Exploiting Spatial and Spectral Image Regularities for Color Constancy

Barun Singh and William T. Freeman
MIT Computer Science and Artificial Intelligence Laboratory

David H. Brainard
University of Pennsylvania Department of Psychology

October 10, 2003

Abstract

We study the problem of color constancy–inferring from an image the spectrum of the illumination and the reflectance spectra of all the depicted surfaces. This estimation problem is underdetermined: many surface and illumination spectra can be described as a linear combination of 3 basis functions, giving 3 unknowns per pixel, plus 3 parameters for a global illumination. A trichromatic visual system makes fewer measurements than there are unknowns.

We address this problem by writing the reflectance spectra of small groups of pixels as linear combinations of "spatio-spectral" basis functions. These aggregated surface reflectance spectra require fewer parameters to describe than the sum of the spectral parameters for the individual surface pixels, giving us more measurements than unknown parameters.

We explore this problem in a Bayesian context, showing when the problem is over or underdetermined based on analyzing the local curvature characteristics of the log-likelihood function. We show how using the spatio-spectral basis functions might give us improved reflectance and illumination spectral estimates when applied to real image data.

1 Introduction

Color is important in our understanding of the visual world and provides an effective cue for object detection and recognition. In general, however, the observed color of an object differs from its true color due to factors such as lighting and orientation. Color constancy refers to the ability to perceive the color of an object as approximately constant regardless of the color of the light falling upon it [10]. We are able to reliably use color for a variety of vision tasks largely because of our ability to perform color constancy well.

Inherently, the color constancy problem is underdetermined. The image formation process consists of an illumination reflecting off of a surface, then being passed through a set of sensors. The input to the sensors is the product of the illumination spectrum and the surface's reflectance spectrum. Different combinations of illumination and reflectance spectra can produce the same spectrum impinging on the sensor, and various sensor inputs can result in the same sensor responses. Due to these ambiguities, it may be impossible to uniquely separate the effect of the illumination and the surface reflectances in an image. The goal of computational color constancy [21, 17, 6, 18, 3, 20, 13, 14, 24, 2, 9], therefore, is to determine an optimal separation of reflectance and illumination under some metric. Given the sensor responses, we seek an estimate of the reflectance and illumination spectra.

This would appear to involve estimating very many numbers. If we specify a 300 nm spectrum at 10 nm intervals, even assuming a single illumination spectrum over the whole image, we would need $31(N + 1)$ numbers to specify the reflectance spectra for N pixels plus one illuminant. In order to simplify the color constancy problem, low-dimensional linear models are used to describe illumination and reflectance spectra [16, 7, 19, 15]. A sufficiently large portion of the energy in these spectra (typically 99%) - can be described using as low as three-dimensional models for both surfaces and illuminants. An additional constraint used by some approaches is that surfaces and illuminants must be physically realizable, e.g. their spectra cannot be negative and surface reflectances must be less than or equal to 1 [13, 2, 9].

Even after these simplifications, the color constancy problem remains underdetermined and existing approaches must make further assumptions in order to obtain a solution. Buchsbaum's Gray World algorithm assumes that the mean reflectance of all images is the same, and the illumination is estimated using this mean [6]. Maloney and Wandell's Subspace method requires that a two-dimensional model describe surface reflectances for the case of trichromatic sensors [20]. Gamut mapping methods [12, 11] exploit the observed color gamut to estimate the illuminant spectrum, and can be combined with realizability constraints. Since a Bayesian approach [2] can incorporate the various assumptions above into its prior probabilities, Bayesian decision theory [1] provides a principled framework which we will use for studying the color constancy problem.

A parameter counting argument reveals the underlying problem we address in this work: if we use a 3-dimensional model for the surface reflectance at each pixel, and a 3-dimensional model for the illumination spectrum, for the case of trichromatic color sensors, we have more unknowns than observations. We measure 3 numbers at each position, yet have to estimate 3 numbers at each position, plus the 3 numbers describing the illuminant. The problem is always underdetermined, requiring, in a Bayesian solution, more reliance on the prior probabilities or loss function than on the likelihood function to estimate the best answer. One would prefer the situation where the data likelihood function dominated the other terms.

In this paper, we make the problem overdetermined by introducing linear basis func-

tions that describe both spectral and spatial variations in the surface reflectances. These spatio-spectral basis functions exploit image regularities which allow us to specify surfaces reflectances using fewer numbers, yielding more image measurements than unknown parameters.

In the next sections, we introduce the notation and linear basis function analysis, then explore the Bayesian approach. We show with simple examples that the structure of the posterior indicates conditions under which color constancy is still undetermined by the data. We then apply our approach to real images, verifying (using hyperspectral image data) that our surface reflectance estimates improve through using the spatio-spectral basis functions.

2 Linear models

We assume there is one global illumination in each scene and that the light spectrum leaving each surface is the term-by-term product of the illumination and local reflectance spectrum. For a surface reflectance $S(\lambda)$ and illumination $E(\lambda)$, the response at position x of a photoreceptor with a spectral response of $R_k(\lambda)$ is:

$$y_k^x = \sum_\lambda R_k(\lambda)E(\lambda)S^x(\lambda), \tag{1}$$

The illuminant and surface spectra can be written as linear combinations of the illumination basis functions $E_i(\lambda)$ and reflectance basis functions $S_j(\lambda)$, with coefficients e_i and s_j^x, respectively, at position x:

$$E(\lambda) = \sum_{i=1}^{L} E_i(\lambda)e_i, \tag{2}$$

$$S^x(\lambda) = \sum_{j=1}^{L} S_j(\lambda)s_j^x, \tag{3}$$

where L is defined as the number of elements in the λ vector (i.e., the number of wavelength samples). Doing so allows us to write the rendering equation as

$$y_k^x = \sum_\lambda R_k(\lambda) \sum_{i=1}^{L} E_i(\lambda)e_i \sum_{j=1}^{L} S_j(\lambda)s_j^x. \tag{4}$$

Summing over λ, we get a bilinear form,

$$y_k^x = \sum_{i=1}^{L}\sum_{j=1}^{L} e_i G_{ij,k} s_j^x, \tag{5}$$

where $G_{ij,k} = \sum_\lambda R_k(\lambda)E_i(\lambda)S_j(\lambda)$.

