2) yields the following EM equations:

$$\begin{aligned} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^t) = & \sum_{r \in R} P(X_r = 1 | \mathbf{y}, \boldsymbol{\theta}^t) [\kappa_1 \cos(y_r - \mu_1) - \ln I_0(\kappa_1) - \ln 2\pi] + \\ & \sum_{r \in R} P(X_r = 2 | \mathbf{y}, \boldsymbol{\theta}^t) [\kappa_2 \cos(y_r - \mu_2) - \ln I_0(\kappa_2) - \ln 2\pi] + \\ & \sum_{r \in R} \sum_{\mathbf{x}_r \cup \eta_r} P(\mathbf{x}_r \cup \eta_r | \mathbf{y}, \boldsymbol{\theta}^t) \left[-\ln Z(\eta_r, \beta) + \beta \sum_{s \in \eta_r} \delta(x_s - x_r) \right] \quad (6) \end{aligned}$$

Notice that each of the three terms in the sum can be maximized independently. The first two yield analytical solutions. The final requires a numerical method; we employed the Nelder-Mead simplex. (Note that the initial conditions were chosen empirically, and remain fixed across all images.)

3.3 Pixel-Wise Classification

Having determined $\boldsymbol{\theta}$ with SCEM, classification becomes straightforward. We simply estimate $P(x_r | \mathbf{y}, \boldsymbol{\theta})$ with (3). Employing $P(x_r | \mathbf{y}, \boldsymbol{\theta})$ for classification is called maximum posterior marginal (MPM) estimation. Instead of MPM, we could have chosen maximum *a posteriori* (MAP) estimation, implementing it with iterated conditional modes [7] or simulated annealing [11]. However, MPM is the natural choice since $P(x_r | \mathbf{y}, \boldsymbol{\theta})$ is already calculated during the SCEM iteration. As a final step, the probabilities are converted to hard classifications via thresholding.

Table 1. Comparison of AUC values for both algorithms across sub-images; comparison of SCEM-estimated beta $\hat{\beta}$ with β derived from ground-truth

Stain	Area Under the ROC Curve		Markov Parameter		
	$P(X_r=1 y_r, \theta_{hue})$	$P(X_r=1 \mathbf{y}, \theta)$	β	$\hat{\beta}$	$ \beta - \hat{\beta} /\beta$
HE	0.761	0.823	2.826	2.936	0.0391
H↓E	0.787	0.860	3.149	2.936	0.0677
H↑E	0.735	0.823	3.111	2.845	0.0853
↓HE	0.638	0.781	2.976	2.741	0.0789
↓H↓E	0.717	0.834	2.765	2.867	0.0367
↑HE	0.803	0.867	2.978	2.823	0.0520
↑H↑E	0.814	0.881	3.102	2.845	0.0829
Mean	0.748	0.838	2.987	2.856	0.0632

4 Results of Nuclear Segmentation

4.1 Dataset

We cut seven consecutive sections from a block of paraffin embedded GI tissue. To simulate the common variability resulting from differing staining protocols, we dyed the specimens as follows: HE, H↓E, H↑E, ↓HE, ↓H↓E, ↑HE, and ↑H↑E, where ↑ and ↓ indicate over- and under-staining of the specified dye. Each slide was then digitized using an Aperio whole-slide scanner at 40x apparent magnification (0.25 μm per pixel). From each image, we extracted the same 768x793 sub-image for which we manually delineated the nuclear regions, establishing the “ground-truth” (Figure 1). We used these seven sub-images and their associated ground-truths to evaluate our algorithm.

4.2 Results

For each sub-image, we perform the segmentation algorithm described in the previous section. For comparison, we also apply an additional algorithm, identical to the first with the exception that it ignores spatial information. That is, $P(x_r|\mathbf{x}_{\eta_r}, \theta) = P(x_r|\theta)$. In this instance, SCEM devolves into standard EM and $P(x_r|\mathbf{y}, \theta)$ reduces to $P(x_r|y_r, \theta_{hue})$.

