Highlight and Shading Invariant Color Image Segmentation Using Simulated Annealing

Paul Fieguth and Slawo Wesolkowski

Systems Design Engineering
University of Waterloo
Waterloo, Ontario, Canada, N2L-3G1
{swesolko,pfieguth}@uwaterloo.ca
http://ocho.uwaterloo.ca/~pfieguth/

Abstract. Color constancy in color image segmentation is an important research issue. In this paper we develop a framework, based on the Dichromatic Reflection Model for asserting the color highlight and shading invariance, and based on a Markov Random Field approach for segmentation. A given RGB image is transformed into a R'G'B' space to remove any highlight components, and only the vector-angle component, representing color hue but not intensity, is preserved to remove shading effects. Due to the arbitrariness of vector angles for low R'G'B' values, we perform a Monte-Carlo sensitivity analysis to determine pixel-dependent weights for the MRF segmentation. Results are presented and analyzed.

1 Introduction

In recent years the problem of color constancy — the perception of objects in the real world without illumination effects — has been a major research subject in the image science and technology communities. In spite of shading and highlight effects, humans are quite able to perceive object surfaces in a scene, a difficult task for computer systems. An algorithm for color image segmentation, which is invariant to shading and highlight effects, has recently been introduced [23], developed in the context of the Dichromatic Reflection Model of Shafer [12].

In [23] the authors describe a principal component analysis and vector angle clustering-based approach for color image segmentation. In this method, the prototype vector is described as the principal vector (as opposed to principal curve) of the RGB color cluster and the calculation of the distance from this "cluster center" to a pixel in the image is done using the vector angle. The number of clusters is selected and the algorithm chooses the most optimal (in the Mean Squared Error-sense) multi-vector fit to the data [3]. The illumination invariances are well captured by this method, however there are several drawbacks:

1. For small (black) RGB values the algorithm breaks down and produces extremely noisy angles.
2. All colors must fit into a predetermined number of clusters.
3. Border areas composed of composite colors are classified arbitrarily.

Certainly a wide variety of color-segmentation approaches have been proposed. In particular, methods based on color clustering have seen considerable interest, including k-means [15, 21], fuzzy k-means [7], and morphology-based clustering [9]. The most notable drawback of such clustering methods is that they normally do not take any spatial relationships into account, and determine the segmentation strictly on a pixel-by-pixel basis, normally using the Euclidean distance. We will demonstrate for the problems of our interest, specifically the segmentation of images involving illumination effects, some degree of spatial dependence is *crucial* in formulating an adequate approach. The ability for Markov/Gibbs methods to model spatial dependencies will make them a very natural fit to our context.

Ad-hoc local methods have also been proposed for color image segmentation such as [5, 18, 20]. In [5], the authors present a method based on the calculation of principal components of local non-overlapping regions to estimate the region color. The method is also said to be highlight and shading invariant. [18] describes a method based on a region growing technique using the Euclidean distance as a similarity measure which is tested on images of homogenous color. [20] presents another region growing technique in which each region is defined by two values: the color gradient (calculated using the Euclidean distance) between two adjacent pixels and the maximum distance between two colors within this region. The first algorithm suffers from having to quantize the region segmentation information while the last two use the Euclidean distance. All three methods are based on various heuristics.

The focus of the present paper is to formulate a color image processing and segmentation technique in the context of the Dichromatic Reflection Model [12, 19], which is introduced in Section 2. The crucial question is how to measure the *similarity* of two colors. Most previous methods assess the relationship between two multispectral (including color) pixels based on the Euclidean distance [7, 9, 21, 22]. The Euclidean distance is often chosen for its simplicity, mathematical tractability, and is well-suited to feature spaces having an isotropic distribution (for color, a good example is the CIE Luv space [11]). However in the case of color images, where each pixel is represented as a RGB vector, the Euclidean distance is a particularly poor measure of color similarity because the RGB space is *an*-isotropic, especially when lighting effects such as specular reflection and shading are present in the image.

In this paper, we propose to use the Dichromatic Reflection Model to transform the RGB image into a different space in which shading and specular reflection are normalized. In this context, highlight and shading invariant color image segmentation means the finding of regions, homogenous in color, irrespective of illumination effects.

Therefore, given the Dichromatic Reflection Model, why can the transformed pixels not be clustered effectively using k-means [10] or other related techniques? The problem is that sufficiently dark shades of any color all look alike (i.e., black), and similarly specular reflections or highlights converge to the same color (the color of the illuminating light, normally white). For example, Figure 1 clearly

Fig. 1. Original RGB color scene image, showing highlights and shading, captured using white light.

illustrates highlights (glossy white image patches) and shading (intensity variations on the surface of each fruit, taken under white light illumination. Consequently, some sort of *spatial* model is *essential* in order to perform segmentation, to assign a highlight pixel to a colored group based on its surrounding context.