The number of basis functions used in the linear models of (2) and (3) are equal to the number of wavelength samples, so that the model fully describes the entire illumination and reflectance spectra exactly. In general, however, we wish to approximate the true illuminant $E(\lambda)$ and surface spectra $S^x(\lambda)$ using lower dimensional linear models. Thus, we can define

$$\tilde{E}(\lambda) = \sum_{i=1}^{d_E} E_i(\lambda)e_i, \tag{6}$$

$$\tilde{S^x}(\lambda) = \sum_{j=1}^{d_S} S_j(\lambda)s_j^x, \tag{7}$$

where d_E is the dimensionality of the illuminant approximation $\tilde{E}(\lambda)$, d_S is the dimensionality of the surface approximation $\tilde{S^x}(\lambda)$ and $0 < (d_S, d_E) < L$. In addition, we can decompose (5) as:

$$\begin{aligned} y_k^x &= \sum_{i=1}^{d_E}\sum_{j=1}^{d_S} e_i G_{ij,k} s_j^x + \sum_{i=1}^{d_E}\sum_{j=d_S+1}^{L} e_i G_{ij,k} s_j^x + \sum_{i=d_E+1}^{L}\sum_{j=1}^{d_S} e_i G_{ij,k} s_j^x + \sum_{i=d_E+1}^{L}\sum_{j=d_S+1}^{L} e_i G_{ij,k} s_j^x \\ &= \sum_{i=1}^{d_E}\sum_{j=1}^{d_S} e_i G_{ij,k} s_j^x + w_k^x, \end{aligned} \tag{8}$$

where w_k^x is the error due to the use of lower dimensional models to approximate the true illumination and reflectance, which we will refer to as "observation noise" for the remainder of this paper. A summary of the notation used in this paper is provided in Table 1.

Equation (8) is perhaps best understood by viewing color constancy as an estimation problem. The goal is to estimate $\tilde{S^x}(\lambda)$ and $\tilde{E}(\lambda)$ given y_k^x. Thus, we view the "true" illuminants as having d_E degrees of freedom as defined by the first d_E illuminant basis functions, with remaining basis functions contributing in the form of noise. Analogously for surface reflectance, the first d_S basis functions define the space of "true" reflectances and the remaining basis functions add noise. Note that while the illumination "noise" and reflectance "noise" are orthogonal to the "true" illumination and reflectance, the projections of the "true" data and the "noise" into the sensor response space are not orthogonal.

3 Bayesian Approach to Color Constancy

The Bayesian approach to color constancy utilizes a probabilistic framework to examine the problem. This framework is based on three fundamental probability distributions - the prior, the likelihood, and the posterior. The prior describes the probability that a certain set of pa-

Symbol	Symbol Definition	Symbol	Symbol Definition
N	number of pixels in an image	m	number of pixels along one dimension of a square block being considered within an image
λ	wavelength	γ	index into spatio-spectral basis functions
L	number of wavelength samples (length of λ vector)	K	number of sensor responses available at any pixel
$E(\lambda)$	spectral radiance of illuminant	$\boldsymbol{E}(\gamma)$	spatio-spectral radiance of illumination in $m \times m$ block
$E_i(\lambda)$	ith basis function of illumination	$\boldsymbol{E}_i(\gamma)$	ith spatio-spectral illumination basis function
e_i	scalar weight of ith illumination basis function	$\boldsymbol{e}_i$	scalar weight of ith spatio-spectral illumination basis function
e	row vector of all e_i	$\boldsymbol{e}$	row vector of all $\boldsymbol{e}_i$
d_E	dimensionality of illumination representation	$\boldsymbol{d_E}$	dimensionality of spatio-spectral illumination representation
$\tilde{E}(\lambda)$	approximation to illuminant from d_E dimensional linear model	$\tilde{\boldsymbol{E}}(\gamma)$	approximation to spatio-spectral illuminant from $\boldsymbol{d_E}$ dimensional linear model
$S^x(\lambda)$	surface reflectance at point x	$\boldsymbol{S^x}(\gamma)$	spatio-surface reflectance for pixel block $\boldsymbol{x}$
$S_j(\lambda)$	jth basis function of surface reflectance	$\boldsymbol{S}_j(\gamma)$	jth spatio-spectral surface reflectance basis function
s_j^x	scalar weight of jth surface reflectance basis function at position x	$\boldsymbol{s}_j^{\boldsymbol{x}}$	weight of jth spatio-spectral surface reflectance basis function for pixel block $\boldsymbol{x}$
s^x	column vector of s_j^x for all j	$\boldsymbol{s^x}$	column vector of $\boldsymbol{s}_j^{\boldsymbol{x}}$ for all j
s	matrix whose columns are given by s^x	$\boldsymbol{s}$	matrix whose columns are given by $\boldsymbol{s^x}$
d_S	dimensionality of surface reflectance representation	$\boldsymbol{d_S}$	dimensionality of spatio-spectral surface reflectance representation
$\tilde{S^x}(\lambda)$	approximation to surface reflectance at position x from d_S dimensional linear model	$\tilde{\boldsymbol{S^x}}(\gamma)$	approximation to spatio-spectral surface reflectance at pixel block $\boldsymbol{x}$ from $\boldsymbol{d_S}$ dimensional linear model
$R_k(\lambda)$	spectral sensitivity of kth sensor	$\boldsymbol{R}_k(\gamma)$	spatio-spectral sensitivity of kth sensor
$R(\lambda)$	matrix of $R_k(\lambda)$ for all k	$\boldsymbol{R}(\gamma)$	matrix of $\boldsymbol{R}_k(\gamma)$ for all k
y_k^x	scalar response of kth sensor at x	$\boldsymbol{y}_k^{\boldsymbol{x}}$	scalar response of kth sensor for pixel block $\boldsymbol{x}$
y^x	vector of y_k^x for all k	$\boldsymbol{y^x}$	vector of $\boldsymbol{y}_k^{\boldsymbol{x}}$ for all k
y	matrix of y^x vectors for all x	$\boldsymbol{y}$	matrix of $\boldsymbol{y^x}$ vectors for all $\boldsymbol{x}$
$G_{ij,k}$	3-d tensor relating illumination and reflectance weights to sensor responses	$\boldsymbol{G}_{ij,k}$	3-d tensor relating spatio-spectral illumination and reflectance weights to spatio-spectral sensor responses
w_k^x	observation noise contained in y_k^x	$\boldsymbol{w}_k^{\boldsymbol{x}}$	observation noise contained in $\boldsymbol{y}_k^{\boldsymbol{x}}$

Table 1: Symbol Notation

rameters is the correct estimate without knowing anything about the data. This distribution is written as $P(e, s)$, where e and s are the global illumination weights and the surface weights for all positions in an image. The likelihood, given by $P(y|e, s)$, governs the relationship between the parameters being estimated (e and s) and the observed data (in this case the set of all sensor responses in an image, y). The posterior describes what is known about the parameters being estimated after having observed the data and can be calculated using Bayes' rule as

$$P(e, s|y) = \frac{1}{Z} P(y|e, s) P(e, s), \tag{9}$$

where the normalization constant $\frac{1}{Z}$ is independent of the parameters to be estimated. (In addition to these probabilities, the Bayesian formulation requires that we specify a cost function that defines the penalty associated with a parameter estimate of $(\tilde{e}, \tilde{s})$ when the true parameter values are (e, s). For estimating the relative surface and illumination spectra, independent of any overall scale factor, we find that the loss function has little effect [2], so we use a minus delta loss function, which corresponds to MAP estimation wherein we maximize the posterior probability given by equation (9).)

The underdetermined nature of the general color constancy problem implies that there should be an infinite set of solutions which maximize the posterior. To see that this is the case, we must examine the properties of equation (9).