Both of these algorithms create a probability map (either $P(X_r=1|y_r, \theta_{hue})$ or $P(X_r=1|\mathbf{y}, \theta)$) for each image. Please see Figure 1 for illustrations. Using each map, we generate a single receiver operator characteristic (ROC) curve and then measure the area under the curve (AUC). The resulting AUC values are listed in Table 1. Both algorithms perform consistently across all images, validating their robustness to color nonstandardness. However, including spatial constraints increases AUC performance substantially. Table 1 also reports our SCEM estimates of β along with the values obtained from the ground-truth (estimated using pseudo-likelihood). The relative difference between the two estimates never exceeds 9%. (Note that for MCMC estimation, we use $b=5$, $m=15$, and $c=4$; all values were chosen empirically.)

5 Conclusion

Color nonstandardness poses a significant problem in the computerized analysis of histopathology. In this work, we presented a novel Bayesian color segmentation algorithm that is highly robust to such nonstandardness; this algorithm employed a novel spatially-constrained extension of the EM algorithm, called spatially-constrained EM (SCEM), to dynamically estimate the probability density functions that describe the color and spatial properties of salient objects. Evaluating this algorithm over a dataset consisting of sections of gastrointestinal tissue stained using different protocols (to simulate severe color nonstandardness) we found that 1) the algorithm performed consistently across images and 2) the inclusion of spatial constraints increased AUC performance by over 12%.

An important goal of this paper was to introduce the SCEM algorithm. Unlike previous approaches for incorporating Markov priors into the EM framework, SCEM does not limit the type of Markov model (as in [5]) nor the parameters that can be estimated (as in [4]). On a final note, we should mention that SCEM is not specific to nuclear segmentation, or even image segmentation, but instead is applicable to any task requiring the unsupervised estimation of MRF models.

References

1. Finlayson, G., et al.: Color by correlation: A simple, unifying framework for color constancy. *IEEE Trans. on Patt. Anal. and Mach. Intel.* 23, 1209–1221 (2001)
2. Madabhushi, A., Udupa, J.: New methods of mr image intensity standardization via generalized scale. *Medical Physics* 33(9), 3426–3434 (2006)
3. Monaco, J., et al.: Image segmentation with implicit color standardization using cascaded em: Detection of myelodysplastic syndromes. In: *ISBI (2012)*
4. Comer, M.L., Delp, E.J.: The em/mpm algorithm for segmentation of textured images: analysis and further experimental results. *IEEE Trans. Imag. Proc.* 9(10), 1731–1744 (2000)
5. Nikou, C., et al.: A class-adaptive spatially variant mixture model for image segmentation. *IEEE Trans. on Imag. Proc.* 16(4), 1121–1130 (2007)
6. Marroquin, J.L., Vemuri, B.C., Botello, S., Calderon, F., Fernandez-Bouzas, A.: An accurate and efficient bayesian method for automatic segmentation of brain mri. *IEEE Trans. on Med. Imag.* 21(8), 934–945 (2002)
7. Besag, J.: On the statistical analysis of dirty pictures. *J. of the Roy. Stat. Soc. Series B (Methodological)* 48(3), 259–302 (1986)
8. Monaco, J., Madabhushi, A.: Weighted maximum posterior marginals for random fields using an ensemble of conditional densities from multiple markov chain monte carlo simulations. *IEEE Trans. on Med. Imag.* 30, 1352–1364 (2011)
9. Marroquin, J., Mitter, S., Poggio, T.: Probabilistic solution of ill-posed problems in computational vision. *J. of the Amer. Stat. Assoc.* 82(397), 76–89 (1987)
10. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the em algorithm. *J. of the Roy. Stat. Soc. Series B* 39(1), 1–38 (1977)
11. Geman, S., Geman, D.: Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. *IEEE Transactions on Pattern Recognition and Machine Intelligence* 6, 721–741 (1984)