We propose to define the spatial context using a Gibbs/Markov approach, as outlined in Section 3. Certainly others have used Markov random fields for image segmentation [1, 6, 24]; however, normally these methods involve Gauss-Markov random fields, where the GMRF defines a spatial texture for the R, G, B components, from which segmentation can proceed as a separate hypothesis-testing procedure applied to the GMRF likelihood [8]. Our approach is quite different: we wish to find the segmented image directly as the result of energy minimization of some appropriately-defined Gibbs random field. Furthermore the regions are not distinguished on the basis of texture, rather on shading and highlight invariant color. That said, textured surfaces where the pixel variations are due to local shading effects (such as the surface of an orange) will be segmented correctly, since the normalized color is similar for all such pixels; whereas textures with intrinsically different colors (such as marble or paisley) are not the focus of our approach.

The formulation of our Gibbs model will be similar to others used for segmentation [4, 6] except for a number of variations due to the peculiarities of our transformed space. We demonstrate the advantages of constructing an energy function for Markov Random Field-driven image segmentation using a measure related to the inner vector product.

This paper first describes the Dichromatic Reflection model and a development of an optimization criterion for segmentation. Next, results on an artificial image and a real scene image are presented and analyzed. Finally, conclusions and directions for future work are given.

2 Color Theory

The Dichromatic Reflection Model [12, 19] will be used in this paper to show highlight and shading invariance properties of the new algorithm. First, the DRM will be introduced. Next, the highlight invariance property will be briefly explained. Finally, how shading invariance is achieved will be described.

2.1 Preliminaries

The Dichromatic Reflection Model purports to separate light reflected from objects into two different types:

1. specular reflection or highlight characterized visually by a glossy appearance and describing light that is reflected in a mirror-like fashion from a surface;
2. diffuse or body reflection which is the light reflected from a surface in all directions, giving a surface its usual colored appearance.

This model has been described for a variety of materials [16]; the focus here will be on inhomogeneous dielectric materials such as plastics. The presentation of the DRM follows closely that given in [23]. First, light reflected from an object surface o (called the color signal) is described as a function $C^o(\lambda, x)$ of wavelength λ and pixel location x:

$$C^o(\lambda, x) = \text{Body Reflection} + \text{Interface Reflection} \tag{1}$$
$$= \alpha(x) S^o(\lambda) E(\lambda) + \beta(x) E(\lambda) \tag{2}$$

where $E(\lambda)$ is the spectral power distribution of a light source, $S^o(\lambda)$ is the spectral-surface reflectance of an object o, $\alpha(x)$ is the shading factor and $\beta(x)$ is a scalar factor for the specular reflection term. The following set of equations can then represent the sensor responses for a camera using R, G, and B coordinates:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \int C^o(\lambda, x) \begin{bmatrix} R_R(\lambda) \\ R_G(\lambda) \\ R_B(\lambda) \end{bmatrix} d\lambda \tag{3}$$

where $R_i(\lambda)$, $(i = R, G, B)$ are the spectral sensitivity functions of the camera in the visible spectrum. Substituting (2) into (3), we have

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \alpha(x) \int S^o(\lambda, x) E(\lambda) \begin{bmatrix} R_R(\lambda) \\ R_G(\lambda) \\ R_B(\lambda) \end{bmatrix} d\lambda + \beta(\lambda) \int E(\lambda) \begin{bmatrix} R_R(\lambda) \\ R_G(\lambda) \\ R_B(\lambda) \end{bmatrix} d\lambda \tag{4}$$
$$= \alpha(x) \boldsymbol{c}_b + \beta(x) \boldsymbol{c}_i \tag{5}$$

where $\boldsymbol{c}_b$ is the body color vector and $\boldsymbol{c}_i$ is the illumination color vector. These color vectors are normalized into a unit vector length.

For the sensor outputs R, G, and B to be white balanced, it is necessary to satisfy the following condition:

$$\int E(\lambda)R_R(\lambda)d\lambda = \int E(\lambda)R_G(\lambda)d\lambda \tag{6}$$

$$= \int E(\lambda)R_B(\lambda)d\lambda \tag{7}$$

As long as the illuminant $E(\lambda)$ is a constant white over the visible wavelengths, and the spectral sensitivity functions $R_i(\lambda)$, $(i = R, G, B)$ have the same area, then the above condition obviously holds. However, if the illuminant is not white, a color balancing step [23] is needed where the three sensor outputs are adjusted to be equal to each other. In this paper it will be assumed that the illumination light is white or the image has been white balanced.