Assuming independent and identically distributed Gaussian observation noise with variance σ^2 in the photoreceptor responses at each patch, the posterior probability is:

$$P(e, s|y) = \frac{1}{Z} \prod_{x,k} e^{-(y_k^x - \sum_{i,j} e_i G_{ij,k} s_j^x)^2/(2\sigma^2)} P(e, s). \tag{10}$$

If the posterior is flat along any dimension in this region, there is a family of possible solutions to the maximization problem, whereas if the posterior is not flat along any dimensions, a finite number of local maxima must exist. To illustrate this point, we will consider a very simple toy example in which $d_E = 1$, $d_S = 1$ and $P(e, s)$ is uniform over some range $[\alpha, \beta] \times [\alpha, \beta]$. The posterior is then given by

$$P(a, b|y) = \begin{cases} \frac{1}{Z} e^{-(y - es)^2/(2\sigma^2)} & \text{if } \alpha < e, s < \beta \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

Figure 1 illustrates this distribution. There is a ridge of points, defined by the hyperbola $es = 1$, that maximize the posterior probability. At any one of these points we can see that there is one direction of zero curvature and one direction of negative curvature. In the following section, we describe a method of performing this sort of curvature analysis to examine the structure of the color constancy problem in a less trivial setting.

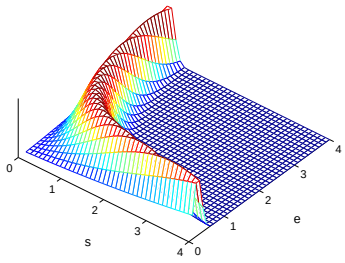


Figure 1: Depiction of the posterior for the toy color constancy problem, $y = es$. Note that at any local maximum of the posterior, there is zero curvature along one principle direction, and negative curvature along the other, indicating a local 1-d subspace of the parameters which explain the observed data equally well.

3.1 Local Curvature Analysis

We explore how the structure of the posterior probability reveals whether or not the color constancy problem is solvable. We assume in this analysis that the prior probability $P(e, s)$ varies slowly over the region being considered. In this case, the local structure of the posterior can be analyzed by examining the local structure of the likelihood function,

$$L(y|e, s) = -\frac{1}{2\sigma^2} \sum_{x,k} \left(y_k^x - \sum_{i,j} e_i G_{ij,k} s_j^x \right)^2. \tag{12}$$

The local structure we are interested in is the curvature of the log likelihood function, which tells us how moving in the parameter space affects our ability to explain the observed data. We will analyze how this structure changes when we introduce low-level spatial information into the model, and use this analysis to motivate the use of such information for color constancy.

The curvature of the log likelihood can be obtained from an eigen decomposition of its Hessian matrix [8]. The Hessian is a matrix of all second order derivatives and thus, for an N-pixel image, is a $(d_E + Nd_S) \times (d_E + Nd_S)$ matrix. The eigenvectors of the Hessian matrix give principal directions of curvature and the corresponding eigenvalues give the extent of curvature in each of these directions.

At a maximum of the log likelihood, all eigenvalues of the Hessian must be non-positive (meaning that there cannot be any positive curvature at this point since the log likelihood cannot increase as we move away from the local maximum). If there are any eigenvalues equal to zero, this implies that there are directions of zero curvature - e.g., there are ways we can vary the estimated parameters without affecting the likelihood of the data. For a unique maximum, all eigenvalues of Hessian must be negative at the maximum.

The amount of curvature of the Hessian also tells us something of how robust we can expect a parameter estimate to be in the presence of observation noise. Large eigenvalues of the Hessian imply a highly curved log likelihood, which, under the assumption of a relatively flat

prior, implies a highly curved posterior distribution. The larger the curvature of the posterior (i.e., the more "peaky" it is) the more robust we can expect it to be to observation noise in the observed data.

This notion of calculating the curvature of the Hessian of the log-likelihood is also related to the Fisher Information Matrix, except that we examine the Hessian of the log likelihood at a particular point, whereas the Fisher Information Matrix takes an expectation of the Hessian with respect to the parameters being estimated [23].

To help us understand the structure of the color constancy problem for real-world images, we first examine representative special cases in toy examples. Appendix B gives a full derivation of the Hessian for the log likelihood function given by (12).

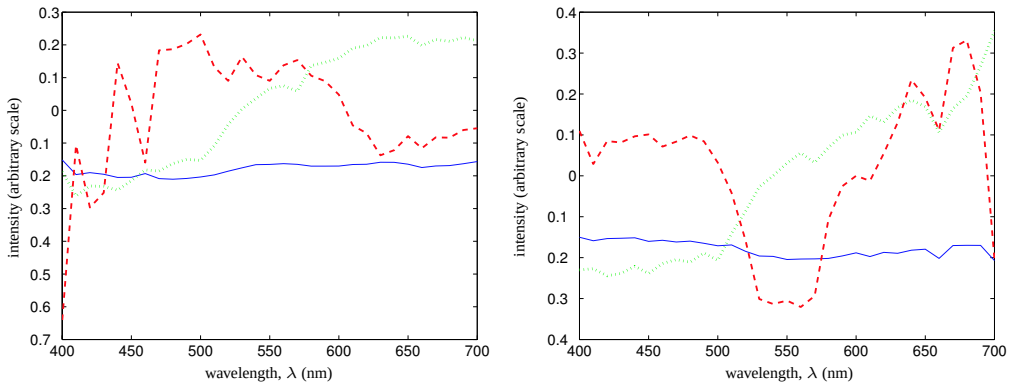
3.1.1 Underdetermined Case

We constructed a toy problem in which the sensor responses for all points in an image are generated using equation (8) with the noise term set to zero. To perform simulations, we used a hyperspectral data set of 28 natural scenes collected at Bristol University by Brelstaff, et. al. as described in [22, 4, 5]. Each of these images is 256×256 pixels in size and contains sensor responses to 31 spectral bands, ranging from 400 nm to 700 nm in 10 nm intervals. Each scene also contains a Kodak greycard at a known location with a constant reflectance spectrum of known intensity. The scene illuminant is approximated as the spectrum recorded at the location of the greycard divided by its constant reflectance. Our dataset therefore consists of 256x256x28 reflectance spectra and 28 illumination spectra.

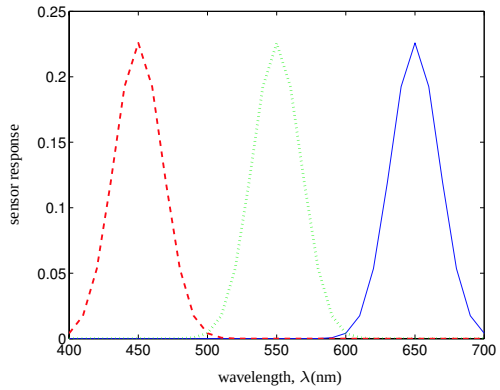
The first three illumination and reflectance basis functions obtained by applying principal components analysis (PCA) to this data are plotted in Figure 2(a) and 2(b), respectively (PCA is performed on all 28 illumination spectra and approximately half of the reflectance spectra). We assume, without loss of generality, a Gaussian model for sensor responses centered at 650 (red), 550 (green), and 450 nm (blue) as shown in Figure 2(c). Sample images from the dataset, after being passed through these Gaussian sensors, are shown in Figure 3.

To construct our toy problem, we draw 10 sample reflectance spectra from our dataset and project those spectra onto a 3-dimensional space spanned by the first three reflectance basis functions shown in Figure 2(b). We perform an analogous operation on one sample illumination spectrum and pass the low dimensional illumination and reflectances through the sensors of Figure 2(c) to yield sample image values. This is a toy example because the illumination and reflectance are described fully by 3-dimensional linear models, which is not the case for real images.

Evaluating the Hessian at the correct location in parameter space (i.e., the maximum of the log-likelihood) results in the eigenvalue spectrum shown in Figure 8, which shows that there are 3 eigenvalues equal to zero, and thus 3 directions of zero curvature. The three associated eigenvectors give the directions of zero curvature in parameter space. Any linear combination



(a) Illumination Basis Functions (b) Reflectance Basis Functions



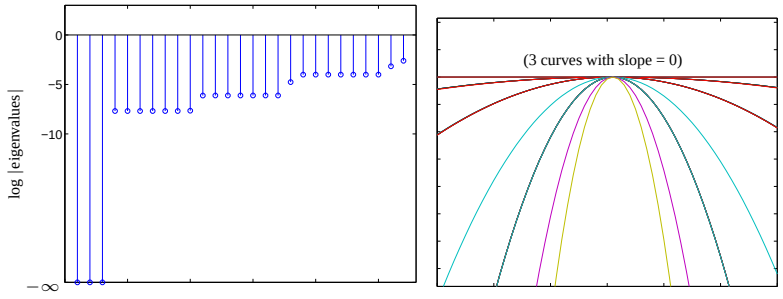
(c) Sensor Responses

Figure 2: Basis functions and sensor responses used in toy examples. Basis functions are obtained from a hyperspectral dataset.

Figure 3: Sample images from dataset of 28 hyperspectral images after being passed through the sensors of Figure 2(c). The hyperspectral images contain information from 31 spectral bands at each pixel, and the greycards in each of the images is used to find the true illumination of the scene.

of these eigenvectors is also a direction of zero curvature.

It is interesting to note that there are 3 illumination weights and also three eigenvectors with corresponding zero eigenvalues. Each of these eigenvectors is of dimension $d_E + Nd_S$, with three of those dimensions corresponding to the illuminant weights. The three eigenvectors span the space of the 3-dimensional illuminant space, meaning that for any set of illumination coefficients, there exists some set of reflectance weights such that the likelihood associated with that set of parameters is equal to the likelihood associated with the correct set of parameters. In other words, this means that we can move in any direction in the 3-dimensional space of the illuminant and still find surface reflectance values that render to the observed sensor values. One example of this is a multiplicative scaling up of the illumination and a corresponding scaling down of the surface reflectances.

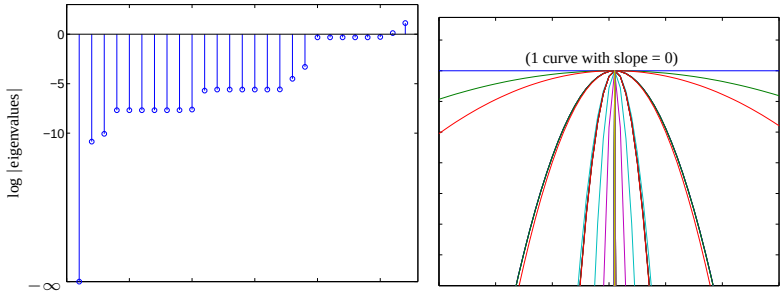


(a) Eigenvalue spectrum (b) Graphic depiction

Figure 4: Eigenvalue spectrum and corresponding curvature of the log likelihood for a toy example of the color constancy problem. Note that all eigenvalues are non-positive (the log of the absolute value is used on the right side plot for display purposes). Each of the curves on the left shows a projection on the log likelihood function onto a particular direction of principle curvature. The directions of zero curvature and eigenvalues at zero show that the problem is underdetermined

3.1.2 Sufficient Sensors

If there are 4 independent photoreceptor responses available instead of 3, we would expect for a better behaved log-likelihood function. However, there should still be one direction of zero curvature in the Hessian, corresponding to a simple scaling of the illumination and surfaces. To see this, we use the basis functions and sensor responses given in Figure 8 as well as a fourth sensor response (chosen to be random and thus independent of the other three) to construct another toy example. We again generate 10 random points and find that the Hessian behaves as expected, as shown in Figure 5. [[[[Not only does the Hessian have only one eigenvalue equal to zero, the other eigenvalues have a much larger value than the eigenvalues of the Hessian in the underdetermined case.]]]]



(a) Eigenvalue spectrum (b) Graphic depiction

Figure 5: Eigenvalue spectrum and corresponding curvature of the log likelihood for a case when the number of sensors exceeds the dimensionality of the surface model. Again, all eigenvalues are non-positive. There is now only one direction of zero curvature, corresponding to an unknown scale factor.

3.2 Bayesian Color Constancy in the Underdetermined Case

In general, the Bayesian approach to color constancy is able to produce a unique parameter estimate in the underdetermined case despite multiple degrees of ambiguity in the likelihood function. This is because the prior or the loss function will favor one solution over the others [2].

Often, however, the prior may be inaccurate for the particular image being examined. Ideally, we would like for our parameter estimate to be dominated by the effect of the likelihood function. That is, we would like for the estimate to be based on the data itself, not on prior assumptions about the data or how much we value particular parts of the data.

4 Spatio-spectral Basis Functions

We now introduce an alternative formulation of the finite-dimensional models presented in Section 2 that model both spatial and spectral properties of groups of pixels in natural images. The hypothesis is that by taking characteristic spatial variations of natural images into account, the number of parameters necessary to describe the reflectance of an image will be reduced so that the problem is no longer underdetermined.

Certain physical phenomena, such as interreflections, may generate characteristic spatial and spectral signatures which allow estimation of the desired spectra from image data. Natural images, like foliage, may exhibit characteristic spatial changes in color. The spatio-spectral basis functions allow us to exploit these, or other, regularities in the visual world in order to solve the color constancy problem. We expect this approach will work best in richly textured images, and worst in flat, color Mondrian-type images [21].

4.1 Modified Linear Models

Instead of using linear models to describe the reflectance of individual pixels, we will now use these models to describe groups of pixels. Without loss of generality, we can group pixels into $m \times m$ blocks and use the same basic formulation as was developed in Section 2. In order to do so, it is necessary to convert blocks of pixels into vector format. We do this by rasterizing the pixels within the block. The reflectance of a block of pixels is defined as a vector of length $m^2 L$ consisting of the reflectance of each pixel within the block stacked on top of each other in raster order. The same process is used to describe the sensor response of the block as a vector of length $m^2 K$ and the illumination as a vector of length $m^2 L$.