2.2 Highlight Invariance

To remove the effects of highlights it is necessary to transform the pixel coordinates according to the following transformation [14, 23]:

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} R \\ G \\ B \end{bmatrix} - AVG \tag{8}$$

where AVG represents the average value of R, G and B. In this transformation, the reflectance variation caused by interface reflection is removed by projecting the observed reflectance in an n-dimensional vector space along the illumination vector onto an (n-1)-dimensional subspace that is perpendicular to the illumination vector [14]. From a practical point of view, a histogram of the RGB pixels making up a homogeneously-colored region containing a highlight patch would show two connected clusters (one for the homogenous color and one for the highlight).

For example, Figure 2 shows such a distribution in the RGB space of pixels from Figure 1. The four clusters appear highly spread-out and are non-linear (do not lie along a straight line in RGB space), because each cluster is composed of both body and specular reflections.

The transformation (8) transforms each set of nonlinear clusters into a single linear cluster representing the body reflection. This is well illustrated in Figure 3, where the original nonlinear clusters now appear as linear groupings. Given that the RGB components are assumed to be white balanced, the application of (5)–(8) eliminates the interface reflection term and reduces to

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \alpha(x) \int S^o(\lambda, x)E(\lambda)\frac{1}{3}\begin{bmatrix} 2R_R(\lambda) - R_G(\lambda) - R_B(\lambda) \\ -R_R(\lambda) + 2R_G(\lambda) - R_B(\lambda) \\ -R_R(\lambda) - R_G(\lambda) + 2R_B(\lambda) \end{bmatrix} d\lambda \tag{9}$$

$$= \alpha(x) \int S^o(\lambda, x)E(\lambda)\begin{bmatrix} R'_R(\lambda) \\ R'_G(\lambda) \\ R'_B(\lambda) \end{bmatrix} d\lambda \tag{10}$$

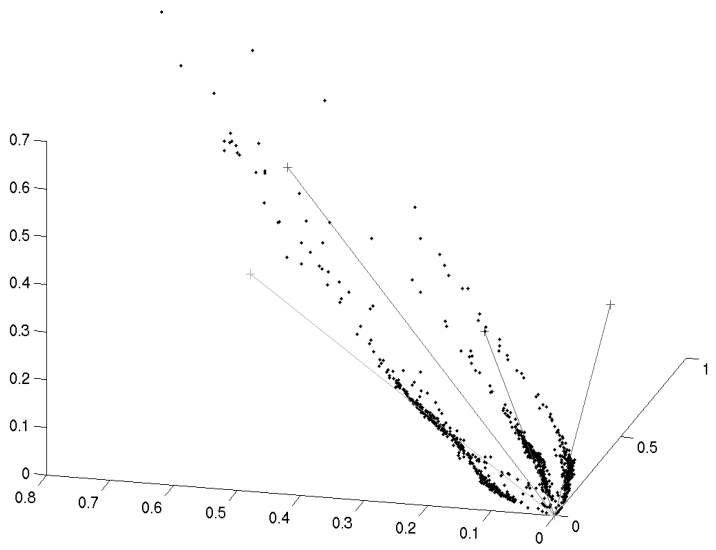


Fig. 2. Distribution of pixels in the RGB space from Figure 1. The straight lines are the principal vectors obtained with the best MSE fit from [23]. Both the red and orange fruits have been clumped into one larger cluster. Whereas three of the four clusters depicted in the image correspond to fruit colors, the fourth represents all of the highlight areas.

This formulation is dependent on the shading factor (illumination) and the body reflection (material color), which makes this color representation highlight invariant. Individual elements of the pixel vector in the new representation will be shifted according to the average of the body reflection term. This results in the new space having negative coordinates. Equivalently the spectral sensitivity functions, $R'_R(\lambda)$, $R'_G(\lambda)$, and $R'_B(\lambda)$, in the new system also have negative values. Three properties were derived from this representation. The first property says that all RGB colors fall into one of six quadrants. The second one says that all gray values (including saturated highlight areas) naturally collapse to the (0,0,0) point. Finally, the third property demonstrates that the same color can only exist in quadrants that have at least one adjacent edge.