The basis functions used to model the reflectance of blocks of pixels are now referred to as *spatio-spectral* reflectance basis functions, since they describe both the spectral and spatial characteristics of a block of pixels.

We shall denote a group of pixels by $\boldsymbol{x}$, so that the illumination and reflectance of a block of pixels is given by:

$$\boldsymbol{E}(\gamma) = \sum_{i=1}^{m^2 L} \boldsymbol{E}_i(\gamma)\boldsymbol{e}_i, \tag{13}$$

$$\boldsymbol{S}^{\boldsymbol{x}}(\gamma) = \sum_{j=1}^{m^2 L} \boldsymbol{S}_j(\gamma)\boldsymbol{s}_j^{\boldsymbol{x}}, \tag{14}$$

where $\boldsymbol{E}_i(\gamma)$ and $\boldsymbol{S}_j(\gamma)$ are the spatio-spectral illumination and reflectance basis functions, $\boldsymbol{e}_i$ and $\boldsymbol{s}_j^{\boldsymbol{x}}$ are the weights associated with these basis functions, $\boldsymbol{E}(\gamma)$ is the illumination of all blocks in the image, and $\boldsymbol{S}^{\boldsymbol{x}}(\gamma)$ is the reflectance of the block of pixels $\boldsymbol{x}$. Note that the elements of the scene are now written as a function of γ rather than λ. This is due to the fact that the spatio-spectral representation contains information about both the frequency and spatial characteristics of the scene. Approximating these models with fewer dimensions, we can define

$$\tilde{\boldsymbol{E}}(\gamma) = \sum_{i=1}^{\boldsymbol{d_E}} \boldsymbol{E}_i(\gamma)\boldsymbol{e}_i, \tag{15}$$

$$\tilde{\boldsymbol{S}}^{\boldsymbol{x}}(\gamma) = \sum_{j=1}^{\boldsymbol{d_S}} \boldsymbol{S}_j(\gamma)\boldsymbol{s}_j^{\boldsymbol{x}}, \tag{16}$$

where $\tilde{\boldsymbol{E}}(\gamma)$ is the approximate illumination of all blocks in an image, constructed using a $\boldsymbol{d_E}$ dimensional model, and $\tilde{\boldsymbol{S}}^{\boldsymbol{x}}(\gamma)$ is the approximated reflectance for the block $\boldsymbol{x}$, constructed using a $\boldsymbol{d_S}$ dimensional model.

We define an $m \times m$ block of pixels as having $m^2 K$ sensor outputs, with K sensor outputs per pixel. Thus, we define the sensor responses of the group of pixels as the block diagonal

matrix

$$\boldsymbol{R}(\gamma) = \begin{bmatrix} R(\lambda) & 0 & \dots & 0 \\ 0 & R(\lambda) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & R(\lambda) \end{bmatrix} \tag{17}$$

with m^2 blocks along the diagonal, where

$$R(\lambda) = \begin{bmatrix} R_1(\lambda) & R_2(\lambda) & \dots & R_K(\lambda) \end{bmatrix}.$$

We let $\boldsymbol{R}_k(\gamma)$ refer to the kth column of the matrix $\boldsymbol{R}(\gamma)$.

Following a derivation analogous to that presented in Section 2, we can write the sensor output in a bilinear form with noise as:

$$\boldsymbol{y}_k^{\boldsymbol{x}} = \sum_{i=1}^{\boldsymbol{d_E}} \sum_{j=1}^{\boldsymbol{d_S}} \boldsymbol{e}_i \boldsymbol{G}_{ij,k} \boldsymbol{s}_j^{\boldsymbol{x}} + \boldsymbol{w}_k^{\boldsymbol{x}}, \tag{18}$$

where $\boldsymbol{y}_k^{\boldsymbol{x}}$ is the sensor response of the block of pixels $\boldsymbol{x}$, $\boldsymbol{G}_{ij,k}$ is defined as

$$\boldsymbol{G}_{ij,k} = \sum_{\gamma} \boldsymbol{R}_k(\gamma) \boldsymbol{E}_i(\gamma) \boldsymbol{S}_j(\gamma), \tag{19}$$

and $\boldsymbol{w}_k^{\boldsymbol{x}}$ is the noise introduced from using a lower-dimensional linear model.

4.2 Descriptive Power of Spatio-Spectral Basis Functions

The advantage of incorporating spatial information into the linear models used to describe scenes is that it provides more descriptive power using a fewer number of parameters. It is straightforward to show that even if natural images displayed the characteristic that individual pixels in an image are independent of one another, and incorporating spatial information provides no benefit, spatio-spectral basis functions provide equal descriptive power to standard spectral basis functions. For real images, spatio-spectral basis functions in fact describe images much better than standard spectral basis functions, due to the correlation that exists between neighboring pixels in natural images.

The amount of improvement in descriptive power provided by the new method can be tested by using real images to find linear reflectance bases for blocks of various sizes (including $m = 1$, which corresponds to the traditional case of examining each pixel independently) and, in each case, calculating the squared reconstruction error as the dimensionality of the reflectance model is varied. To do this, we use the dataset of hyperspectral images described earlier in Section 3.1.1. Approximately half of the data in the 28 scenes (selected randomly) is used to find the linear model, and the squared reconstruction error is calculated using the

remaining data.

The first three components of the reflectance basis corresponding to $m = 1$ was shown in Figure 2(b), and Figure 6 plots the first 12 spatio-spectral reflectance basis functions for $m = 2$. These basis functions were extracted from the hyperspectral data using the calibration gray card in each image of the dataset, in the same manner as described in Section 3.1.1. The correlation between neighboring pixels in the $\times 2$ block is apparent when examining these basis functions.

Figure 6 shows the first 12 components of the reflectance basis corresponding to $m = 2$. A comparison of squared reconstruction error when considering block sizes of $m =$1, 2 and 4 is shown in Figure 7.

It can be seen that using a block size of $m = 2$ dramatically improves performance. For example, the data shows that the error rate obtained when describing each pixel with 3 basis functions can be achieved by using only 1.25 basis functions per pixel when describing an image with 2×2 blocks (i.e., 5 basis functions for the block). Increasing the block size beyond $m = 2$ shows even further improvement in the ability to describe the data using fewer parameters, since more of the image correlations can be described by the larger blocks.

4.3 Curvature Analysis Using Spatio-Spectral Basis Functions

To see the advantage of using spatial information in the context of Bayesian color constancy, we must once again analyze the local curvature of the posterior probability. Following the derivations in Section 3, we once again make the assumption of a locally flat prior distribution, allowing us to simplify the problem and examine the log likelihood function, now written as

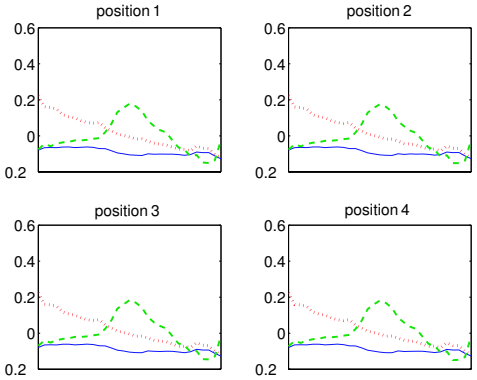
$$L(\boldsymbol{y}|\boldsymbol{e},\boldsymbol{s}) = \sum_{\boldsymbol{x},k} \left(\boldsymbol{y}_k^{\boldsymbol{x}} - \sum_{i,j} \boldsymbol{e}_i \boldsymbol{G}_{ij,k} \boldsymbol{s}_j^{\boldsymbol{x}} \right)^2. \tag{20}$$

We again consider two toy problems, now for the case of 2×2 pixel blocks instead of single pixels. Figures **??** and **??** show the eigenvalues of the Hessian matrix in the underdetermined and overdetermined cases, respectively. In both cases we have 12 sensor responses available and we consider 12 pixels (3 blocks) and an illumination projected down into three dimensions. For the underdetermined case, reflectances are 12 dimensional whereas in the overdetermined case they are 11 dimensional.

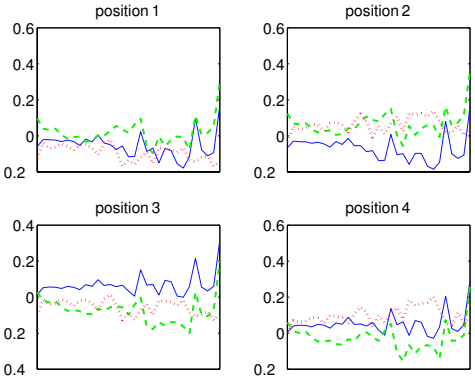
5 Experiments with Natural Images

We look at groups of 40 pixels, grouped into 10 sets of 2x2 pixel blocks. The 10 sets were selected from randomized positions in the image.

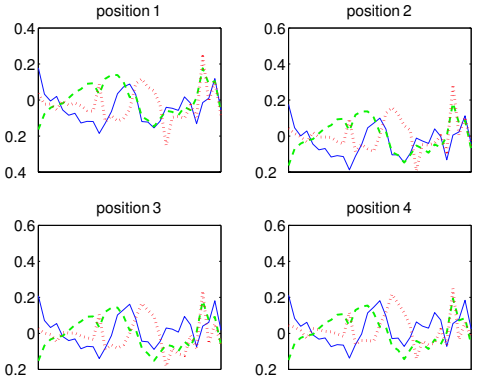
In order to get the priors, we project our dataset of reflectances and illuminations onto the



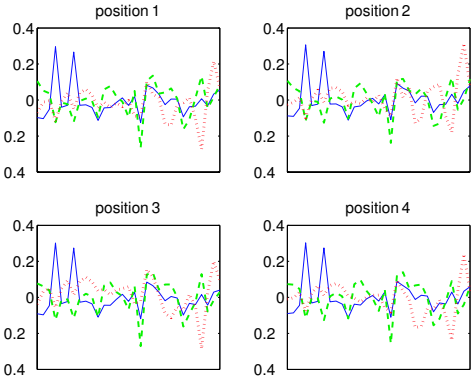
(a) basis functions 1, 2, and 3



(b) basis functions 4, 5, and 6



(c) basis functions 7, 8, and 9



(d) basis functions 10, 11, and 12

Figure 6: The first 12 spatio-spectral basis functions for 2x2 pixel blocks. Each basis function is plotted for wavelength values from 400 to 700 nm on the x-axis. The first 3 spatio-spectral basis functions, (a), show only spectral variations and no spatial variation. The next 3 basis functions, (b), correspond to spatial derivatives. The final 6 indicate a more complex relationship of spatial and spectral variations, (c) and (d).

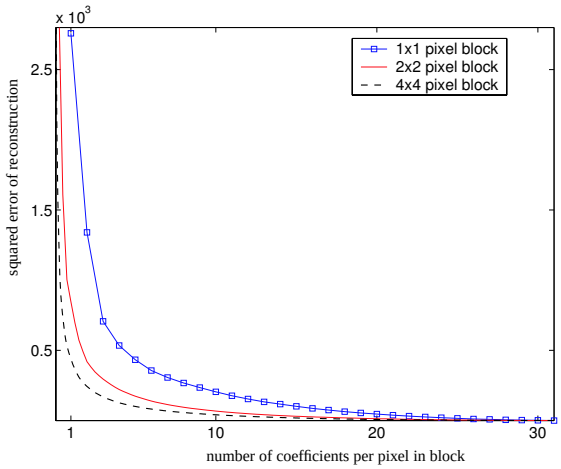
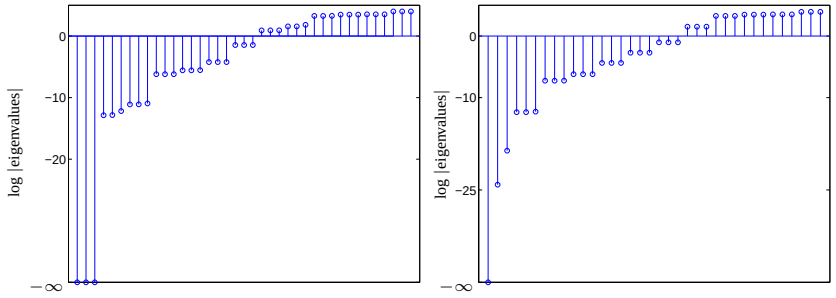


Figure 7: The squared reconstruction error using linear models of varying dimensions for block sizes of m =1, 2, 3, and 4. The x-axis is equivalent to the number of basis functions used to describe the $m \times m$ block divided by m^2. Using block sizes of $m = 2$ or larger greatly reduces the number of coefficients needed to describe each pixel.

basis functions given in Figures 2(a) (for illumination) and 6 (for spatio-spectral reflectance). We fit truncated Gaussians to the data and use these as our priors.

Our goal is now to find the set of illumination and reflectance coefficients that minimize the posterior probability. In order to do this, we need to search over the space of all possible illumination and surface reflectance coefficients. Since the number of surface coefficients scales with the number of pixels in the image, this quickly becomes an intractable problem.

Following [2], in order to make the problem more feasible, we limit our search to look over the space of illumination coefficients and solve for the surfaces at each iteration using a deterministic relationship. In the underdetermined case, where the number of sensor responses



(a) Underdetermined case (b) Overdetermined case

Figure 8: Eigenvalue spectra of the log likelihood function when using 2×2 blocks in the underdetermined case (left) and overdetermined case (right). As was the case when considering single pixels, in the underdetermined case, there are three directions of zero curvature and in the overdetermined case there is only one direction of zero curvature.

at each pixel is equal to the number of surface coefficients at each pixel, we solve for surfaces that produce the sensor responses exactly given an illumination (this is done by solving equation (8) assuming no sensor noise through a simple matrix inversion). This is equivalent to constraining our search to locations of maximum probability in the likelihood. In the case where we have more sensor responses than surface coefficients, we take the pseudo-inverse to find the surface coefficients corresponding to a given illumination.