2.3 Shading Invariance

Insuring a shading invariance property of the algorithm means that the shading factor shown first in (2) needs to be eliminated from the representation obtained using (10). The simplest way to do this is to normalize the new color vectors to

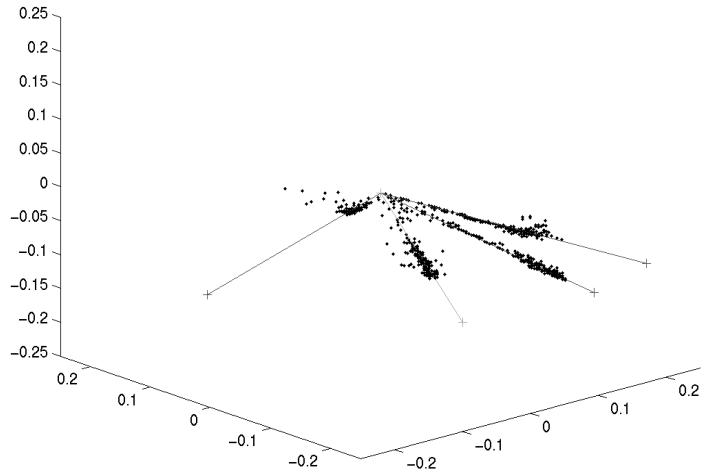


Fig. 3. Distribution of pixels in the R'G'B' space from Figure 1; compare with the RGB distribution in Figure 2. The straight lines are the principal vectors obtained with the best MSE fit from [23]. The alignment of the four cluster prototypes with the four color clusters is clearly seen; each of the four cluster prototypes corresponds to a colored fruit.

unit length [14]. First, reformulate (10) as

$$\boldsymbol{c}' = \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \alpha(x) \int S^o(\lambda, x) E(\lambda) \begin{bmatrix} R'_R(\lambda) \\ R'_G(\lambda) \\ R'_B(\lambda) \end{bmatrix} d\lambda = \alpha(x) \begin{bmatrix} c^o_R \\ c^o_G \\ c^o_B \end{bmatrix} \tag{11}$$

where c^o_R, c^o_G, c^o_B represent the non-factorable terms of (8). Now normalizing the color vector, $\boldsymbol{c}'$, we obtain:

$$\frac{\boldsymbol{c}'}{|\boldsymbol{c}'|} = \frac{\alpha(x) \begin{bmatrix} c^o_R \\ c^o_G \\ c^o_B \end{bmatrix}}{(\alpha^2(x)[(c^o_R)^2 + (c^o_G)^2 + (c^o_B)^2])^{1/2}} \tag{12}$$

$$= \frac{\begin{bmatrix} c^o_R \\ c^o_G \\ c^o_B \end{bmatrix}}{((c^o_R)^2 + (c^o_G)^2 + (c^o_B)^2)^{1/2}} \tag{13}$$

The shading factor has been eliminated and hence this representation is intensity invariant. This operation puts all vectors on the unit hypersphere, except for the null vector (0,0,0) for which this operation is undefined. Since the Euclidean distance between two normalized transformed color vectors does not reflect accurately the perceptual difference between the two vectors, we propose to factor the invariance operation directly into the similarity measure calculation by using one minus the cosine of the vector angle $\theta_{\boldsymbol{c}',\boldsymbol{d}'}$ between two transformed color vectors $\boldsymbol{c}', \boldsymbol{d}'$; the similarity measure then becomes

$$\Omega(\theta_{\boldsymbol{c}',\boldsymbol{d}'}) = 1 - \frac{<\boldsymbol{c}', \boldsymbol{d}'>}{|\boldsymbol{c}'||\boldsymbol{d}'|} \tag{14}$$

$$= 1 - \frac{\alpha_c(x)\alpha_d(x)(c_R^o d_R^o + c_G^o d_G^o + c_B^o d_B^o)}{(\alpha_c^2(x)(c_R^o c_R^o + c_G^o c_G^o + c_B^o c_B^o))^{1/2}(\alpha_d^2(x)(d_R^o + d_G^o + d_B^o))^{1/2}} \tag{15}$$

$$= 1 - \frac{(c_R^o d_R^o + c_G^o d_G^o + c_B^o d_B^o)}{(c_R^o c_R^o + c_G^o c_G^o + c_B^o c_B^o)^{1/2}(d_R^o + d_G^o + d_B^o)^{1/2}} \tag{16}$$

So if $\boldsymbol{c}'$ and $\boldsymbol{d}'$ are similar in orientation then (16) will be close to zero. Both vectors will be deemed close irrespective of the shading factors $\alpha_c(x)$ and $\alpha_d(x)$ associated with them. Therefore, this method is also shading invariant. In practice, the color vectors R', G', B' are normalized as in (13), which reduces (16) to a simple dot product calculation for each pixel comparison.