Unfortunately, applying this technique of constraining the search space to real images gives numerical problems. In these images, the noise term in equation (8), which corresponds to the surface and illumination components not included in a low-dimensional model, causes the surface coefficients that maximize the likelihood term given the true illuminant to be very different from the true surfaces coefficients. This makes it hard to properly analyze the posterior probability corresponding to a given illuminant.

To avoid the numerical issues mentioned above, we test our algorithm on images constructed by first projecting the true surface reflectances down to 11 dimensions out of 124 in the 2×2 block model, multiplying the reflectance image by the true illuminant, and passing the result through the Gaussian sensors shown in Figure 2(c). To provide a fair comparison between our algorithm using spatial information and te existing Bayesian and Gray World algorithm that rely on single pixels, we also find a new set of single-pixel reflectance basis functions and prior probabilities that apply to this modified dataset.

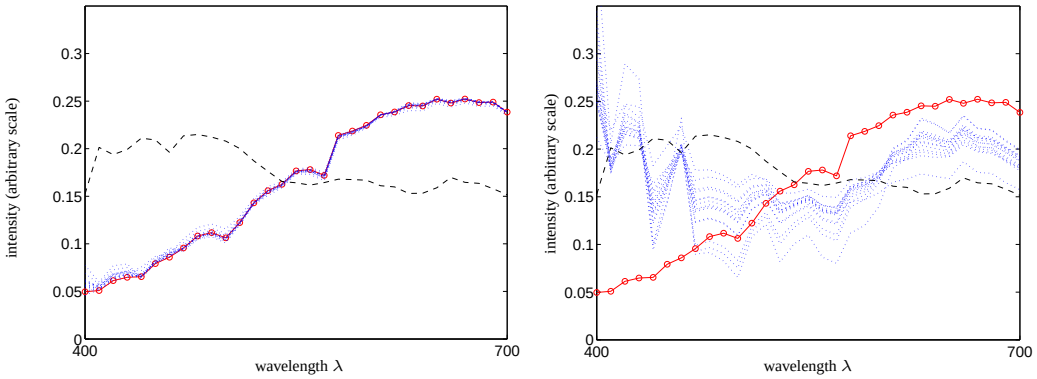
Figure 9 shows that there is very little visual difference between the image obtained using the true reflectance and the image obtained after projecting the reflectance onto the lower-dimensional subspace. We need to be clear: these are pre-processed images, not "real" images, although they are tantalizingly close to real images. We show the benefits of using spatio-spectral basis functions in solving color constancy using the pre-processed images, and expect that extensions to real images will follow in work by us or by others.

(a) Sample image from dataset (b) Image after projection

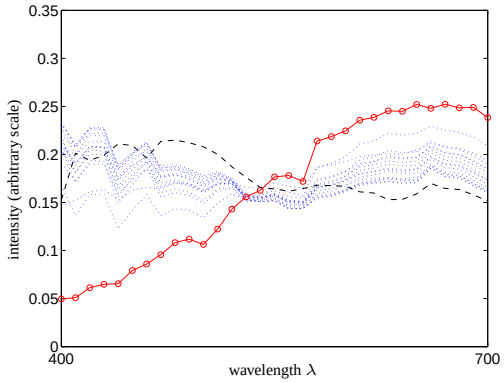
Figure 9: The visual effect of projecting reflectance to lower dimensions is very small.

Figure 10 shows results for our approach using spatio-spectral basis functions in a Bayesian

approach with the standard Bayesian approach and the Gray World approach. Each algorithm is run on 15 random groups of 40 pixels chosen from the image. The particular image we consider is not fit very well by the priors, which is why the standard Bayesian approach, examining one pixel at a time, and the Gray World approach both result in incorrect estimates. Despite the low prior probabilities, however, the Bayesian approach using spatio-spectral basis functions allows us to locate the correct illuminant for each draw. This illustrates that the likelihood term does dominate our estimate, as we would expect.



(a) using spatio-spectral basis functions (b) standard Bayesian approach



(c) Gray World algorithm

Figure 10: Results of color constancy applied to a natural image. Each algorithm was run 15 times, with results shown in dotted lines. The mean illumination from the prior is given in the dashed line as reference, and the true illuminant is marked by circles.

6 Conclusions

We have studied the effect of using spatio-spectral surface reflectance basis functions to find a low-dimensional estimate of surface reflectance characteristics.

These basis functions can convert the color constancy problem from an under-determined problem to an over-determined one. We can see this in the curvature characteristics of the

log-likelihood function at parameter local maxima.

For natural images, pre-processed to restrict the surface reflectance variations to be explained by an average of 11 basis functions per 2×2 pixel block, we find that using spatio-spectral basis functions allows for very accurate estimates of the global illumination spectrum (and therefore of surface reflectance spectra, too). For the same images, using purely spectral basis functions results in an underdetermined problem, with much higher estimation error for the optimal Bayesian estimate.

Acknowledgments

We thank Antonio Torralba for many helpful conversations. This work was funded by the Nippon Telegraph and Telephone Corporation as part of the NTT/MIT Collaboration Agreement. Barun Singh was supported by a National Science Foundation graduate fellowship.

APPENDIX: The Hessian of the Log Likelihood Functions with Bilinear Forms

$H_{e_l e_m} = \frac{\partial^2 L}{\partial e_l \partial e_m}$

$$\frac{\partial L}{\partial e_l} = \frac{\partial}{\partial e_l} \sum_{x,k} (y_k^x - \sum_{i,j} e_i G_{ij,k} s_j^x)^2 \quad \text{(A-1)}$$

$$= \sum_{x,k} \frac{\partial}{\partial e_l} (y_k^x - \sum_{i,j} e_i G_{ij,k} s_j^x)^2 \quad \text{(A-2)}$$

$$= -2 \sum_{x,k} \Big((y_k^x - \sum_{i,j} e_i G_{ij,k} s_j^x) (\frac{\partial}{\partial e_l} \sum_{i,j} e_i G_{ij,k} s_j^x) \Big) \quad \text{(A-3)}$$

$$= -2 \sum_{x,k} \Big((y_k^x - \sum_{i,j} e_i G_{ij,k} s_j^x) \sum_j G_{lj,k} s_j^x \Big) \quad \text{(A-4)}$$

$$\frac{\partial 2L}{\partial e_l \partial e_m} = -2 \frac{\partial}{\partial e_m} \sum_{x,k} \Big((y_k^x - \sum_{i,j} e_i G_{ij,k} s_j^x) \sum_j G_{lj,k} s_j^x \Big) \quad \text{(A-5)}$$

$$= -2 \sum_{x,k} \frac{\partial}{\partial e_m} \Big((y_k^x - \sum_{i,j} e_i G_{ij,k} s_j^x) \sum_j G_{lj,k} s_j^x \Big) \quad \text{(A-6)}$$

$$= 2 \sum_{x,k} \Big((\sum_j G_{lj,k} s_j^x) (\sum_j G_{mj,k} s_j^x) \Big) \quad \text{(A-7)}$$

$H_{s_l^n s_m^p} = \frac{\partial^2 L}{\partial s_l^n \partial s_m^p}$

$$\frac{\partial L}{\partial s_l^n} = \frac{\partial}{\partial s_l^n} \sum_{x,k} (y_k^x - \sum_{i,j} e_i G_{ij,k} s_j^x)^2 \tag{A-8}$$

$$= \frac{\partial}{\partial s_l^n} \Big(\sum_{x \neq n} \sum_k (y_k^x - \sum_{i,j} e_i G_{ij,k} s_j^x)^2 + \sum_{x=n} \sum_k (y_k^x - \sum_{i,j} e_i G_{ij,k} s_j^x)^2 \Big) \tag{A-9}$$

$$= \frac{\partial}{\partial s_l^n} \sum_k (y_k^n - \sum_{i,j} e_i G_{ij,k} s_j^n)^2 \tag{A-10}$$

$$= -2 \sum_k \Big((y_k^n - \sum_{i,j} e_i G_{ij,k} s_j^n) (\frac{\partial}{\partial s_l^n} \sum_{i,j} e_i G_{ij,k} s_j^n) \Big) \tag{A-11}$$

$$= -2 \sum_k \Big((y_k^n - \sum_{i,j} e_i G_{ij,k} s_j^n) (\sum_i e_i G_{il,k}) \Big) \tag{A-12}$$

$$\frac{\partial^2 L}{\partial s_l^n s_m^p} = 0 \tag{A-13}$$

$$\frac{\partial^2 L}{\partial s_l^n s_m^n} = \frac{\partial}{\partial s_m^n} \Big(-2 \sum_k \Big((y_k^n - \sum_{i,j} e_i G_{ij,k} s_j^n) (\sum_i e_i G_{il,k}) \Big) \Big) \tag{A-14}$$

$$= 2 \sum_k \Big(\Big(\frac{\partial}{\partial s_m^n} (\sum_{i,j} e_i G_{ij,k} s_j^n) \Big) (\sum_i e_i G_{il,k}) \Big) \tag{A-15}$$

$$= 2 \sum_k \Big((\sum_i e_i G_{im,k}) (\sum_i e_i G_{il,k}) \Big) \tag{A-16}$$

$H_{e_l s_m^p} = \frac{\partial^2 L}{\partial e_l \partial s_m^p}$

$$\frac{\partial 2L}{\partial e_l \partial s_m^p} = \frac{\partial}{\partial e_l} \Big(-2 \sum_k \Big((y_k^p - \sum_{i,j} e_i G_{ij,k} s_j^p) (\sum_i e_i G_{im,k}) \Big) \Big) \tag{A-17}$$

$$= -2 \sum_k \Big(\Big(\frac{\partial}{\partial e_l} (y_k^p - \sum_{i,j} e_i G_{ij,k} s_j^p) \Big) \Big(\sum_i e_i G_{im,k} \Big) + \Big(y_k^p - \sum_{i,j} e_i G_{ij,k} s_j^p \Big) \Big(\frac{\partial}{\partial e_l} \sum_i e_i G_{im,k} \Big) \Big) \tag{A-18}$$

$$= 2 \sum_k \big((\sum_j G_{lj,k} s_j^p) (\sum_i e_i G_{im,k}) - (y_k^p - \sum_{i,j} e_i G_{ij,k} s_j^p) G_{lm,k} \big) \tag{A-19}$$

References

[1] J. O. Berger. *Statistical decision theory and Bayesian analysis*. Springer, 1985.

[2] D. H. Brainard and W. T. Freeman. Bayesian color constancy. *Journal of the Optical Society of America A*, 14(7):1393–1411, 1997.

[3] D H Brainard and B A Wandell. Analysis of the retinex theory of color vision. *Journal of the Optical Society of America A*, 3(10):1651–1661, 1986.

[4] G. J. Brelstaff, C. A. Parraga, T. Troscianko, and D. Carr. Hyperspectral camera system: acquistion and analysis. *SPIE*, 2587, 1995.

[5] G. J. Brelstaff, C. A. Parraga, T. Troscianko, and I. Moorhead. Color and luminance information in natural scenes. *JOSA-A*, 1997.

[6] G. Buchsbaum. A spatial processor model for object colour perception. *J. Franklin Inst.*, 310:1–26, 1980.

[7] CIE. Colorimetry. Bureau Central de la CIE, 1986.

[8] M. DoCarmo. *Differential Geometry of Curves and Surfaces*. Prentice Hall, 1976.

[9] M. D'Zmura, G. Iverson, and B. Singer. Probabilistic color constancy. In R. D. Luce et. al., editor, *Geometric representations of perceptual phenomena: papers in honor of Tarow Indow's 70th birthday*. Lawrence Erlbaum, Hillsdale, NJ. in press.

[10] R. M. Evans. *The perception of color*. Wiley, New York, 1974.

[11] G. Finlayson and S. Hordley. Improving gamut mapping color constancy. *Image and Vision Computing*, 9:1774–1783, 2000.

[12] D. Forsyth. A novel approach to color constancy. *International Journal of Computer Vision*, 5:5–36, 1990.

[13] D. A. Forsyth. A novel approach to colour constancy. In *Proc. 1st Intl. Conf. Computer Vision*, pages 9–18. IEEE, 1988.

[14] R. Gershon and A. D. Jepson. The computation of color constant descriptors in chromatic images. *Color Res. Appl.*, 14:325–334, 1989.

[15] T. Jaaskelainen, J. Parkkinen, and S. Toyooka. A vector-subspace model for color representation. *Journal of the Optical Society of America A*, 7:725–730, 1990.

[16] D. B. Judd, D. L. MacAdam, and G. W. Wyszecki. Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.*, page 1031, 1964.

[17] E H Land. The retinex theory of color vision. *Scientific American*, 237(6):108–128, 1977.

[18] E. H. Land. Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image. *Proc. Nat. Acad. Sci. USA*, 80:5163–5169, 1983.

[19] L. T. Maloney. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal of the Optical Society of America A*, 3(10):1673–1683, 1986.

[20] L T Maloney and B A Wandell. Color constancy: a method for recovering surface spectral reflectance. *Journal of the Optical Society of America A*, 3(10):29–33, 1986.

[21] J. J. McCann, S. P. McKee, and T. H. Taylor. Quantitative studies in retinex theory: a comparison between theoretical predictions and observer responses to the 'color mondrian' experiments. *Vision Research*, 16:445–458, 1976.

[22] C. A. Parraga. Spatichromatic infomation content of natural scenes. Master's thesis, University of Bristol UK, 1995.

[23] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I*. Wiley-Interscience, 2001.

[24] H. J. Trussell and M. J. Vrhel. Estimation of illumination for color correction. In *Proc. ICASSP*, pages 2513–2516, 1991.