2.4 Angle Accuracy

The principal problem with the vector angle formulation derives from the consequences of using (8), which collapses all graylevel values to the origin in the transformed domain, and (13), which performs a vector normalization. Stated more plainly, a collection of noisy, nearly black pixels will be normalized to vastly different transformed locations. This is strictly a reflection of the large degree of sensitivity in the definition of a "hue" for nearly black pixels.

Standard clustering approaches either require such pixels to be rejected (needing an arbitrary rejection threshold) or incorporate them, leading to misleading conclusions. The elegance of the MRF approach to a segmentation algorithm, set up in the next section, is that the penalty term associated with a vector angle can be continuous, rather than discrete admit/reject.

The noise sensitivity of the similarity measure (16) is easily computed, as a preprocessing step, using Monte-Carlo means. In particular, if we model two pixels as noisy

$$\boldsymbol{c} = \boldsymbol{c}_{exact} + \text{noise} \tag{17}$$

$$\boldsymbol{d} = \boldsymbol{d}_{exact} + \text{noise} \tag{18}$$

then the variance $\text{var}(\Omega(\theta_{\boldsymbol{c}',\boldsymbol{d}'}))$ can be efficiently computed; clearly if this variance in angle difference is small then the accuracy of the angle calculation will be deemed high and will be weighted more heavily in the Gibbs energy.

3 Markov Random Fields

The modeling problems in this paper are addressed from the computational viewpoint. There are two primary concerns: how to define an objective function for the optimal solution of the image segmentation problem, and how to find its optimal solution. Given the various uncertainties in the imaging process, it is reasonable to define the desired solution in an optimization sense, such that the "perfect" or "exact" solution to our segmentation problem is interpreted as the optimum solution to the optimization objective.

Some forms of contextual constraints are eventually necessary when trying to interpret visual information. The spatial and visual contexts of the objects in an image scene are necessary for the understanding of the scene; the context of object features at a lower level of representation allow the recognition of the objects; the context of primitives at an even lower level lets the object features be identified; and finally the context of image pixels at the lowest level of abstraction allows for the extraction of those primitives. To create a reliable and effective image analysis system the use of contextual constraints is unavoidable and therefore indispensable.

Gibbs Random Fields (GRFs) [4, 24] provide a natural way of modeling context dependencies between, for example, image pixels of correlated local features [6]. The practical use of GRF models is largely possible due the improved insights and understanding provided by the Hammersley Clifford theorem [6], which allows Markov random field (MRF) modeling to be reinterpreted as an energy function minimization. The second motivating development is the improved insight and available methods for Gibbs sampling and Simulated Annealing.

The MRF-based segmentation model is defined by the contextual relationships within the local neighborhood structure. Since our goal is the assertion of local constraints, rather than an accurate modeling of spatial textures, as in other GMRF color-segmentation research [8], we shall only be concerned with first order random fields, both simplifying the model and limiting the computational complexity.

Suppose we are given a color image on a pixel lattice $\mathcal{L} = \{i, j\}$. As just discussed in Section 2, each pixel $\{RGB\}_{i,j}$ is transformed to its normalized representation $\boldsymbol{c}'_{i,j}$.

If we precompute the adjacent-pixel vector-angle criteria

$$\psi_{i,j} = \Omega(\theta_{\boldsymbol{c}'_{i,j},\boldsymbol{c}'_{i+1,j}}) \quad \phi_{i,j} = \Omega(\theta_{\boldsymbol{c}'_{i,j},\boldsymbol{c}'_{i,j+1}}) \tag{19}$$

then a Gibbs energy E for segmentation can be formulated as follows:

$$E[\{l(i,j)\}] = \sum_{i,j} \alpha \left[\psi^2_{i,j}\delta_{l(i,j),l(i+1,j)} + \phi^2_{i,j}\delta_{l(i,j),l(i,j+1)}\right] + \beta\left[(1-\delta_{l(i,j),l(i,j+1)}) + (1-\delta_{l(i,j),l(i,j+1)})\right] \tag{20}$$

where each pixel (i, j) is assigned an integer label $0 \leq l(i, j) < N$, and where α, β control the relative constraints on the homogeneity of a single region and the degree of region fragmentation, respectively.

Fig. 4. Boundary length problem: both regions have the same boundary length, although very different volumes.

The model (20) is intuitive and easily implemented. As mentioned before, it deviates from previous models used for image segmentation in that the inner vector product Ω is used to calculate the minimum energy instead of the Euclidean distance in R, G, B. However it misses one essential point: not all of the vector angles are computed with the same accuracy. Even a small amount of pixel noise on a dark or highlight region results in nearly totally random vector angles, which (20) would choose to separate into single-pixel regions. Given the covariance of the vector angle difference, computed by analytic or Monte-Carlo means as discussed in Section 2.4, we introduce weights

$$w_{i,j} = \frac{1}{\mathrm{var}(\Omega(\theta_{\boldsymbol{c'}_{i,j},\boldsymbol{c'}_{i+1,j}}))} \quad v_{i,j} = \frac{1}{\mathrm{var}(\Omega(\theta_{\boldsymbol{c'}_{i,j},\boldsymbol{c'}_{i,j+1}}))} \tag{21}$$

to assert the degree of confidence of the terms in the energy:

$$E[\{l(i,j)\}] = \sum_{i,j} \alpha \left[w_{i,j}\psi_{i,j}^2 \delta_{l(i,j),l(i+1,j)} + v_{i,j}\phi_{i,j}^2 \delta_{l(i,j),l(i,j+1)} \right] + \beta \left[(1 - \delta_{l(i,j),l(i,j+1)}) + (1 - \delta_{l(i,j),l(i,j+1)}) \right] \tag{22}$$

Model (22) is a very credible segmentation criterion, representing a considerable advance beyond standard vector-angle methods, and yet (22) is little more complicated than a standard Ising/Potts model [24] and so is well-understood and easily implemented.

The primary drawback with (22) is that it is strictly a local, pixel-neighbor model and suffers from the same problems as other region-growing approaches: two vastly differently colored pixels may be grouped into a single region if they are linked by noisy or intermediately-colored pixels. A second undesired effect is that N constrains only the number of region labels, not the number of regions; that is, in regions of noise or color-gradients, (22) can generate a proliferation of small regions. Finally, the label criterion, controlled by β, measures boundary length, rather than region volume (see Figure 4). Therefore, in regions where the vector-angle criterion is vague (that is, in saturated or dark regions), a large number of pixels may have to be flipped to see *any* change in the energy, implying that only the slowest of annealing schedules will successfully converge.

A global model can overcome these drawbacks. If we associate with label l a global transformed color a'_l then each region is forced to be well defined:

$$E[\{l(i,j), \boldsymbol{a}'\}] = \sum_{i,j} \Omega(\theta_{\boldsymbol{a'}_{l(i,j)},\boldsymbol{c'}_{i,j}})^2 + \beta \left[(1 - \delta_{l(i,j),l(i,j+1)}) + (1 - \delta_{l(i,j),l(i,j+1)}) \right] \tag{23}$$

For the purposes of this paper, we propose to fix the region colors $\{a_l\}$; that is, the sampling and annealing takes place only over the label indices $\{l(i,j)\}$ themselves. The $\{a_l\}$ would be found by a preceding step, such as vector quantization [21].

A final modification mirrors that of (22): the degree to which the region color is to be asserted at each pixel should be spatially-varying, now for two reasons:

1. The color-dependent effect of noise, particularly for dark and highlight pixels.
2. We are normally not interested in pixels in regions of high color gradient; at the very least, these pixels should not unduly influence the Gibbs energy by being inconsistent with the region color $\boldsymbol{a}$.

If we let

$$u_{i,j} = \min\left\{ \frac{1}{\mathrm{var}(\Omega(\theta_{\boldsymbol{c'}_{i,j},\boldsymbol{c'}_{i+1,j}}))}, \frac{1}{\mathrm{var}_{\mathcal{N}}(\Omega(\theta_{\boldsymbol{c'}_{i,j},\boldsymbol{c'}_{i+1,j}}))} \right\}, \tag{24}$$

that is, the variances are the pointwise one, based on a noise model, and a spatial one, computed over a local neighborhood $\mathcal{N}$, then our segmentation model becomes

$$E[\{l(i,j), \boldsymbol{a}_l\}] = \sum_{i,j} u_{i,j}\Omega(\theta_{\boldsymbol{a'}_{l(i,j)},\boldsymbol{c'}_{i,j}})^2 + \beta\left[(1-\delta_{l(i,j),l(i,j+1)}) + (1-\delta_{l(i,j),l(i,j+1)})\right] \tag{25}$$

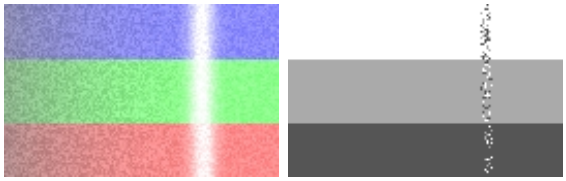
This gives us a concise and coherent representation of the color image segmentation problem by incorporating both local and global constraints. The global constraints are defined by global color region labels obtained through some vector quantization process such as the one presented in [23]. Local constraints are included by virtue of using pixel level constraints in the MRF model.

Model (25) is a tradeoff between a completely local region growing approach, where many spurious regions can be created, and a global color clustering approach where regions of differing color can be inadvertently merged. Furthermore, the use of vector angle accuracy weights (21) allows the less reliable calculation of vector angle for small R'G'B' values to be appropriately modulated.

4 Results

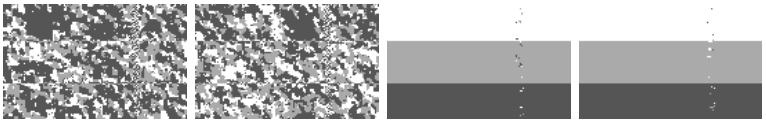
The Gibbs Sampler [4] will be used to optimize both (22) and (25). To make comparison as straightforward as possible, all MRF results were initialized from a random start, although in practice initializing from an MPC or other segmentation could accelerate convergence. For the global model (25) the label colors $\boldsymbol{a}$ are determined using the algorithm presented in [23].

Results were prepared an artificial image of colored bands, shown in Figure 5. The artificial image varies in intensity horizontally (i.e., from left to right and a saturated highlight is present near the right border). Some additive uniform uniformly distributed noise was added to this image.



(a) (b)

Fig. 5. Color band image: (a) Original, (b) MPC segmentation.



(a) Model (20) (b) Model (22) (c) Model (23) (d) Model (25)

Fig. 6. Color Band image: Results of four proposed MRF models.

The MPC result on the artificial image is shown in Figure 5(b). The highlight part is clearly a mixture of the three other segmentation classes due to having a nearly null vector representation in the R', G', B' space, and the absence of spatial constraints prevents the ambiguity from being corrected. For the MRF models, the results in Figure 6(a) and Figure 6(b) clearly illustrate the problems of boundary length discussed in Section 3, because of the lack of region-defining constraints such as characteristic region vector, boundary length or area size constraints. It is interesting to note that under careful examination, regions generated on both sides of the border between each color band pair are seldom part of the same class. Figure 6(c) demonstrates the type of result that is obtained using Model (23). As desired, very few highlight parts remain as other highlight areas have been subsumed into their adjacent regions. The remaining few misclassified pixels are due most probably to a too-rapid annealing schedule.

The free parameter β clearly controls the significance of the color-angle dot product in relation to the spatial label contribution in the energy term; clearly in the limit of a small value of β, the MRF result converges to that of MPC. Finally, Figure 6(d) shows the results for the same color bands, but now where the vector angle calculation is weighted in terms of the accuracy to which to the angle can be determined (which is affected by darkness or degree of highlight), as in (25).

5 Conclusions

We have presented a Markov Random Field-based model for shading and highlight invariant color image segmentation. The model's invariance properties have been verified using the Dichromatic Reflection Model. Furthermore, the model is based on a vector angle difference measure between color vectors and includes

weights to take into account the reliability of calculating angles between various vector pairs.

The MRF model used is a compromise between local-only or global-only color image segmentation methods. It combines the best of both worlds: the ability of the global methods to create well segmented regions and the ability of local methods to adapt to the local variations in pixel values.

There are three immediate considerations for future work. It is not obvious that it is desirable to fix the region colors in models (23), (25). The obvious advantage of doing so is the computational consideration, however the disadvantage is that any error in the vector-quantization step is locked in place and cannot be removed. Instead, the region colors $\{a_l\}$ can be variable, determined as part of the annealing process. Although this requires the Gibbs sampling of continuous values, the effort can remain reasonable if the variables are accurately initialized: $\{a_l\}$ from vector quantization or k-means [10], and the pixel labels $\{l(i,j)\}$ from (22).

Furthermore, the limitation, as illustrated in Figure 4, of using the boundary length as an energy metric for each segmented region, should be revisited. The most obvious choice would be to prefer *larger* regions, where region size is measured by the number of pixels in the region. Although much more robust than boundary length, the number of pixels is a non-local criterion, and is therefore computationally much less convenient.

Finally, parameter estimation to obtain proper convergence of the MRF models is essential. In this paper, parameter estimation was ad-hoc. A formalized parameter estimation technique needs to be applied to fully evaluate the advantages of the MRF models over vector quantization and region growing-based methods when applied to real scene images.

References

1. S. A. Barker, and P. J. W. Rayner, "Unsupervised Image Segmentation Using Markov Random Fields," in M. Pelillo and E. R. Hancock (ed), *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp. 165-178, Springer-Verlag: 1997.
2. R. Chellappa, and A. Jain, *Markov Random Fields: Theory and Application*. Academic Press, New York, 1993.
3. R. D. Dony, and S. Haykin, "Image segmentation using a mixture of principal components representation," *IEE Proc. VISP*, vol. 144, pp. 73-80, April 1997.
4. S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans-PAMI*, Vol. 6, No. 6, 1984.
5. G.J. Klinker, S.A. Shafer and T. Kanade, "A Physical Approach to Color Image Understanding," *Inter'l J. of Computer Vision*, Vol. 4, No. 1, pp. 7-38, 1990.
6. S. Z. Li, "Modeling Image Analysis Problems Using Markov Random Fields," in C.R. Rao and D.N. Shanbhag (ed), *Stochastic Processes: Modeling and Simulation, Vol. 20 of Handbook of Statistics*. Elsevier Science, 2000, pp. 1-43.
7. Y.W. Lim, and S.U. Lee, "On the color image segmentation algorithm based on the thresholding and fuzzy c-means techniques," *Pattern Recognition*, vol. 23, no. 9, pp. 1235-1252, 1990.

8. D. K. Panjwani, and G. Healey, "Markov Random Field Models for Unsupervised Segmentation of Textured Color Images," *IEEE Trans-PAMI*, Vol. 17, No. 10, 1995.
9. S. H. Park, I. D. Yun, and S.U. Lee, "Color Image Segmentation Based on 3-D Clustering: Morphological Approach," *Pattern Recognition*, vol. 31, no. 8, pp. 1061-1076, 1998.
10. R.J. Schalkoff, *Pattern Recognition: Statistical, Structural and Neural Approaches.* John Wiley & Sons, Inc., New York, 1992.
11. L. Shafarenko, M. Petrou, and J. Kittler, "Automatic watershed segmentation of randomly textured color images," *IEEE Trans. on Image Processing*, vol. 6, pp. 1530-1544, November 1997.
12. S.A. Shafer, "Using color to separate reflection components," TR-136, Computer Sciences Dept., University of Rochester, NY, April 1984.
13. S. Tominaga and B. Wandell, "The standard reflectance model and illuminant estimation", *J. of Optical Society of America A*, Vol. 6, No.4, pp. 576-584, April 1989.
14. S. Tominaga, "Surface Identification Using the Dichromatic Reflection Model," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 13, No. 7, pp. 658-670, July 1991.
15. S. Tominaga, "Color Classification of Natural Color Images," *Color Research and Application*, Vol. 17, No. 4, pp. 230-239, 1992.
16. S. Tominaga, "Dichromatic Reflection Models for a Variety of Materials," *Color Research and Application*, Vol. 19, No. 4, pp.277 - 285, 1994.
17. S. Tominaga, "Spectral imaging by a multichannel camera," *Journal of Electronic Imaging*, vol. 8, no. 4, pp. 332-342, 1999.
18. A. Tremeau, and N. Borel, "A Region Growing and Merging Algorithm to Color Segmentation," *Pattern Recognition*, vol. 30, no. 7, pp. 1191-1203, 1997.
19. B.A. Wandell. *Foundations of Vision*, Sinauer Associates, Inc. Publishers, Sunderland, MA, 1995.
20. W. Wang, C. Sun, and H. Chao, "Color Image Segmentation and Understanding through Connected Components," *IEEE International Conference on Systems, Man, and Cybernetics*, vol. 2, pp. 1089-1093, October 1997.
21. S. Wesolkowski, M.E. Jernigan, R.D. Dony, "Global Color Image Segmentation Strategies: Euclidean Distance vs. Vector Angle," in Y.-H. Hu, J. Larsen, E. Wilson and S. Douglas (eds.), *Neural Networks for Signal Processing IX*, IEEE Press, Piscataway, NJ, 1999, pp. 419-428.
22. S. Wesolkowski, *Color Image Edge Detection and Segmentation: A Comparison of the Vector Angle and the Euclidean Distance Color Similarity Measures*, Master's thesis, Systems Design Engineering, University of Waterloo, Canada, 1999.
23. S. Wesolkowski, S. Tominaga, and R.D. Dony, "Shading and Highlight Invariant Color Image Segmentation Using the MPC Algorithm," *SPIE Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts VI*, San Jose, USA, January 2001, pp. 229-240.
24. G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer-Verlag, Berlin, Germany, 1995.